The development of tools to allow genetic and genomic analysis of Frankia

Tania Rawnsley Spenlinhauer

University of New Hampshire, Durham

Follow this and additional works at: https://scholars.unh.edu/dissertation

Recommended Citation
https://scholars.unh.edu/dissertation/415

This Dissertation is brought to you for free and open access by the Student Scholarship at University of New Hampshire Scholars' Repository. It has been accepted for inclusion in Doctoral Dissertations by an authorized administrator of University of New Hampshire Scholars' Repository. For more information, please contact nicole.hentz@unh.edu.
THE DEVELOPMENT OF TOOLS TO ALLOW GENETIC AND GENOMIC ANALYSIS OF FRANKIA

BY

TANIA RAWNSLEY SPENLINHAUER
Bachelor's of Science, University of New Hampshire, 1994

DISSERTATION

Submitted to the University of New Hampshire
in Partial Fulfillment of
the Requirements for the Degree of

Doctor of Philosophy
in
Microbiology

December, 2007
INFORMATION TO USERS

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleed-through, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

UMI Microform 3290108
Copyright 2008 by ProQuest Information and Learning Company. All rights reserved. This microform edition is protected against unauthorized copying under Title 17, United States Code.

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
This dissertation has been examined and approved.

Dissertation Director, Louis S. Tisa, Associate Professor of Microbiology

Robert Zsigay, Professor Emeritus of Microbiology

Elise Sullivan, Clinical Assistant Professor of Microbiology

Estelle Hrabak, Associate Professor of Plant Biology

Arnab Sen, Professor of Microbiology
University of North Bengal

November 30, 2007
Date
DEDICATION

I dedicate this dissertation to my father, Bruce J. Rawnsley, for his unconditional love and support. I am proud of who he was as a man, husband and father and hope that I have made him proud of who I have become as a woman, wife and mother.
ACKNOWLEDGEMENTS

This dissertation was supported by funding from NHAES Hatch 377, Department of Energy, National Science Foundation EF-0333173, and USDA 2003-35319-13787. I would also like to thank my advisor, Dr. Louis S. Tisa, for his guidance and encouragement, my committee members, Dr. Robert Zsigay, Dr. Elise Sullivan, Dr. Estelle Hrabak, and Dr. Arnab Sen, the entire Department of Microbiology, especially, Robert Mooney, for all of their advise and support. Above all I would like to thank my entire family for always being there for me, especially my loving and devoted husband, Eric Spenlinhauer and our two beautiful children, Brady and Tia.
### TABLE OF CONTENTS

| DEDICATION | iii |
| ACKNOWLEDGEMENTS | iv |
| LIST OF TABLES | vii |
| LIST OF FIGURES | viii |
| ABSTRACT | x |

<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. INTRODUCTION AND LITERATURE REVIEW</td>
<td>1</td>
</tr>
<tr>
<td>Frankia and Actinorhizal Plants</td>
<td>1</td>
</tr>
<tr>
<td>Frankia Isolates Used in This Study</td>
<td>5</td>
</tr>
<tr>
<td>Overview of Frankia Genetics</td>
<td>7</td>
</tr>
<tr>
<td>Overview of Frankia Genomics</td>
<td>8</td>
</tr>
<tr>
<td>Research Goals</td>
<td>9</td>
</tr>
<tr>
<td>II. DEVELOPMENT OF A PHYSICAL MAP FOR THREE FRANKIA STRAINS AND A PARTIAL GENETIC MAP FOR FRANKIA EU1C</td>
<td>10</td>
</tr>
<tr>
<td>Introduction</td>
<td>10</td>
</tr>
<tr>
<td>Materials and Methods</td>
<td>12</td>
</tr>
<tr>
<td>Results</td>
<td>14</td>
</tr>
</tbody>
</table>
LIST OF TABLES

Chapter II.

Table 1. List of primer sets used to generate specific gene probes ..................13
Table 2. Sizes of individual restriction fragments for Frankia gDNA digested with different restriction endonucleases .............................................16
Table 3. Sizes of individual restriction fragments used for the determination of the physical maps of Frankia Ccl3 and EAN1pec ............................19
Table 4. Sizes of individual restriction fragments used for Frankia Eu1c physical map determination .................................................................21

Chapter III.

Table 1. Relationship between actinorhizal plants and groups of Frankia strains ..................................................................................................28
Table 2. Summary of genome characteristics ..............................................................................................................................29
Table 3. Summary of genes involved in size differentiation of the three Frankia strains ..............................................................................30
Table 4. Categories of deleted genes assessed by BlastClust analysis .....30
Table 5. Top 20 families of duplicated genes in each Frankia strain assessed by BlastClust analysis ..............................................................31

Chapter IV.

Table 1. Strains used .........................................................................................50
Table 2. Primers used in this study ........................................................................51
Table 3. Transfer of Tn916 to Frankia Eu1c. From Myers 2001 ..................51
### LIST OF FIGURES

#### Chapter II.

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>PFG analysis of <em>Frankia</em> Eul1c digested with restriction enzymes</td>
<td>15</td>
</tr>
<tr>
<td>2</td>
<td>PFG analysis of <em>Frankia</em> Ccl3 digested with restriction enzymes</td>
<td>17</td>
</tr>
<tr>
<td>3</td>
<td>PFG analysis of digested and undigested <em>Frankia</em> DNA</td>
<td>17</td>
</tr>
<tr>
<td>4</td>
<td>The <em>Frankia</em> chromosome has a circular topology</td>
<td>17</td>
</tr>
<tr>
<td>5</td>
<td>Physical maps for the circular chromosomes of <em>Frankia</em> Ccl3 and EAN1pec</td>
<td>18</td>
</tr>
<tr>
<td>6</td>
<td>Physical and genetic map of <em>Frankia</em> Eul1c chromosome</td>
<td>20</td>
</tr>
</tbody>
</table>

#### Chapter III.

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Neighboring-joining phylogenetic tree calculated with ClustalX 1.83</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td>from 16S rRNA gene sequences</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Present-day native distribution of actinorhizal plant hosts</td>
<td>29</td>
</tr>
<tr>
<td>3</td>
<td>Genome maps of the three <em>Frankia</em> strains</td>
<td>30</td>
</tr>
</tbody>
</table>

#### Chapter IV.

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Photograph of recipient and transconjugant colonies on medium</td>
<td>52</td>
</tr>
<tr>
<td></td>
<td>with NH₄Cl as the nitrogen source</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>PCR amplification of the transposon Tn916</td>
<td>52</td>
</tr>
<tr>
<td>3</td>
<td>PCR amplification of the glutamine synthetase gene found in <em>Frankia</em> strains</td>
<td>53</td>
</tr>
<tr>
<td>4</td>
<td>PCR amplification of the conjugal transfer protein found in <em>E. faecalis</em></td>
<td>53</td>
</tr>
<tr>
<td>5</td>
<td>DGGE of the 16S rRNA for recipient <em>Frankia</em> Eul1c and EAN1pec, transconjugants, and <em>E. faecalis</em></td>
<td>54</td>
</tr>
<tr>
<td>6</td>
<td>PFGE of <em>Frankia</em> and transconjugants digested with restriction</td>
<td></td>
</tr>
</tbody>
</table>
endonuclease...54

Figure 7. Southern blot hybridization of *Frankia* and transconjugants digested with *Scal*...55

Appendix

Figure A1. Cloning vector pBeloBAC11...80

Figure A2. Partial digest of *Frankia* isolate Ccl3 by restriction enzyme *Sau3A*I run on a 1.0% gel, 5.5 V/cm, 5-20sec switch time for 16.5 h....80
ABSTRACT
THE DEVELOPMENT OF TOOLS TO ALLOW GENETIC AND GENOMIC
ANALYSIS OF FRANKIA

By
Tania Rawnsley Spenlinhauer

University of New Hampshire, December, 2007

Frankia, a nitrogen-fixing actinomycete, forms a symbiotic association with a variety of woody dicotyledonous plants. The lack of standard genetic tools for this important bacterium has hindered studies on the molecular biology of this symbiosis. We are interested in developing tools for the genetic analysis of Frankia physiology and its interactions with its host plants. Our approach has focused on the development of a physical and genetic map of the Frankia chromosome by macrorestriction analysis. Agarose-embedded chromosomal DNA plugs were prepared from three Frankia isolates: Eul1c, EAN1pec and Ccl3. The genome sizes of the three strains were determined by pulse-field gel electrophoresis (PFGE) and estimated to be 5.5Mb for Ccl3, 9.1Mb for EAN1pec and 8.5Mb for Eul1c. Several members of the actinomycetales have linear chromosomes; however, our experiments on topology show that the Frankia chromosome is a single circular chromosome. The physical map was achieved with the enzymes Asel, Sspl and Swal for Ccl3 and EAN1pec and the enzymes Asel and Pmel for Eul1c isolate. Also a partial genetic map was constructed for isolate, Eul1c, by locating the known
Frankia genes including: ginA, ginII, a truncated hemoglobin, and the 16srRNA. The development of these physical maps of Frankia has provided a stepping-stone to facilitate the sequencing of the entire genome of isolates, EAN1pec and Ccl3.

Another important step in further understanding the genetics of this important symbiosis is establishing a protocol for genetic transfer and a reliable system for standardize mutagenesis. This was achieved through the use of a conjugative transposon, Tn916, which was introduced into Frankia isolate Eul1c by mating with Enterococcus faecalis GC110. The generation of stable tetracycline and novobiocin resistant transconjugants suggests that the transposon inserted into the genome. Results from PCR and Southern Blot Hybridization experiments confirmed the insertion of the tetM gene. The transposon insertion sites have been mapped to the physical map of the Frankia chromosome that we developed by macrorestriction analysis. The development of the physical map along with the first successful transfer and expression of foreign DNA into Frankia is a major stepping-stone in the advancement of knowledge of this important prokaryote.
CHAPTER I

INTRODUCTION AND LITERATURE REVIEW

_Frankia_ and Actinorhizal Plants

Many human activities, including gravel operations, mining, and extensive deforestation, deplete vital natural resources that are needed for the survival of various ecological communities. Within these nutrient-poor habitats, one of the principle limiting factors in the growth of agronomic and forest crops is a low level of available nitrogen (Benoit and Berry, 1990). Many pioneer community plants are only able to survive these nutrient-poor environments by forming a symbiotic relationship with indigenous soil microorganisms (Benoit & Berry, 1990, Dommergues, 1995). One of these important relationships involves the soil microorganism, _Frankia_, and its interactions with its actinorhizal host plants (Benson and Silvester, 1993, Schwencke and Carú, 2001).

_Frankia_, a member of the bacterial family actinomycetes, is distinguished by its unique ability to induce nitrogen fixation in root nodules of certain nonleguminous plants (Berry 1994, Clawson _et al._ 1998, Torrey and Tjepkema 1979). These bacteria grow symbiotically with over 200 species of woody dicotyledonous plants in eight families known as actinorhizal plants (Benson 1988, Schwencke and Carú, 2001). This symbiotic relationship provides a significant contribution to the nitrogen budget of the planet. The ecological range of actinorhizal plants is quite diverse and
varies from dry sandy, desert areas to wetlands and forests (Benson and Silvester 1993, Schwencke and Carú, 2001).

The ability of actinorhizal plants to survive in such diverse and nutrient-poor conditions permits an immense potential to use these plants in many areas of ecological growth and prosperity (Schwencke and Carú, 2001, Simpson et al., 1996). Actinorhizal plants are of economic significance with respect to land reclamation, soil stabilization, landscaping, fuel, nurse cropping, and have been used for commercial lumber (Akkermans et al., 1991). Frankia also has the ability to bind and sequester several toxic heavy metals, including chromate, selenite, copper and lead (Richards et al., 2002). This property gives Frankia the potential for bioremediation and phytoremediation applications especially on heavy metal contaminated lands. Currently, in Western Europe, actinorhizal plants are used extensively for reclamation of industrial wastelands and mine spoils (Wheeler et al., 2000).

In order to take full advantage of its economic and ecological potential, it is essential that the physiology and genetics of Frankia be thoroughly investigated. Since the first isolation of Frankia in 1978 (Torrey and Tjepkema 1979), many advances have occurred in the knowledge and understanding of actinorhizal nitrogen fixation regarding the physiology, morphology, and biochemistry of these bacteria (Benson and Silvester 1993), whereas, little endeavor has been made to investigate the genetics of Frankia.

Although many advances have occurred in understanding the symbiosis between Frankia and its host plant there is little known about the molecular aspects
of this interaction. Many attempts to understand this process have been conducted by comparing it with the well studied *Rhizobium*-legume symbiosis, assuming that symbiotic signal molecules may be conserved (Schwencke and Carú, 2001). The formation of root nodules in both *Frankia* and *Rhizobium* is a complex process that occurs in several stages, which require a series of interactions and chemical signaling between the bacteria and host root cells. These signals include chemoattractants, gene activators, phytohormones, and many others. *Frankia* is known to use two general pathways for infecting its host plants: (1) root hair infection or intracellular infection, as seen in *Alnus*, *Myrica*, *Comptonia*, and *Casuarina* (Callaham and Torrey 1977, Berry, McIntyre and McCully 1986, Schwencke and Carú, 2001); and (2) intercellular penetration, as seen in *Elaeagnus*, *Ceanothus*, and *Cercocarpus* (Liu and Berry, 1991, Schwencke and Carú, 2001). Since it has been shown that a single isolate of *Frankia* can use different infection pathways when nodulating different host species indicating that the mechanism of infection is controlled by the host plant (Racette and Torrey 1989). However, the signal molecules involved are still largely unknown.

The first stage of intracellular infection is root hair curling and branching which favors the binding of *Frankia* cells to the root hair wall (Berry and Torrey 1983) and allows for penetration via an infection thread into the cortical tissue of the host plant (Callaham and Torrey 1977). The infection thread that is composed of a mass of bacteria is encapsulated by plant cell-derived material (Lalonde and Knowles, 1975). In response to this invasion, the cortical cells of the host plant are induced to divide and produce a prenodule (Berry and Sunell, 1990). *Frankia* filaments continue to
infect more cells and cause large clusters of bacteria-infected plant cells and the
host plant forms lateral root nodules.

In the intercellular pathway the infection process occurs in the absence of
roots hairs and the infective Frankia hyphae penetrate the root epidermal cells
directly (Miller and Baker, 1985, Benson and Silvester, 1993). Once in the root, the
bacterium invades the intercellular spaces toward the nodule lobe. The end of
infection in either pathway leads to the growth of several lateral root-like structures.
Within these mature root nodules the effective Frankia is able to form vesicles, which
become the site for nitrogen fixation.

In addition to forming a symbiotic relationship with plants, another striking
feature of Frankia is its ability to develop two unique morphological structures:
vesicles and spores. Spores are the reproductive structure of Frankia and are
formed in multilocular membrane-bound sporangia which are produced terminally at
the mycelial tips (Callaham et al., 1978). Each sporangium contains hundreds of
refractile spores. When mature, the spores are released form the sporangia and are
presumed to aid in the survival and dissemination of Frankia (Benson and Silvester,
1993). The size and shape of the spores vary considerably among different strains
as well as the ability to produce spores either in culture or in planta (Benson and
Silvester, 1993, Krumholz et al., 2003). In addition to the important ecological role in
the life cycle of filamentous organisms, spores provide a large population of uni-
genomic haploid cells and are a powerful tool for their genetic studies (Krumholz et
al., 2003).
Vesicles are ovoid, club-shaped or pear-shaped structures that form terminally on short side branches of hyphae and are the site of nitrogen fixation (Benson and Silvester, 1993). Because *Frankia* is able to fix nitrogen under aerobic conditions, adaptive mechanisms must exist for oxygen protection. Vesicles are normally initiated only when the nitrogen source is limited. The developmental structure of the vesicles is specifically designed for physiological compartmentalization to protect the oxygen-labile nitrogenase enzyme (Huss-Danell, 1997). As a result, each vesicle is surrounded by an external multilamellate lipid envelope which presumably functions as an oxygen barrier (Lamont *et al.*, 1988, Huss-Danell, 1997).

**Frankia Isolates Used in this Study**

*Frankia* isolates have broadly reflects the host plants from which they were derived. Since the availability of 16S rRNA sequences, the phylogeny and taxonomy of the genus *Frankia* has been deduced (Benson and Clawson, 2000 and Normand *et al.*, 1996). Analyses show that known strains form a cohesive lineage in the family *Frankiaceae* within the bacterial Class *Actinobacteria* (Normand *et al.*, 1996 and Lechevalier, 1994). Most of the isolates fall into one of four host-plant specificity groups: (1) strains that infect *Alnus* and *Myrica* species; (2) strains that infect *Casuarina* and *Myrica* species; (3) strains that infect *Elaeagnus* and *Myrica* species and; (4) strains that infect only *Elaeagnus* species (Benson and Silvester 1993). Recently, the phylogeny of the genus *Frankia* has been reconstructed by comparative sequence analysis of the 16S rRNA gene, the genes for nitrogen fixation (*nif* genes) and by other genes (Benson and Clawson, 2000). It was found
that strains generally fall into one of three major groups or clusters (Cluster 1, Cluster 2, and Cluster 3) each having different and sometime overlapping plant specificity, physiological properties and symbiotic interactions (An et al., 1985; Benson and Clawson, 2000; Dobritsa and Stupar, 1989; Fernandez et al., 1989; Normand et al., 1996). Cluster 1 Frankia strains form nodules on members in the order Fagales, including the Betulaceae, Myricaceae and Casuarinaceae (Benson and Clawson, 2000). Cluster 2 Frankia strains only infect members of the Coriariaceae, Datiscaeae, Rosaceae and Ceanothus of the Rhamnaceae. These strains have not been isolated in pure culture despite many attempts to do so and may therefore be obligate symbionts (Benson and Clawson, 2000). Cluster 3 strains form effective nodules on members of the Myricaceae, Rhamnaceae, Elaeagnaceae and Gymnostoma of the Casuarinaceae (Benson and Clawson, 2000). The isolates, Ccl3, EAN1pec and Eu11c, used in this study represent two of the three clusters. Strain Ccl3 belongs to Cluster 1, Eu11c and EAN1pec belong to Cluster 3, and thus provide for comparative study among the two groups. In addition, the physiology of these strains have been intensely studied for vesicle functionality and sporulation (Krumholz et al., 2003, Tisa et al., 1987, Tjepkema and Murry, 1989).

Strains EAN1pec and Eu11c have also been used in our genetic studies for transformation using our transposon system using the transposon Tn916 (Myers et al., 2002, Rawnsley and Tisa, 2003). These strains have antibiotic resistance markers that provide a selectable trait. Previous studies have shown that strain EAN1pec is resistant to kasugamycin, novobiocin and lincomycin, while strain Eu11c
is resistant to novobiocin and strain Ccl3 is resistant to neomycin, kasugamycin, and gentamycin (Tisa et al., 2003).

These strains are also infective and will nodulate their host plant. Strain Ccl3 will form an effective nodule structure that is capable of nitrogen fixation; however, it is much more restrictive in its host range than either EAN1pec or Eu11c. Strains EAN1pec and Eu11c have a broad host range capability. However, strain EAN1pec will form an effective nodule structure that is capable of nitrogen fixation while strain Eu11c produces an ineffective nodule, meaning Eu11c is present inside of the nodule but is unable to fix nitrogen and unable to produce vesicles.

**Overview of Frankia Genetics**

The lack of genetic knowledge of the bacterium Frankia has hindered the study of this important symbiosis. The molecular genetics is largely unexplored due to slow growth rate (30-80h doubling time) leading to low biomass yields resulting in difficulty isolating high quality genomic DNA. Most Frankia strains are grown and maintained in liquid culture and generally grow slower on solid media. Colonies from spores or mycelial fragments become visible to the unaided eye about 7 to 10 days, or longer, after plating under the best of conditions. Genetic techniques used on other actinomycetes have had limited success with Frankia due to a variety of reasons, including low growth rates, poor spore germination, and lack of useful vectors (Benson and Silvester, 1993). Only a few Frankia genes have been cloned and sequenced and very little is known about gene expression (for review see Lavire and Cournoyer, 2003). Crucial protocols for gene transfer and mutagenesis have
not been firmly established and would provide an important stepping-stone for genetic analysis.

The few genes that have been sequenced can be divided according to their role into several key activities: gene translation (\(rrn\) and \(tRNA^{pro}\) gene), proteosomes (\(pcr\) genes), assimilation of ammonium (\(glnA\) and \(glnI\)), protection against superoxide ions (\(sodF\)), nitrogen fixation (\(nif\) genes), and plasmid replication genes (Lavire and Cournoyer, 2003).

**Overview of *Frankia* Genomics**

Reassociation kinetic analysis estimated the genome sizes of *Frankia* strains Arl4 and EuMc as 12 and 8.7 Mb, respectively (An *et al.*, 1985). These estimations are similar to those determined through restriction mapping and pulse field gel electrophoresis analysis of several *Streptomyces* spp. However, actinomycetes represent a wide range of morphological diverse species ranging from single rods to complex mycelial forms with diversity in the genome size and structure. Preliminary studies have shown that nearly all filamentous strains such as *Nocardia asteroids*, *Actinoplanes philippensis* and *Streptomyces* spp. seem to possess a chromosome with a linear topology, whereas the non-mycelium forming species, such as *Rhodococcus opacus*, or *Mycobacterium* and *Corynebacterium* spp. possess a circular chromosome that is much smaller in comparison to the filamentous actinomycete (Redenbach *et al.*, 2000). However, additional analyses with more representative strains need to be carried out to verify this preliminary correlation. A elucidation of the size, structure and organization of the genome of the actinomycete *Frankia*, would enable a better understanding of this important bacterium and its
interaction with its host plants (Lavire and Cournoyer, 2003). Lastly, determining the size and structure of the genome would provide the preliminary information needed to help facilitate the sequencing of the entire genome.

**Research Goals**

My approach is focused on developing molecular tools to facilitate easier genetic manipulation of this bacterium, along with generating new knowledge about the genome. This will be achieved through a number of approaches. First, the development of a physical and genetic map of the *Frankia* chromosome will provide tremendous insight on the organization and structure of this complex genome and provide the tools needed to facilitate the sequencing of the genome. The sequencing of two of the *Frankia* chromosome has provided an enormous amount of information to elucidate the complex genetic system of this bacterium. Second, confirming the feasibility of a transposon mutagenesis protocol and mapping the location of the mutations to the physical map is a vital protocol for future genetic analysis. The generation of these vital genetic tools presents an opportunity to further investigate all aspects of this symbiosis including infectivity, vesicle development, sporulation, and nitrogen fixation, as well as offering a means for comparative studies among other nitrogen fixing plant symbionts.
CHAPTER II

DEVELOPMENT OF A PHYSICAL MAP FOR THREE \textit{FRANKIA} STRAINS AND A PARTIAL GENETIC MAP FOR \textit{FRANKIA} EuI1C

\textbf{Introduction}

The data provided in this published paper was supported in part by Hatch grant 377, by USDA/NRICP grant 2003-0127 and by the College of Life Sciences and Agriculture, University of New Hampshire-Durham. The experiments and data shown in this journal article was provided entirely by my research for my dissertation and performed at the University of New Hampshire. Permission was granted by Physiologia Plantarum to allow for this reprint within this dissertation.
Development of a physical map for three *Frankia* strains and a partial genetic map for *Frankia* Eul1c

Tania Rawnsley and Louis S. Tisa*

Department of Microbiology, University of New Hampshire, Durham, NH 03824-2617, USA

**Introduction**

*Frankia* are nitrogen-fixing actinobacteria (gram-positive filamentous bacteria) that form a symbiotic association with over 200 different species of plants belonging to eight different plant families, which are only distantly related to each other (for a review see Benson and Silvester 1993, Huss-Daniel 1997, Mullin and Dobritsa 1996, Schwencke and Caru 2001, Wall 2000). These actinorhizal plants are able to grow on poorly fertilized soil and are often the pioneer species in plant community development. As a result, they are of potential significance with respect to land reclamation, reforestation, soil stabilization and fuel (Benson and Silvester 1993, Richards et al. 2002, Schwencke and Caru 2001).

Symbiotic interactions between *Frankia* and the host plant are not well understood. Outside of morphological descriptions, virtually nothing is known about the infection process or subsequent development of the actinorhizal symbiosis. Besides the ability to form a symbiotic association with plants, *Frankia* are developmentally complex and produce three cell types: hyphae, sporangia and vesicles (Benson and Silvester 1993). Spores are formed in sporangia at the mycelial tips and presumed to aid in the survival and dispersal of *Frankia*. Vesicles act as specialized structures for nitrogen fixation and are formed inside plant nodules or in culture under limiting conditions. Their shape is strain dependent and host-plant influenced. These pleiomorphic abilities suggest a complex regulation of cell development responding to environmental stimuli including interactions with their host plants. Although certain physiological and developmental requirements need to be satisfied for a successful mutualism, the interactions between *Frankia* and its host plant differ from those of other root-based nitrogen-fixing associations, such as rhizobia and legumes (Vessey et al. 2005). For example, the equivalents of rhizobial Nod factors have not yet been identified for *Frankia*. As a result of these unknown interactions it is necessary to gain more information not only about the physiology and biochemistry of this...
organism but also about the genetics and genomics of this symbiont.

Genetic analysis of Frankia has been limited (Lavire and Courmoyer 2003, Mullin and An 1990, Normand and Lalonde 1986, Simonet et al. 1990). Most efforts have been restricted to gene cloning via hybridization to genes from other organisms, to phylogenetic analysis based on selected gene sequences and to plasmid characterization. Genetic tools are only beginning to be developed (Caru and Cabello 2000). Because very little is known about the size, structure and organization of the genome, we were interested in developing a physical map of the Frankia chromosome by the use of macrorestriction analysis and pulsed-field gel electrophoresis (PFGE). Macrorestriction analysis and PFGE have been used successfully for the physical mapping of many bacterial genomes (Dudez et al. 2002, Pandza et al. 1997, Pisabarro et al. 1998, Ramos-Diaz and Ramos 1998, Sun et al. 2003). Here, we report the physical mapping of Frankia genome for three strains (Eul1c, EAN1pec and Ccl3). Two of these strains (Eul1c and EAN1pec) represent one of the three major lineages of Frankia, while strain Ccl3 represents another lineage. The host compatibilities of these strains differ greatly. Strain Ccl3 is restricted to Casuarina and Allocasuarina sp.; the host range of the other two strains is more diverse. During the middle of this study, genome sequencing for two of these strains (EAN1pec and Ccl3) was initiated and has been completed for strain Ccl3.

Materials and methods

Growth conditions
Frankia strains Ccl3 (Zhang et al. 1984), EAN1pec (Lalonde et al. 1981) and Eul1c (Baker et al. 1980) were grown and maintained in basal medium with NH₄Cl as the nitrogen source, as described previously (Tisa et al. 1999). For strain Eul1c, 20 mM glucose was used as carbon and energy source, while 20 mM fructose was used for strain EAN1pec. For Ccl3, 5 mM propionate was used. Unless stated otherwise, standing cultures of strains EAN1pec, Ccl3 and Eul1c were incubated at 25°C, 28°C and 30°C, respectively. Streptomyces coelicolor NRRL B-16638 was grown in Yeast Extract Malt Extract (YEME) medium (Hopwood et al. 1985).

Cultures grown for DNA preparation were grown in media supplemented with 0.1% glycine. This pretreatment with glycine increases the susceptibility of the hyphae to lysozyme and increases the yield of genomic DNA (gDNA) extracted (Okanishi et al. 1974, Tisa and Ensign 1987).

Preparation of gDNA in agarose plugs

gDNA for PFGE was prepared according to a modified method of Kieser et al. (2000). Cultures were grown for 7 days in growth medium supplemented with 0.1% glycine. Cells were harvested by centrifugation at 12 000 g for 15 min and rinsed in Hepes-EDTA (HE) buffer (10 mM HEPES, 1 mM EDTA, pH 8.0). The cells were finally resuspended in HE buffer to give an optical density at 600 nm of 1.9-2.1, and 625 μl of the suspension was mixed with 375 μl of molten 2.0% Seakem® LE (FMC, Rockland, ME) agarose to a final concentration of 0.75%. The molten suspension was poured into plastic molds by using a pipette tip with the end cut off and allowed to solidify. The agarose blocks were incubated at 37°C for 2 h in a lysozyme solution (2 mg ml⁻¹ lysozyme in HE buffer). The agarose plugs were removed and suspended in NDS (1% SDS, 1 mg ml⁻¹ proteinase K, 0.5 M EDTA, pH 8.0) and incubated at 50°C for 24 h. Proteinase K activity was inhibited by treatment with 1 mM phenylmethylsulfonyl fluoride (PMSF). The proteinase K-treated plugs were incubated twice in HE buffer containing 1 mM PMSF for 2 h each treatment. After the PMSF treatment, the plugs were incubated 2 x in HE buffer for 2 h each wash and stored at 4°C in HE buffer until needed.

Restriction endonuclease treatment

The above treated DNA-agarose plugs were cut in thirds along the width. Prior to enzymatic treatment, the slices of the DNA-agarose plugs were incubated twice in the 1 x appropriate restriction enzyme core buffer recommended by the supplier for at least 2 h each incubation at 4°C (all restriction enzymes were purchased from New England BioLabs, Ipswich, MA). After this equilibrium treatment, the DNA-agarose slices were transferred to
a tube containing 200 μl of fresh 1× core buffer and 50 U of restriction enzyme and incubated for 24 h at the optimum temperature for the restriction enzyme, as recommended by supplier. Restriction digests were terminated by heat inactivation.

For the double digest experiments, the DNA–agarose plug was treated with the first enzyme. After 24 h incubation, the DNA–agarose plug was removed and washed in 0.5 M EDTA, pH 8.0. The restriction enzyme-treated DNA–agarose plug was removed and suspended in the second 1× core buffer and incubated twice for at least 2 h at 4°C for each time period. After this equilibrium treatment, the restriction enzyme-treated DNA–agarose plugs were transferred to a tube containing 200 μl of fresh 1× core buffer and 50 U of the second restriction enzyme and incubated for 24 h at optimal temperature. Restriction digests were terminated by heat inactivation.

The restriction enzyme-treated DNA–agarose plugs were stored briefly at 4°C in PFGE buffer (0.5× TBE [50 mM Tris-HCl, 45 mM boric acid, 2.5 mM EDTA, pH 8.2 for 10× Tris-Borate-EDTA (TBE) buffer]) and used as soon as possible for PFGE.

**PFGE analysis**

BioRad CHEF-DR II system was used for all PFGE analysis. All agarose gels were prepared in 0.5× TBE buffer (2.5 mM Tris-HCl, 2.25 mM boric acid, 0.125 mM EDTA, pH 8.2) supplemented with 100 μM thiourea with varying concentrations of Seaplaque GTG agarose (Cambrex, East Rutherford, NJ). DNA samples were separated by the use of contour-clamped homogenous electric field with 0.5× TBE supplemented with 100 μM thiourea as the running buffer. Pulse times and gel concentrations were optimized with respect to the length of DNA to be separated. Yeast chromosome PFGE, lambda PFGE, mid-range PFGE, low-range PFGE markers (New England BioLabs) and Saccharomyces pombe yeast chromosome PFGE markers (BioRad, Richmond, CA) were used as linear PFGE DNA standards.

The PFGE results were analyzed through the use of QUANTITY ONE software (BioRad, Richmond, CA) and by comparing the fragments with the DNA standards.

**Polymerase chain reaction**

The oligonucleotide polymerase chain reaction (PCR) primers used in this study were purchased from Integrated DNA Technologies Inc. and are listed in Table 1. PCR amplifications were performed using HotStart Taq DNA Polymerase (Qiagen, Valencia, CA) following the manufacturer's instruction: 2.5 units of HotStart Taq DNA Polymerase, 1× PCR buffer, 200 μM of each dNTP, 0.3 μM of each primer and <1 ng of DNA template per reaction. PCR conditions were as follows: 15 min at 95°C for initial activation step, 1 min at 94°C for denaturation, 1 min at 50-68°C (5°C below Tm of primers) for annealing, 1 min at 72°C for extension for 30 cycles and 10 min at 72°C for a final extension.

**Topology experiments**

The presence of proteins covalently bound to DNA extremities was investigated to determine the topology of the chromosomal DNA. This was achieved by preparing the PFGE DNA with two different procedures. Purified DNA was prepared according to the procedure in Kieser et al. (2000) including the proteinase K treatment. A second of DNA samples were prepared without the proteinase step as described previously (Leblond et al. 1996, Lin et al. 1993). Instead of the proteinase treatment, an SDS step (2% SDS incubated at 50°C for 48 h) was used to remove non-covalently bound proteins from the DNA. Proteinase-treated and -untreated DNA samples were subjected to electrophoresis after digestion with AseI.

A second method was used to attempt to visualize the terminal protein-bound DNA fragments that have been shown to exist on linear chromosome, such as the linear chromosomes of the actinomycete Streptomyces species

<table>
<thead>
<tr>
<th>Target</th>
<th>Primer</th>
<th>Sequence (5'–3')</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>HboI</td>
<td>thhOr</td>
<td>GTCGACGGGGAAGGAGACCTTC</td>
<td>Neimann et al. (2005)</td>
</tr>
<tr>
<td></td>
<td>thhOr</td>
<td>CTGGCCGCAATCCGACCCCGG</td>
<td>Neimann et al. (2005)</td>
</tr>
<tr>
<td>GlnI</td>
<td>fGps19</td>
<td>TACATCTGGATTCACGGCAC</td>
<td>Hahn et al. (1999)</td>
</tr>
<tr>
<td></td>
<td>fGps4T7</td>
<td>GCCGACGCCGACAGTAGTA</td>
<td>Hahn et al. (1999)</td>
</tr>
<tr>
<td>GlnA</td>
<td>DB44</td>
<td>TTCCTCTCAGACGACCCG</td>
<td>Clawson et al. (2004)</td>
</tr>
<tr>
<td></td>
<td>DB54</td>
<td>GCCGCTCCGATGGAACTG</td>
<td>Clawson et al. (2004)</td>
</tr>
<tr>
<td>16S rRNA</td>
<td>RDB1</td>
<td>CCCAGCTGGAGGTTTATACACCAGGAA</td>
<td>Benson et al. (1996)</td>
</tr>
<tr>
<td></td>
<td>S18R</td>
<td>GNNATACCAGGGGCGCTGCTGAG</td>
<td>Lane et al. (1985)</td>
</tr>
</tbody>
</table>
tube. The DNA, which appeared as a white precipitate, was discarded. The beads were resuspended in 200 μl 1% SDS and incubated for 30 min at 50°C. After the suspension was centrifuged at 15,500 g for 1 min, the supernatant was removed and saved in a new tube. The above SDS step was repeated, and the supernatant fluids were pooled. The pooled supernatant fluids were centrifuged 2 x at 15,500 g for 1 min each to ensure total removal of the beads. After the supernatant was removed and placed in a new tube, 100 μl of 0.5 M EDTA, pH 8.0, and a final concentration of 300 μg ml⁻¹ of proteinase K were added to the supernatant fluid. The suspension was incubated for 2 h at 37°C to remove the terminal proteins. Proteinase K was removed by standard phenol:chloroform extraction methods (Sambrook et al. 1989), and NaCl was added to a final concentration of 0.2 M. The DNA was precipitated by adding an equal volume of isopropanol. The DNA pellet was dissolved in 10 μl of TE buffer and loaded onto a 0.8% agarose gel. After electrophoresis, the gel was stained with ethidium bromide.

Hybridization experiments

After PFGE, DNA fragments were transferred to nylon membranes (Hybond-N; Amersham, Piscataway, NJ) by capillary transfer as described previously (Sambrook et al. 1989) and fixed by heating at 80°C for 1 h. Probes for specific genes were prepared by PCR and were labeled by the use of [α-32P] deoxycytidine triphosphate (Amersham) and the Random primed DNA labeling kit (Amersham), according to the manufacturer's instructions. Hybridization with the labeled probes was carried out at 68°C with RapidHyb buffer (Amersham) under high stringency conditions, as specified by the manufacturer.

Results

Macrorestriction analysis and chromosome-size estimation

The genome size was determined by cleaving Frankia gDNA into relatively few fragments with restriction enzymes and separating the fragments by PFGE. The high G + C content (68–71%) of the Frankia DNA (An et al. 1983, Fernandez et al. 1989) required the use of restriction enzymes with recognition sites containing only A and T nucleotides, such as Asel (ATTAAAT), Spl (TAATTA), Swal (ATTAAAAAT) and Pmel (GTTTAAAAAC), to generate relatively few fragments that were suitable for the PFGE experiments. All fragments produced through digestion with these rare cutting restriction endonucleases required optimal adjustments of the pulse time programs. Pertinent examples are given in what follows.
For the sake of clarity, each Frankia strain will be described separately.

**Restriction analysis of Frankia Eul1c**

Frankia Eul1c DNA was digested with several different restriction enzymes, and all fragments were visualized by separating under the three different pulse conditions (Fig. 1). Digestion with Asel generated 14 fragments ranging in size from 20 to 2500 kb and SspI digestion produced 25 fragments ranging in size from 20 to 1100 kb (Table 2). Digestion with Swal yielded no visible bands, but digestion with Pmel formed seven fragments ranging in size from 40 to 5000 kb. Digestion with I-Ceul formed three fragments ranging in size from 700 to 5000 kb. The sums of the fragment sizes for Asel, SspI, Pmel and I-Ceul digestion were 8950, 7845, 7425, and 8200 kb, respectively, and gave an average of 8105 ± 845 kb for the genome size.

The intron-encoded restriction endonuclease, I-Ceul, recognizes a highly conserved 19-bp sequence within rrl genes for the large rRNA subunit (23S) gene (Liu et al. 1993) and has been used as an indicator for the number of rRNA (rrn) operons in a genome (Dudez et al. 2002, Kuwahara et al. 2002). Digestion of the Frankia Eul1c genome with I-Ceul produced three DNA fragments, which suggests the presence of three rRNA operons in Frankia Eul1c chromosome.

**Restriction analysis of Frankia Ccl3**

Digestion of Frankia Ccl3 gDNA with Asel produced 23 DNA fragments that ranged in size from 10 to 860 kb, while SspI digestion yielded 24 fragments, which ranged in size from 5 to 520 kb (Table 2). The separation of these fragments required three different pulse times 30 min, 25 s and 10 s as shown in Fig. 2. Digestion with SspI resulted in 11 singlet and 13 doublet (or multiple) bands (Fig. 2). The doublet bands were determined by using software and through the visualization of a thicker band and were confirmed by double digestion. Digestion with Swal yielded no visible bands under the three optimized pulsed-field times used in Fig. 2. One faint band was visualized on a pulse time of 30 min with a voltage of 2 V cm⁻¹ and a run time of 72 h (Fig. 3, panel A). Because the uncut chromosome was unable to migrate into the gel matrix and the Swal-digested chromosome was able to migrate into the gel matrix, this indicates the chromosome is circular in topology. The sums of the fragment sizes for Asel and SspI digests were 5395 kb and 5465 kb, respectively (Table 2) and gave an average of 5430 ± 35 kb for the genome size.

**Restriction analysis of Frankia EAN1pc**

The digestion of Frankia EAN1pc DNA with Asel produced 33 fragments that ranged in size from 5 to 750 kb, and digestion with Swal resulted in the formation of three large fragments that ranged from 1000 to 3000 kb. SspI digestion yielded 23 fragments ranging in size from 5 to 2750 kb (Table 2). All fragments were separated by using the three different pulse times as mentioned above. Experiments with Frankia EAN1pc were more problematic than the other two strains, because the digested DNA occasionally produced smearing patterns during PFGE. The sums of the fragment sizes for Asel and Swal digestion were 9005, 9210 and 9000 kb, respectively, and resulted in an average of 9101 ± 109 kb for the genome size.

**Determination of the Frankia chromosome topology**

Because many members of the actinomycetes are known to possess linear chromosomes (Redenbach et al. 2000), we were interested in determining the topology of the Frankia chromosome. Two strategies were employed to reveal information about the topology of the chromosome. The first approach takes advantage of the mobility differences for DNA topologies observed with PFGE (Levene and Zimm 1987). Large circular DNA molecules are unable to enter the agarose gels during PFGE and remain within the well of the gel. However, large linear
Table 2. Sizes of individual restriction fragments for Frankia gDNA digested with different restriction endonucleases. Sizes of fragments were estimated on the basis of at least two of the different PFGE conditions used to resolve the fragments. The letter 'd' indicates multiple bands.

<table>
<thead>
<tr>
<th>Strain</th>
<th>Cci3 (kb)</th>
<th>SspI (kb)</th>
<th>Pmel (kb)</th>
<th>AseI (kb)</th>
<th>SspI (kb)</th>
<th>Asel (kb)</th>
<th>SspI (kb)</th>
<th>Pmel (kb)</th>
<th>I-Ceul (kb)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>860</td>
<td>520</td>
<td>750</td>
<td>3000 (d)</td>
<td>2750</td>
<td>2500</td>
<td>1100</td>
<td>5000</td>
<td>5000</td>
</tr>
<tr>
<td>B</td>
<td>700</td>
<td>490</td>
<td>660</td>
<td>2000</td>
<td>1860</td>
<td>2000</td>
<td>795</td>
<td>1600</td>
<td>2500</td>
</tr>
<tr>
<td>C</td>
<td>430 (d)</td>
<td>400</td>
<td>650</td>
<td>1000</td>
<td>750</td>
<td>1000</td>
<td>645</td>
<td>480</td>
<td>700</td>
</tr>
<tr>
<td>D</td>
<td>400</td>
<td>320</td>
<td>600</td>
<td>2000</td>
<td>1860</td>
<td>2000</td>
<td>795</td>
<td>1600</td>
<td>2500</td>
</tr>
<tr>
<td>E</td>
<td>340</td>
<td>290</td>
<td>420</td>
<td>1000</td>
<td>900</td>
<td>1000</td>
<td>645</td>
<td>480</td>
<td>700</td>
</tr>
<tr>
<td>F</td>
<td>280</td>
<td>270</td>
<td>400</td>
<td>1000</td>
<td>900</td>
<td>1000</td>
<td>645</td>
<td>480</td>
<td>700</td>
</tr>
<tr>
<td>G</td>
<td>250</td>
<td>250</td>
<td>360</td>
<td>1000</td>
<td>900</td>
<td>1000</td>
<td>645</td>
<td>480</td>
<td>700</td>
</tr>
<tr>
<td>H</td>
<td>240</td>
<td>235</td>
<td>300</td>
<td>1000</td>
<td>900</td>
<td>1000</td>
<td>645</td>
<td>480</td>
<td>700</td>
</tr>
<tr>
<td>I</td>
<td>210</td>
<td>215 (d)</td>
<td>280</td>
<td>1000</td>
<td>900</td>
<td>1000</td>
<td>645</td>
<td>480</td>
<td>700</td>
</tr>
<tr>
<td>J</td>
<td>185</td>
<td>190</td>
<td>260</td>
<td>1000</td>
<td>900</td>
<td>1000</td>
<td>645</td>
<td>480</td>
<td>700</td>
</tr>
<tr>
<td>K</td>
<td>155</td>
<td>180</td>
<td>250</td>
<td>1000</td>
<td>900</td>
<td>1000</td>
<td>645</td>
<td>480</td>
<td>700</td>
</tr>
<tr>
<td>L</td>
<td>140</td>
<td>130</td>
<td>240 (d)</td>
<td>1000</td>
<td>900</td>
<td>1000</td>
<td>645</td>
<td>480</td>
<td>700</td>
</tr>
<tr>
<td>M</td>
<td>130</td>
<td>120 (d)</td>
<td>215 (d)</td>
<td>1000</td>
<td>900</td>
<td>1000</td>
<td>645</td>
<td>480</td>
<td>700</td>
</tr>
<tr>
<td>N</td>
<td>115</td>
<td>100 (d)</td>
<td>200 (d)</td>
<td>1000</td>
<td>900</td>
<td>1000</td>
<td>645</td>
<td>480</td>
<td>700</td>
</tr>
<tr>
<td>O</td>
<td>100 (d)</td>
<td>90 (d)</td>
<td>180</td>
<td>1000</td>
<td>900</td>
<td>1000</td>
<td>645</td>
<td>480</td>
<td>700</td>
</tr>
<tr>
<td>P</td>
<td>75 (d)</td>
<td>85 (d)</td>
<td>160</td>
<td>1000</td>
<td>900</td>
<td>1000</td>
<td>645</td>
<td>480</td>
<td>700</td>
</tr>
<tr>
<td>Q</td>
<td>60 (d)</td>
<td>65 (d)</td>
<td>150</td>
<td>1000</td>
<td>900</td>
<td>1000</td>
<td>645</td>
<td>480</td>
<td>700</td>
</tr>
<tr>
<td>R</td>
<td>55 (d)</td>
<td>55 (d)</td>
<td>140</td>
<td>1000</td>
<td>900</td>
<td>1000</td>
<td>645</td>
<td>480</td>
<td>700</td>
</tr>
<tr>
<td>S</td>
<td>40 (d)</td>
<td>45 (d)</td>
<td>130 (d)</td>
<td>1000</td>
<td>900</td>
<td>1000</td>
<td>645</td>
<td>480</td>
<td>700</td>
</tr>
<tr>
<td>T</td>
<td>30 (d)</td>
<td>40 (d)</td>
<td>100 (d)</td>
<td>1000</td>
<td>900</td>
<td>1000</td>
<td>645</td>
<td>480</td>
<td>700</td>
</tr>
<tr>
<td>U</td>
<td>25 (d)</td>
<td>30 (d)</td>
<td>90</td>
<td>1000</td>
<td>900</td>
<td>1000</td>
<td>645</td>
<td>480</td>
<td>700</td>
</tr>
<tr>
<td>V</td>
<td>20 (d)</td>
<td>20 (d)</td>
<td>85</td>
<td>1000</td>
<td>900</td>
<td>1000</td>
<td>645</td>
<td>480</td>
<td>700</td>
</tr>
<tr>
<td>W</td>
<td>10 (d)</td>
<td>75</td>
<td>5 (d)</td>
<td>1000</td>
<td>900</td>
<td>1000</td>
<td>645</td>
<td>480</td>
<td>700</td>
</tr>
<tr>
<td>X</td>
<td>5 (d)</td>
<td>70 (d)</td>
<td>25</td>
<td>1000</td>
<td>900</td>
<td>1000</td>
<td>645</td>
<td>480</td>
<td>700</td>
</tr>
<tr>
<td>Y</td>
<td>65</td>
<td></td>
<td>20</td>
<td>1000</td>
<td>900</td>
<td>1000</td>
<td>645</td>
<td>480</td>
<td>700</td>
</tr>
<tr>
<td>Z</td>
<td></td>
<td>60 (d)</td>
<td>20</td>
<td>1000</td>
<td>900</td>
<td>1000</td>
<td>645</td>
<td>480</td>
<td>700</td>
</tr>
<tr>
<td>AA</td>
<td>55 (d)</td>
<td></td>
<td>20</td>
<td>1000</td>
<td>900</td>
<td>1000</td>
<td>645</td>
<td>480</td>
<td>700</td>
</tr>
<tr>
<td>BB</td>
<td>50 (d)</td>
<td></td>
<td>20</td>
<td>1000</td>
<td>900</td>
<td>1000</td>
<td>645</td>
<td>480</td>
<td>700</td>
</tr>
<tr>
<td>CC</td>
<td>40 (d)</td>
<td></td>
<td></td>
<td>1000</td>
<td>900</td>
<td>1000</td>
<td>645</td>
<td>480</td>
<td>700</td>
</tr>
<tr>
<td>DD</td>
<td>30 (d)</td>
<td></td>
<td></td>
<td>1000</td>
<td>900</td>
<td>1000</td>
<td>645</td>
<td>480</td>
<td>700</td>
</tr>
<tr>
<td>EE</td>
<td>20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FF</td>
<td>10 (d)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GG</td>
<td>5 (d)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

DNA molecules are able to migrate into the PFGE gels. Linear chromosomes of several Streptomyces strains and several other actinomycetes were identified by this technique (Lin et al. 1993, Redenbach et al. 1996, Pitarro et al. 1998, Reeves et al. 1998, Redenbach et al. 2000). Undigested Frankia gDNA did not migrate into the gel. In Fig. 3, panels A and B show the results for gDNA from Frankia strains Cci3 and EAN1pec. The undigested gDNA did not migrate into the gel matrix, but digestion with Swal allowed the gDNA to enter the gel matrix. In Fig. 3, panel C shows similar results for Frankia Eullc. The undigested gDNA did not migrate into the gel matrix, while gDNA digested with I-Ceul and Pmel were able to enter into the agarose gel. These results suggest that all three Frankia strains have circular chromosomes.

Because many linear chromosomes and plasmids contain terminal proteins covalently bound to the 5' ends of the DNA, protease treatment will affect gel migration of the chromosome ends (Leblond et al. 1996). A comparison of the restriction pattern for both proteinase-treated and -untreated gDNA digested with AseI was determined and yielded identical results (Fig. 4), indicating the absence of terminal proteins and thus suggesting a circular topology. The ends of linear chromosome can be visualized by isolating the terminal protein-bound restriction fragments (Stoll and Cullum 2000). The absence of terminal restriction fragments was also used to confirm the above findings.

S. coelicolor, which possesses a linear chromosome and carries the linear plasmid SCP1 and SLP2, was used as
Fig. 2. PFGE analysis of Frankia Cc3 digested with restriction enzymes. Seaplaque agarose gels (0.8 or 1.0%), loaded with single plugs of strain Cc3 DNA prepared and digested as described in Materials and methods, were run in 0.5x TBE buffer at the conditions indicated to separate different size range fragments. Panel A, high-range PFGE to separate large-size fragments used 0.8% Seaplaque agarose electrophoresed for 24 h at 6.0 V cm−1 with a switch time of 60–110 s. Lane 1, digestion with Asel; lane 2, digestion with SspI; lane 3, yeast chromosome PFGE markers. Panel B, mid-range PFGE with 1.0% Seaplaque agarose run for 24 h at 6.0 V cm−1 with a switch time of 1–25 s. Lane 1, the mid-range PFGE markers; lane 2, digestion with Asel; lane 3, digestion with Asel; lane 4, digestion with SspI. Panel C, low-range PFGE with 1.0% Seaplaque agarose run for 15 h at 6.0 V cm−1 with a switch time of 1–10 s. Lane 1, low-range PFGE markers; lane 2, digestion with Asel; lane 3, digestion with SspI. Arrows indicate the size of the molecular markers.

A positive control to visualize the end proteins (Stoll and Cullum 2000). End bands of S. coelicolor were observed that, however, did not migrate into the gel matrix. This was probably because of the restriction enzyme that was used resulting in end fragments that were too large to enter the gel matrix (data not shown). Visible end bands were not generated for the Frankia strains after using restriction enzymes SphI and HindIII. These results combined with the PFGE analysis confirmed the circular topology of the Frankia chromosome.

Fig. 3. PFGE analysis of digested and undigested Frankia DNA. An 0.8% Seaplaque agarose gel was loaded with plugs of undigested and digested gDNA and ran at 2.0 V cm−1 for 72 h with a switch time of 30 min. Panel A, Frankia Cc3 gDNA; lane 1, S. pombe chromosome markers; lane 2, undigested Cc3 gDNA; lane 3, digestion with SstI; lane 4, digestion with HincII. Panel B, Frankia EAN1pec gDNA; lane 1, S. pombe chromosome markers; lane 2, undigested EAN1pec gDNA; lane 3, digestion with SstI; lane 4, digestion with HincII. Panel C, Frankia Eullc gDNA; lane 1, S. pombe chromosome markers; lane 2, undigested Eullc gDNA; lane 3, digestion with SstI; lane 4, digestion with HincII.

Construction of the physical map for the three strains

Both physical methods and hybridization analysis were used in the construction of a physical map for Frankia strains Cc3, EAN1pec and Eullc. The data obtained from single and double restriction endonuclease digestions of whole gDNA were combined with hybridization experiments to provide information about the linkage of the fragments.

Physical maps for Frankia Cc3 and EAN1pec are shown in Fig. 5. Double digestions of total gDNA of strain Cc3 was performed with combinations of Asel and SstI (Table 3). The single digest fragments A7, A16, A17, A18, A19, A21, A22, A23, A24, A25, A26 and A27 obtained by...
digestion of Asel were not digested by SspI (Table 3). Fragments A6, A11, A12 and A14 obtained by digestions with Asel were digested by SspI at one site and yielded two additional fragments each. Fragments A1, A2, A3, A4, A8, A9, A10, A13, A15 and A20 obtained by digestions with Asel were digested by SspI at multiple sites and yielded many additional fragments. The predicted restriction fragment pattern determined from an analysis of the recently completed genome sequence from strain Cc3 confirms our physical map (data not shown) and further validates a circular chromosome.

Double digestion of total gDNA of Frankia EAN1pec was performed with combinations of Swal and Asel (Table 3). It was determined that fragment A2 obtained by digestion with Asel was digested by Swal at one site and yielded two additional fragments AS1 and AS2 of 340 and 320 kb, respectively. An 8x draft coverage of Frankia EAN1pec has been recently determined (data not shown), and the predicted restriction fragments match the physical map shown in Fig. 5. These sequencing data predict the presence of four Swal restriction sites on the chromosome. For a linear topology, we would expect five fragments with Swal digestion and four fragments for a circular topology. The Swal profile obtained in this study reveals four fragments and thus confirms the circular topology of the chromosome.

Positioning of the known Frankia genes relative to the Eu1c physical map

The positions of three genes, glutamine synthetase, glnA and glnl, and a truncated hemoglobin gene, hboO, were located on the Frankia Eu1c physical map by Southern hybridization experiments, as described in Materials and methods. Probes for glnA and glnl hybridized to the same two fragments, fragments A4 and A5, when hybridized under high stringency conditions, showing that these two genes are clustered in gene organization and indicating that these two fragments are positioned next to each other on the physical map. The hboO probe hybridized to five fragments when double digested with Pmel and Asel, fragments AP7, AP8, AP9 and AP10 under high stringency condition showing the gene organization of these three fragments on the physical map. Two fragments with Asel and correlated with

---

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Table 3. Sizes of individual restriction fragments used for the determination of the physical maps of Frankia Cc!3 and EAN1 pec. gDNA was digested with AseI, SspI or SwaI and subjected to double digestion with the same restriction enzymes. The fragments in bold indicate the fragments that were seen in the single digests as well as the double digests with restriction enzyme. Sizes of fragments were estimated on the basis of at least two of the different PFGE conditions used to resolve the fragments as well as Quantity One software.

<table>
<thead>
<tr>
<th></th>
<th>AseI (kb)</th>
<th>SpI (kb)</th>
<th>AseI-SpI (kb)</th>
<th>SwaI (kb)</th>
<th>AseI-SwaI (kb)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>860</td>
<td>51</td>
<td>520</td>
<td>475</td>
<td>A1 750</td>
</tr>
<tr>
<td>A2</td>
<td>700</td>
<td>52</td>
<td>490</td>
<td>400</td>
<td>A2 660</td>
</tr>
<tr>
<td>A3/A4</td>
<td>430</td>
<td>53</td>
<td>440</td>
<td>290</td>
<td>A7 745</td>
</tr>
<tr>
<td>A5</td>
<td>460</td>
<td>54</td>
<td>430</td>
<td>290</td>
<td>A4 600</td>
</tr>
<tr>
<td>A6</td>
<td>340</td>
<td>55</td>
<td>290</td>
<td>270</td>
<td>A6 420</td>
</tr>
<tr>
<td>A7</td>
<td>290</td>
<td>56</td>
<td>270</td>
<td>215</td>
<td>A6 400</td>
</tr>
<tr>
<td>A8</td>
<td>280</td>
<td>57</td>
<td>225</td>
<td>180</td>
<td>A7 380</td>
</tr>
<tr>
<td>A9</td>
<td>240</td>
<td>58</td>
<td>235</td>
<td>150</td>
<td>A8 300</td>
</tr>
<tr>
<td>A10</td>
<td>210</td>
<td>59/10</td>
<td>215</td>
<td>150</td>
<td>A9 280</td>
</tr>
<tr>
<td>A11</td>
<td>185</td>
<td>51/11</td>
<td>190</td>
<td>150</td>
<td>A10 260</td>
</tr>
<tr>
<td>A12</td>
<td>155</td>
<td>51/2</td>
<td>180</td>
<td>130</td>
<td>A9 240</td>
</tr>
<tr>
<td>A13</td>
<td>140</td>
<td>51/3</td>
<td>130</td>
<td>120</td>
<td>A11 250</td>
</tr>
<tr>
<td>A14</td>
<td>130</td>
<td>51/4/18</td>
<td>120</td>
<td>120</td>
<td>A12/13 240</td>
</tr>
<tr>
<td>A15</td>
<td>115</td>
<td>51/9/20</td>
<td>100</td>
<td>100</td>
<td>A14/15 215</td>
</tr>
<tr>
<td>A16</td>
<td>100</td>
<td>52/12/2</td>
<td>90</td>
<td>100</td>
<td>A16/18 200</td>
</tr>
<tr>
<td>A17</td>
<td>75</td>
<td>52/3/24</td>
<td>85</td>
<td>100</td>
<td>A16/18 180</td>
</tr>
<tr>
<td>A18/19</td>
<td>60</td>
<td>52/5/26</td>
<td>65</td>
<td>90</td>
<td>A21 150</td>
</tr>
<tr>
<td>A20</td>
<td>55</td>
<td>52/7/28</td>
<td>55</td>
<td>90</td>
<td>A21 140</td>
</tr>
<tr>
<td>A21</td>
<td>40</td>
<td>52/9/30</td>
<td>45</td>
<td>85</td>
<td>A22 130</td>
</tr>
<tr>
<td>A22</td>
<td>30</td>
<td>53/1/32</td>
<td>40</td>
<td>85</td>
<td>A23/26 130</td>
</tr>
<tr>
<td>A23/24</td>
<td>25</td>
<td>53/4/35</td>
<td>30</td>
<td>75</td>
<td>A27/28 100</td>
</tr>
<tr>
<td>A25/26</td>
<td>20</td>
<td>53/6/39</td>
<td>20</td>
<td>75</td>
<td>A29  90</td>
</tr>
<tr>
<td>A27</td>
<td>10</td>
<td>54/0/43</td>
<td>10</td>
<td>60</td>
<td>A30  85</td>
</tr>
<tr>
<td></td>
<td>54/4/47</td>
<td>5</td>
<td>22/7/28</td>
<td>55</td>
<td>A31  70</td>
</tr>
<tr>
<td>A510</td>
<td>50</td>
<td>53/3/30</td>
<td>45</td>
<td>55</td>
<td>A32  65</td>
</tr>
<tr>
<td>A511</td>
<td>45</td>
<td>53/4/39</td>
<td>40</td>
<td>55</td>
<td>A32  60</td>
</tr>
<tr>
<td>A512</td>
<td>45</td>
<td>53/5/47</td>
<td>40</td>
<td>55</td>
<td>A32  60</td>
</tr>
<tr>
<td>A513</td>
<td>45</td>
<td>53/6/7</td>
<td>40</td>
<td>55</td>
<td>A32  60</td>
</tr>
<tr>
<td>A514</td>
<td>45</td>
<td>53/7/39</td>
<td>40</td>
<td>55</td>
<td>A32  60</td>
</tr>
<tr>
<td>A515</td>
<td>45</td>
<td>53/8/39</td>
<td>40</td>
<td>55</td>
<td>A32  60</td>
</tr>
<tr>
<td>A516</td>
<td>45</td>
<td>53/9/39</td>
<td>40</td>
<td>55</td>
<td>A32  60</td>
</tr>
<tr>
<td>A517</td>
<td>45</td>
<td>54/0/39</td>
<td>40</td>
<td>55</td>
<td>A32  60</td>
</tr>
<tr>
<td>A518</td>
<td>45</td>
<td>54/1/39</td>
<td>40</td>
<td>55</td>
<td>A32  60</td>
</tr>
<tr>
<td>A519</td>
<td>45</td>
<td>54/2/39</td>
<td>40</td>
<td>55</td>
<td>A32  60</td>
</tr>
<tr>
<td>A520</td>
<td>45</td>
<td>54/3/39</td>
<td>40</td>
<td>55</td>
<td>A32  60</td>
</tr>
<tr>
<td>A521</td>
<td>45</td>
<td>54/4/39</td>
<td>40</td>
<td>55</td>
<td>A32  60</td>
</tr>
<tr>
<td>A522</td>
<td>45</td>
<td>54/5/39</td>
<td>40</td>
<td>55</td>
<td>A32  60</td>
</tr>
<tr>
<td>A523</td>
<td>45</td>
<td>54/6/39</td>
<td>40</td>
<td>55</td>
<td>A32  60</td>
</tr>
<tr>
<td>A524</td>
<td>45</td>
<td>54/7/39</td>
<td>40</td>
<td>55</td>
<td>A32  60</td>
</tr>
<tr>
<td>A525</td>
<td>45</td>
<td>54/8/39</td>
<td>40</td>
<td>55</td>
<td>A32  60</td>
</tr>
<tr>
<td>A526</td>
<td>45</td>
<td>54/9/39</td>
<td>40</td>
<td>55</td>
<td>A32  60</td>
</tr>
<tr>
<td>A527</td>
<td>45</td>
<td>55/0/39</td>
<td>40</td>
<td>55</td>
<td>A32  60</td>
</tr>
<tr>
<td>A528</td>
<td>45</td>
<td>55/1/39</td>
<td>40</td>
<td>55</td>
<td>A32  60</td>
</tr>
<tr>
<td>A529</td>
<td>45</td>
<td>55/2/39</td>
<td>40</td>
<td>55</td>
<td>A32  60</td>
</tr>
<tr>
<td>A530</td>
<td>45</td>
<td>55/3/39</td>
<td>40</td>
<td>55</td>
<td>A32  60</td>
</tr>
<tr>
<td>A531</td>
<td>45</td>
<td>55/4/39</td>
<td>40</td>
<td>55</td>
<td>A32  60</td>
</tr>
<tr>
<td>A532</td>
<td>45</td>
<td>55/5/39</td>
<td>40</td>
<td>55</td>
<td>A32  60</td>
</tr>
<tr>
<td>A533</td>
<td>45</td>
<td>55/6/39</td>
<td>40</td>
<td>55</td>
<td>A32  60</td>
</tr>
<tr>
<td>A534</td>
<td>45</td>
<td>55/7/39</td>
<td>40</td>
<td>55</td>
<td>A32  60</td>
</tr>
<tr>
<td>A535</td>
<td>45</td>
<td>55/8/39</td>
<td>40</td>
<td>55</td>
<td>A32  60</td>
</tr>
<tr>
<td>A536</td>
<td>45</td>
<td>55/9/39</td>
<td>40</td>
<td>55</td>
<td>A32  60</td>
</tr>
</tbody>
</table>

The predicted three rRNA operons determined from digestion with restriction enzyme 1-Ceul.

Discussion

The macrorestriction analysis described here was executed to gain information about the genomic organization of Frankia strains Cc!3, EANTpec and Eu11c. We have estimated by PFGE analysis that the chromosome sizes of the Frankia strains Cc!3, EANTpec and Eu11c were 5.4, 9.1 and 8.1 Mb, respectively. These values are similar to those estimated previously by the less accurate technique of reassociation kinetics (An et al. 1985). However, the extraordinary size discrepancy among the three strains...
was unexpected and may be because of their host-plant specificities. Strain Cci3 had the smallest genome (5.4 Mb) and has a narrow host-range infecting members of the Casuarinaceae family (Zhang et al. 1984). The two larger genomes (9.1 and 8.1 Mb) were associated with the broad host-range *Frankia* strains, EAN1pec and Eui1c. These two Elaeagnus isolates differed in size by 1 Mb and may reflect a loss of genes involved in nodule function. Strain Eui1c infects and nodulates its host plant, but results in an ineffective nodule that is unable to fix N₂ (Baker et al. 1980). Comparative genomic analysis of these two strains may elucidate these important genes.

One surprising finding in our study was that the *Frankia* genome had a circular topology, unlike many other actinomycetes that have large genome sizes that have been shown to have linear topology, such as the closely related *Streptomyces* species (Lin et al. 1993, Leblond et al. 1996). Several lines of evidence support the circular topology of the *Frankia* chromosome. During PFGE, undigested *Frankia* DNA was unable to enter the agarose gels. Under the conditions tested, large linear chromosomes of several *Streptomyces* strains and several other actinomycetes will migrate into the gel (Lin et al. 1993, Pisbarro et al. 1998, Redenbach et al. 1996, Reeves et al. 1998), while circular chromosomes will remain within the well of the gel (Levene and Zimm 1987). The absence of terminal proteins covalently bound to the 5' ends of the chromosome, which was confirmed by two different approaches, further supports the circular topology model.

From this study, both I-CeuI digestion results and Southern hybridization experiments indicate that *Frankia* Eui1c chromosome harbors three *rrn* operons. The data in this study also suggest that *Frankia* EAN1pec and Cci3 possess three and two *rrn* operons, respectively. Preliminary sequencing data for these two strains confirm that prediction. Normand et al. (1992) have shown that *Frankia* Cci3, which is infective on *Casuarina* plants, has two *rrn* operons as does strain ACN14a, which is infective on alder plants. It is curious that the two Elaeagnus isolates harbor three *rrn* operons compared with the two *rrn* operons found with the alder and casuarina isolates. The additional *rrn* genes may be a reflection of the size differences among isolates or provide a selective advantage to the organism under various growth conditions. We favor the second hypothesis for several reasons. The genome sizes of strains Eui1c and ACN14a are similar, but only strain Eui1c contains additional *rrn* genes. Besides being broad host-range isolates, both strains Eui1c and EAN1pec have a more diverse metabolism than the medium host-range alder/casuarina isolates, suggesting a greater flexibility to changing environments.

During this study, one problem with the PFGE analysis was the intermittent occurrence of a band-smearing pattern that was more prevalent with strain EAN1pec. With some actinobacteria, including *Mycobacterium* and *Streptomyces* species, their DNA is susceptible to a Tris-dependent cleavage that occurs during gel electrophoresis (Evans et al. 1994, Ray et al. 1992, Zhang et al. 2004). The addition of thiourea to the Tris running buffer or replacement of Tris with HEPES buffer reduces DNA degradation. Because HEPES buffer has a higher ionic strength, the use of this buffer requires a longer run time to maintain the current within the normal range. To circumvent this problem, a HEPES-based buffer was used for plug preparation and treatment and 100 μM thiourea was added to the running buffer. However, we still frequently observed a smearing pattern with strain EAN1pec, and gels that were subjected to long-term electrophoreses (72 h) also showed a smearing pattern. With the long-term topology gels, the extensive running times may have an effect on the DNA degradation observed.

The results obtained in this study have successfully facilitated in the further genetic analysis of this important plant symbiont by providing valuable initial data to aid in genomic sequencing. During the middle of this study, the genome sequencing of two of these strains, Cci3 and EAN1pec, was initiated and has been completed for Cci3. The physical maps determined in our study were concurred by the predicted physical maps from the

---

Fig. 6. Physical and genetic map of *Frankia* Eui1c chromosome. The *Asel* and *Pmel* fragments are indicated by A and P, respectively, and correspond to the fragments indicated in Table 4. The positions of three genes are shown.
Table 4. Sizes of individual restriction fragments used for Frankia Eu1c physical map determination. gDNA digested with Asel, SspI and Pmel and double digested with Asel–Pmel. The fragments in bold indicate the fragments that were seen in the single digests as well as the double digests with Asel. Sizes of fragments were estimated under at least two of the different PFGE conditions used to resolve the fragments as well as with Quantity One software.

<table>
<thead>
<tr>
<th>Asel (kb)</th>
<th>SspI (kb)</th>
<th>Pmel (kb)</th>
<th>Asel–Pmel (kb)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fragment</td>
<td>Size (kb)</td>
<td>Fragment</td>
<td>Size (kb)</td>
</tr>
<tr>
<td>A1</td>
<td>2500</td>
<td>S1</td>
<td>1100</td>
</tr>
<tr>
<td>A2</td>
<td>2000</td>
<td>S2</td>
<td>795</td>
</tr>
<tr>
<td>A3</td>
<td>1000</td>
<td>S3</td>
<td>645</td>
</tr>
<tr>
<td>A4 (glnA)</td>
<td>900</td>
<td>S4</td>
<td>560</td>
</tr>
<tr>
<td>A5 (glnI)</td>
<td>640</td>
<td>S5</td>
<td>530</td>
</tr>
<tr>
<td>A6</td>
<td>550</td>
<td>S6</td>
<td>500</td>
</tr>
<tr>
<td>A7 (hboQ)</td>
<td>360</td>
<td>S7</td>
<td>440</td>
</tr>
<tr>
<td>A8</td>
<td>320</td>
<td>S8</td>
<td>360</td>
</tr>
<tr>
<td>A9</td>
<td>290</td>
<td>S9</td>
<td>330</td>
</tr>
<tr>
<td>A10</td>
<td>250</td>
<td>S10</td>
<td>270</td>
</tr>
<tr>
<td>A11</td>
<td>170</td>
<td>S11 (hboQ)</td>
<td>260</td>
</tr>
<tr>
<td>A12</td>
<td>100</td>
<td>S12</td>
<td>230</td>
</tr>
<tr>
<td>A13 hboQ</td>
<td>70</td>
<td>S13</td>
<td>210</td>
</tr>
<tr>
<td>A14</td>
<td>20</td>
<td>S14</td>
<td>190</td>
</tr>
<tr>
<td>A15</td>
<td>150</td>
<td>S15</td>
<td>150</td>
</tr>
<tr>
<td>A16</td>
<td>130</td>
<td>S16</td>
<td>115</td>
</tr>
<tr>
<td>A17</td>
<td>115</td>
<td>S17</td>
<td>100</td>
</tr>
<tr>
<td>A18</td>
<td>90</td>
<td>S19</td>
<td>90</td>
</tr>
<tr>
<td>A19</td>
<td>50</td>
<td>S20</td>
<td>50</td>
</tr>
<tr>
<td>A21</td>
<td>40</td>
<td>S21</td>
<td>35</td>
</tr>
<tr>
<td>A22</td>
<td>35</td>
<td>S22</td>
<td>30</td>
</tr>
<tr>
<td>A23</td>
<td>30</td>
<td>S23</td>
<td>25</td>
</tr>
<tr>
<td>A24</td>
<td>25</td>
<td>S24</td>
<td>20</td>
</tr>
<tr>
<td>A25</td>
<td>20</td>
<td>S25</td>
<td></td>
</tr>
</tbody>
</table>

Genome sequence data, and our data confirmed the predicted topology of Frankia chromosome. The establishment of physical and genetic map for Frankia Eu1c will help efforts to understand the genetic basis of nodule formation and function in the actinorhizal symbiosis.

Acknowledgements – This investigation was supported in part by Hatch grant 377, by USDA/NRIICP grant 2003-01127 and by the College of Life Sciences and Agriculture, University of New Hampshire-Durham. This is scientific contribution 2316 from the NH Agricultural Experiment Station. We thank Robert Mooney for his help with the photography.

References


Physical and genetic map of Lactobacillus sakei 23K chromosome. Microbiology 148: 421–431


Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
CHAPTER III

GENOME CHARACTERISTICS OF FACULTATIVELY SYMBIOTIC FRANKIA SP.
STRAINS REFLECT HOST RANGE AND HOST PLANT BIOGEOGRAPHY

Introduction

Through a collaborative project supported by the National Science Foundation Microbial Genome sequencing program, we were able to sequence and compare the genomes of three isolates, including the narrow host range Casuarina strain (Ccl3), a medium host range Alnus strain (ACN14a), and the broad host range Elaeagnus strain (EANIpec). The work on ACN14a was performed at Genoscope, Evry, France and was supported by CNRS/ACI Microbiologie and MRT/ACI IMPBio2004. The work on Ccl3 and EANIpec was performed under the auspices of the U.S. Department of Energy’s Office of Science, Biological, and Environmental Research Program. The genomes of Ccl3 and EANIpec were sequenced using the shotgun sequencing approach carried out at the Joint Genome Institute. However, due to the fastidious handling required for growing Frankia and the difficulty in obtaining high molecular weight genomic DNA, it was vital that I provide the high quality genomic DNA necessary for the completion of the sequencing for both Ccl3 and EANIpec. Also, the preliminary work made available through the research completed for my thesis project was necessary in allowing the funding of the sequencing project to be successful.
The comparisons made in this collaboration paper has not only allowed for us to raise some valid and important hypotheses regarding genome size discrepancies and *Frankia*’s symbiotic host range and host plant biogeography, but also enabled us to confirm the results obtained from my physical and genetic maps. Originally, my thesis project was to provide tools that would facilitate the sequencing of the entire *Frankia* genome and I am pleased to show through this published paper that due to the data provided by my original thesis work we were able to work together with many members of the *Frankia* community and sequence three important *Frankia* isolates. Permission was granted by Cold Spring Harbor Laboratory Press to reprint this article within this dissertation.
Genome characteristics of facultatively symbiotic Frankia sp. strains reflect host range and host plant biogeography

Philippe Normand,1 Pascal Lapierre,2 Louis S. Tisa,3 Johann Peter Gogarten,2 Nicole Alloisio,1 Emilie Bagnarol,1 Carla A. Bassi,2 Alison M. Berry,4 Derek M. Bickhart,2 Nathalie Choine,5,6 Arnaud Couloux,6 Benoit Cournoyer,1 Stephane Cruveiller,7 Vincent Daubin,8 Nadia Demange,6 Maria Pilar Francino,9 Eugene Goltsman,9 Ying Huang,2 Olga R. Kopp,10 Laurent Labarre,7 Alla Lapidus,9 Celine Lavire,1 Joelle Marechal,1 Michele Martinez,9 Juliana E. Mastrunzio,2 Beth C. Mullin,10 James Niemann,3 Pierre Pujic,1 Tania Rawnsley,3 Zoe Rouy,7 Chantal Schenowitz,6 Anita Sellstedt,1 Fernando Tavares,12 Jeffrey P. Tomkins,13 David Vallenet,7 Claudio Valverde,14 Luis G. Wall,14 Ying Wang,10 Claudine Medigue,7 and David R. Benson2,15

1Université de Lyon, Unité Mixte de Recherche, Centre National de la Recherche Scientifique (UMR CNRS), 5557 Ecologie Microbiennne, IFR41 Bio Environnement et Santé, Université Lyon I, Villeurbanne 69622 cedex, France; 2Department of Molecular and Cell Biology, University of Connecticut, Storrs, Connecticut 06279, USA; 3Department of Microbiology, University of New Hampshire, Durham, New Hampshire, 03824, USA; 4Department of Plant Sciences, University of California, Davis, California 95616, USA; 5Institut National de la Recherche Agronomique—Unité de Recherche en Génomique Végétale (INRA-URGV), 91057 Ervy cedex, France; 6Genoscope, Centre National de Séquençage, 91057 Ervy cedex, France; 7Genoscope-CNRS UMR 8030, Atelier de Génomique Comparatve, 91006 Ervy cedex, France; 8Bioinformatics and Evolutionary Genomics Laboratory, UMR CNRS 5558, Université Lyon I, Villeurbanne 69622 cedex, France; 9DOE Joint Genome Institute, Walnut Creek, California 94598, USA; 10Department of Biochemistry & Cellular & Molecular Biology and The Genome Science & Technology Program, The University of Tennessee, Knoxville, Tennessee 37796, USA; 11Department of Plant Physiology, Umeå University, 90187 Umeå, Sweden; 12Instituto de Biologia Molecular e Celular, Universidade do Porto, 4150-180 Porto, Portugal; 13Clemson University Genomics Institute, Clemson, South Carolina 29634, USA; 14Programa Interacciones Biológicas, Departamento de Ciencia y Tecnología, Universidad Nacional de Quilmes, Bernal 818768XU, Argentina

Soil bacteria that also form mutualistic symbioses in plants encounter two major levels of selection. One occurs during adaptation to and survival in soil, and the other occurs in concert with host plant speciation and adaptation. Actinobacteria from the genus Frankia are facultative symbionts that form N2-fixing root nodules on diverse and globally distributed angiosperms in the "actinorhizal" symbioses. Three closely related dades of Frankia sp. strains are recognized; members of each dade infect a subset of plants from among eight angiosperm families. We sequenced the genomes from three strains; their sizes varied from 5.43 Mbp for a narrow host range strain (Frankia sp. strain HFPCcl3) to 7.50 Mbp for a medium host range strain (Frankia alni strain ACN14a) to 9.04 Mbp for a broad host range strain (Frankia sp. strain EANIpec.) This size divergence is the largest yet reported for such closely related soil bacteria (97.8%-98.9% identity of 16S rRNA genes). The extent of gene deletion, duplication, and acquisition is in concert with the biogeographic history of the symbioses and host plant speciation. Host plant isolation favored genome contraction, whereas host plant diversification favoro genome expansion. The results support the idea that major genome expansions as well as reductions can occur in facultative symbiotic soil bacteria as they respond to new environments in the context of their symbioses.

[The genome sequences for Frankia strains Cd3, ACN4a, and EANIpec have been submitted to GenBank under accession nos. CP000249, CT573213, and AA100000000, respectively.]
have symbiotic genes (nod genes) subject to horizontal transfer among α- and some β-Proteobacteria (Chen et al. 1991; Young and Frankia 1996; Moulin et al. 2001). In contrast, all Frankia sp. strains are closely related with no evidence of dissemination of nodulating ability to related actinobacteria (Fig. 1; Normand et al. 1996; Clawson et al. 2004).

In plants, the capacity to form N₂-fixing root nodules occupied by bacteria is retained in a single lineage of angiosperms known as the "N₂-fixing clade" (Solits et al. 1995). Ten families within the Euroid I clade have members that are nodulated (Solits et al. 1995; Swensen 1996; Clawson et al. 2004). Only two of the families have members that associate with nodulating proteobacteria, while eight associate with Frankia sp. strains to form the actinorhizal symbiosis (Table 1).

Frankia strains fall into three closely related clusters. Members of each cluster have distinct host ranges (Table 1; Fig. 1). Cluster 1 strains nodulate plants in the Fabaceae in the Betulaceae and Myricaceae and are often referred to as "Alnus strains" (Normand et al. 1996). A subclade within Cluster 1 is comprised of the narrow host range "Casuarina strains" that under natural conditions nodulate only Casuarina and Allocasuarina species in the Casuarinaceae (Benson et al. 2004). Conversely, Cluster 3 "Elaeagnus strains" are considered to have a broad host range since they nodulate plants from five families in the Fabaceae and Rosales (Benson et al. 2004). Finally, the "Rosaceae strains" form Cluster 2, which is sister to the others; representatives of this cluster have not been isolated and grown in culture. Cluster 2 strains nodulate plants from four families in the Rosales and Cucurbitales (Benson et al. 2004; Vanden Hoovet al. 2004).

To gain insight into the evolutionary trajectory followed by these closely related, yet host-range and geographically divergent, Frankia sp. strains, we sequenced and compared the genomes of three isolates, including a narrow host range Casuarina strain, a medium host range Alnus strain, and a broad host range Elaeagnus strain. The results suggest that gene deletion and duplication have occurred to different extents in the genomes during adaptation to host plants and their environments. The concept of genome contraction echoes the changes known to occur in obligate bacterial pathogens and symbionts (Mira et al. 2001; Ochman and Moran 2001; Moran 2003), but the observation that both contraction and expansion can occur in closely related lineages of facultatively symbiotic soil bacteria in relation to host distribution has not previously been reported.

Results and Discussion

Actinorhizal plant families emerged in the late Cretaceous (~100 million years ago [Mya]) and subsequently adapted to a wide variety of environments (Magallon et al. 1999). Currently, they are globally distributed in climate zones ranging from alpine and subarctic to tropical (Fig. 2) where they add nitrogen and organic material to nutrient-poor soils (Silvester 1976). The native geographical distributions of hosts range from limited in the case of Casuarina sp. to broad in the case of Morella sp. (Fig. 2). The distribution of bacterial symbionts is obviously more difficult to assess, but numerous studies have shown some correlation with plant distribution (for review, see Benson et al. 2004).

Frankia sp. strain HPFC013 (Cc3) represents narrow host range Casuarina strains commonly detected in nodules collected from casuarinas in their native Australia (Fig. 2A) and in areas of the world where casuarina trees have been planted as windbreaks or for erosion control (Simone et al. 1999). Similar strains have not been found in soils in the absence of a suitable host, indicating that the bacteria depend on the plant for their soil propagation (Simone et al. 1999).

Frankia sp. strain ACN14a (ACN) represents Alnus strains that are globally distributed in soils regardless of the presence of a suitable host plant (Benson et al. 2004). This ubiquity parallels the distribution of host plants from the Betulaceae and Myricaceae that have a combined native range spanning all continents except Australia (Table 1; Fig. 2B).

Frankia sp. strain EAN1pc (EAN) represents broad host range Elaeagnus strains that are also globally distributed in soils with or without host plants (Benson et al. 2004). Cognate hosts are the most diverse and have the widest distribution with representatives on all continents including Australia (Table 1; Fig. 2C).

The strains used in this study have 16S rRNA gene sequences that are 97.8% identical between ACN or Cc3 versus EAN, and 98.9% identical between ACN and Cc3 (Fig. 1). This similarity level is frequently observed among bacteria from the same species (Wayne et al. 1987; Greaves et al. 2005), and is typical of the similarity levels found within the genus Frankia (Fig. 1; Clawson et al. 2004).

Genome characteristics

The genomes from ACN and Cc3 have been finished, and that from EAN has been rendered in a single scaffold with some gaps corresponding to regions that have proven difficult to resolve due to sequence repeats and high GC content (Table 2). Nevertheless, unlike Streptomyces (Bentley et al. 2002), all three genomes are circular as demonstrated directly from their sequences (Fig. 3). None of the strains have yielded independently replicating plasmids. Unlike what is observed for obligate symbionts, the coding capacity of the genomes remains quite high at 89% for ACN, 84% for Cc3, and 86% for EAN.

The most striking difference between the three genomes is their sizes, ranging from 5.43 Mb for Cc3 (4493 protein-coding sequences (CDS)), to 7.50 Mb for ACN (6786 CDS), to 9.04 Mb for

8 Genome Research www.genome.org
EAN (7976 CDS) (Table 2). On a total DNA basis, this range is the largest reported for any group of free-living prokaryotes related at the 98%-99% 16S rRNA sequence level described to date. The smallest genome belongs to the narrow host range and geographically limited representative Cdi, and the largest is from strain EAN, which belongs to the broadest host range group. This size correlation raises the hypothesis that genome size and content is driven by the host range and biogeography of the symbioses. We addressed this hypothesis through comparative analysis of the genomes’ contents and structures to determine how the disparate sizes have come about.

The genome maps shown in Figure 3 indicate that the patterns of syntenies are quite similar, with syntenies decreasing as the terminus of replication is approached, corresponding to a high degree of gene rearrangement, duplication, or deletion in this region. Indeed, much of the size differences can be accounted for by expansion in this area of the genomes of EAN and ACN. Genes related to symbiosists shown in Figure 3 include those encoding nitrogenase (nif), uptake hydrogenase (hup), and qualeme biosynthetic (ubx) involved in bacteria-photosynthetic symbioses. Only one gene similar to the common nodulation genes in rhizobia has been found in each strain, but with a general function prediction and relatively low BLAST scores (the product of FRAAL4911, annotated as a chitin deacetylase, resembles NodB from Rhizobium sp. N33 with a score of 3e^-24 but also resembles similar proteins from many other organisms, mainly Bacillus sp.).

Little is known about the genetic basis of the actinorhizal symbiosis, but it is clearly very different from that known to exist among the nodulating prokaryotes.

15 elements and prophase

Insertion elements, transposases, integrated phage, and plasmids tend to reflect the degree of plasticity of genomes (Frost et al. 2005). In the three Frankia strains, integras plus transposases comprise 0.6% (46 of the 7976 CDS), 4.1% (187 of the Cdi CDS), and 3.4% (269) of the EAN CDS (Table 3). Such elements tend to cluster in regions where there is loss of synteny between genomes (Fig. 3). Cci3 has a higher average density of mobile elements (34 per Mb) than EAN (30 per Mb), and both have a much higher density than ACN (6 per Mb). Indeed, of the 33 transposase genes identified in ACN, all but four are found in the same context in Cdi3, and all but six are in EAN, suggesting that such genes in ACN have been inactive for at least the 55 million years since the genomes diverged (Normand et al. 1999; Clawson et al. 2004). Examining regions surrounding phage integrases, we estimate that phage contribute 0.4%, 11.7%, and 7.1% of the CDS of ACN, Cci3, and EAN, respectively. Overall, EAN and Cci3 have had far more dynamic genomes than ACN in their recent history, and this plasticity, conferred by IS elements and phage, has been moving in and out of the genomes, may in large part have driven the size differences observed.

Gene deletions

To examine how the three genomes have evolved to have such different sizes, we used the order of divergence of the three strains (Fig. 1) to estimate how the genome size differences reflect gene deletion, duplication, and acquisition. Using the approximation of 50 MYr per 1% divergence in the 16S rRNA genes
with information from Silvester (1977) and from the Missouri Botanical garden of ACN and CcB.

A similar approach can be used for identifying deleted genes from ACN. However, genes absent from EAN, but present in the other two strains, could either have been lost in EAN or acquired by horizontal gene transfer (HGT) in the progenitor of ACN and CcB.

Table 2. Summary of genome characteristics

<table>
<thead>
<tr>
<th></th>
<th>ACN</th>
<th>CcB</th>
<th>EAN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accession</td>
<td>CT173213</td>
<td>CP000249</td>
<td>AE000000</td>
</tr>
<tr>
<td>Size in bp</td>
<td>7,491,934</td>
<td>5,433,628</td>
<td>9,035,218</td>
</tr>
<tr>
<td>Genes in COGs</td>
<td>4502 (67%)</td>
<td>2564 (57%)</td>
<td>4815 (60%)</td>
</tr>
<tr>
<td>RNA</td>
<td>46</td>
<td>46</td>
<td>47</td>
</tr>
<tr>
<td>%GC</td>
<td>72.83%</td>
<td>70.07%</td>
<td>70.94%</td>
</tr>
</tbody>
</table>

*Numbers are from accession numbers indicated as of November 13, 2006. Numbers are from http://www.genoscope.cns.fr.

Several genes lost by CcB are concerned with metabolic activities of potential importance to survival or symbiosis. These include genes encoding the DNA repair enzymes AP endonuclease, photolysase, DNA-formamidopyrimidine glycosylase, DNA alkylation repair, and RadC; two cellulosomes that might be involved in survival or infection; gas vesicle proteins whose loss could signal adaptation to dry environments where Casuarina sp. grow; general metabolism enzymes (NAD-dependent glutamate dehydrogenase, PEP carboxylase); and a large number of regulatory and solute transport proteins. Among the latter, there is only one iron siderephore gene cluster in CcB as compared with two in ACN and three in EAN. More directly related to symbiosis, CcB has lost one of the two copies of the sic (squalene hopene cyclase) genes involved in synthesizing bacteriochlorophyll lipids that comprise the envelope of Frankia vesicles and provide protection for nitrogenase against oxygen. Unlike ACN and EAN, oxygen protection is conferred by secondary plant cell walls when CcB is in symbiosis (Berg and McDowell 1988), perhaps making bacteriochlorophyll synthesis less of a priority.

In general, the classes of genes lost by CcB (DNA repair, metabolic enzymes, regulatory proteins) resemble those known to be lost by bacterial endosymbionts of animals (Mira et al. 1999), the clade containing EAN diverged an estimated 115 Mya from the clade containing CcB and ACN, which diverged from each other about 55 Mya (Fig. 1). Therefore, orthologous genes present in ACN and EAN but absent in CcB may be assumed to have been lost from CcB after it diverged from ACN. A similar approach can be used for identifying deleted genes from ACN. However, genes absent from EAN, but present in the other two strains, could either have been lost in EAN or acquired by horizontal gene transfer (HGT) in the progenitor of ACN and CcB.

Table 2. Summary of genome characteristics

<table>
<thead>
<tr>
<th></th>
<th>ACN (finished)</th>
<th>CcB (finished)</th>
<th>EAN (draft-1 scaffold)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size in bp</td>
<td>7,491,934</td>
<td>5,433,628</td>
<td>9,035,218</td>
</tr>
<tr>
<td>Genes in COGs</td>
<td>4502 (67%)</td>
<td>2564 (57%)</td>
<td>4815 (60%)</td>
</tr>
<tr>
<td>RNA</td>
<td>46</td>
<td>46</td>
<td>47</td>
</tr>
<tr>
<td>%GC</td>
<td>72.83%</td>
<td>70.07%</td>
<td>70.94%</td>
</tr>
</tbody>
</table>

*Numbers are from accession numbers indicated as of November 13, 2006. Numbers are from http://www.genoscope.cns.fr.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
2001; Ochman and Moran 2001), and indicate that Cc3 is evolving toward a greater dependence on its host. However, Cc3 can still grow on minimal medium, so such strains have not yet been committed to an obligate symbiotic existence.

Gene duplication, acquisition, and ORFans
Gene duplication is a major means by which soil bacteria adapt to new niches, or to the availability of new substrates (Francino 2005; Konstantinidis and Tiedje 2005). Gene acquisition is known to be similarly involved in bacterial adaptation to new environments, particularly in the emergence of pathogens (Mira et al. 2001; Ochman and Moran 2001) and in the evolution of mutualistic bacteria in the legume symbiosis (Young and Haukka 1996; Chen et al. 2001; Moulin et al. 2004). Frankia symbionts have adopted both to living in diverse soils in most parts of the world and to living in root nodules from phylogenetically diverse angiosperms.

We defined duplicates as having the lowest BLAST E-value with a gene from the same genome when compared with genomes from other Frankia strains, Acidithiobacillus and Kineococcus.

Table 3. Summary of genes involved in size differentiation of the three Frankia strains

<table>
<thead>
<tr>
<th>Category</th>
<th>Cc3</th>
<th>ACN</th>
<th>EAN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deleted genes</td>
<td>466</td>
<td>1054</td>
<td>555</td>
</tr>
<tr>
<td>Duplicated genes</td>
<td>512 (7.5%) 444 (6.8%)</td>
<td>1355 (18.5%)</td>
<td></td>
</tr>
<tr>
<td>Strain-specific genes</td>
<td>709 (10.4%)</td>
<td>420 (5.9%)</td>
<td>934 (11.7%)</td>
</tr>
<tr>
<td>ORFans</td>
<td>854 (12.3%)</td>
<td>158 (2.3%)</td>
<td>355 (4.9%)</td>
</tr>
<tr>
<td>Transposases</td>
<td>33 (0.4%)</td>
<td>125 (1.4%)</td>
<td>195 (2.4%)</td>
</tr>
<tr>
<td>Integrase</td>
<td>13 (0.2%)</td>
<td>32 (0.7%)</td>
<td>74 (0.9%)</td>
</tr>
</tbody>
</table>

Deleted genes, duplicated genes, ORFans (no hits in any database), and strain-specific genes (SSG, hits in databases but not in the other two Frankia strains) were detected as described in the text. The number of genes annotated as transposases and inactivated derivatives plus integrases were annotated as described.

Figure 3. Genome maps of the three Frankia strains. Circles, from the outside in, show (1) gene regions related to symbiosis including shcl, hup2, hup1, and nif (2) the coordinates in Mb beginning at 0 = oriC; (3) regions of synteny (syntons) calculated as a minimum of five contiguous genes present in all strains with an identity >20% over 80% of the length of the shortest gene (syntons are tagged with a spectrum-based [red-yellow-green] color code standardized on ACN to indicate regions where syntons have moved in the other strains); (4) IS elements and transposases. Circles were drawn using GenVision Software from DNAStar.

both close relatives to Frankia in the Frankineae, Streptomyces spp., and the NR (non-redundant) database. Using this approach, ~7.5% (512) of the ORFs in ACN, 9.8% (444) in Cc3, and 18.5% (1355) in EAN could be considered duplicates of other genes in the same genomes (Table 3). Core metabolic genes are generally not duplicated, a differential amplification noted in other bacteria (Francino 2005; Konstantinidis and Tiedje 2005). Surprisingly, Cc3, which has sustained strong reducing evolutionary pressures, nevertheless had a slightly higher percentage of duplicates than ACN, an observation that is accounted for by the proliferation of transposase genes in Cc3 (Tables 3, 5). Gene duplication has thus enlarged the EAN genome to a greater ex

Table 4. Categories of deleted genes assessed by BlastClust analysis

<table>
<thead>
<tr>
<th>General category</th>
<th>Cc3</th>
<th>ACN</th>
<th>EAN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypothetical, conserved hypothetical, unknown function</td>
<td>289</td>
<td>158</td>
<td>176</td>
</tr>
<tr>
<td>Transport-associated</td>
<td>113</td>
<td>33</td>
<td>28</td>
</tr>
<tr>
<td>Regulatory</td>
<td>95</td>
<td>21</td>
<td>35</td>
</tr>
<tr>
<td>Short-chain dehydrogenase/reductase</td>
<td>32</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Acyl-CoA dehydrogenase-like</td>
<td>19</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Aminosuccinate monooxygenase</td>
<td>17</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>AMP-dependent synthetase and ligase</td>
<td>14</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Protein kinase</td>
<td>12</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>Amidohydrolase</td>
<td>13</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Enoyl-CoA hydratase/isomerase</td>
<td>10</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>L-carnitine dehydratase/bile acid</td>
<td>9</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Inducible protein F</td>
<td>7</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Alcohol dehydrogenase G015-like</td>
<td>8</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>Cytochrome P450</td>
<td>7</td>
<td>3</td>
<td>11</td>
</tr>
<tr>
<td>Transposases</td>
<td>3</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>Integrase</td>
<td>406</td>
<td>192</td>
<td>276</td>
</tr>
</tbody>
</table>

Deleted genes are defined as genes present in two of the three strains as assessed by BLAST hits below a cutoff of 10^-30 but absent in the third. The general categories correspond to the major groups identified by a BlastClust analysis of missing genes.
reductases, endonucleases, SAM-dependent methyltransferases, transport proteins, and a variety of dehydrogenases are duplicated. In EAN, 132 out of 406 (32.9%) genes are associated with integases, transposases, or reverse transcriptases in the top 20 families, with the remainder annotated as short-chain dehydrogenase/reductase, cytochrome P450s, transport proteins, or regulatory proteins, and dioxygenases.

In summary, EAN has the most duplicated genes in all categories, including those whose products are associated with metabolic processes as well as mobile genetic elements. ACN has the fewest duplicates, and those are of genes involved in general metabolism. Finally, a large portion of all duplicates in CC23 (33% overall) is of transposases. In all strains, the majority of duplicates appeared as two copies of a single gene.

Table 5. Top 20 families of duplicated genes in each Frankia strain assessed by BlastClust analysis

<table>
<thead>
<tr>
<th>Annotated function</th>
<th>Annotated function</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACN</td>
<td>CC23</td>
</tr>
<tr>
<td>Putative serine/threonine protein kinase</td>
<td>Transposase, IS54</td>
</tr>
<tr>
<td>Putative acetohydroxyacid synthase, short-chain dehydrogenase/reductase family</td>
<td>Transposase IS56</td>
</tr>
<tr>
<td>Hypothetical protein; putative HNH endonuclease domain</td>
<td>Transposase, IS54</td>
</tr>
<tr>
<td>Hypothetical protein; putative dehydrogenase</td>
<td>Transposase</td>
</tr>
<tr>
<td>Putative isopenicillin N synthase transport protein (ABC superfamily)</td>
<td>Regulatory protein, MerR/RecA-like</td>
</tr>
<tr>
<td>Putative SAM-dependent methyltransferase</td>
<td>Twin-arginine translocation pathway signal</td>
</tr>
<tr>
<td>Putative alpha-beta-CoA racemase</td>
<td>Transposase, IS54</td>
</tr>
<tr>
<td>Branched-chain amino acid ABC transport, binding protein</td>
<td>Transposase and inactivated derivatives-like</td>
</tr>
<tr>
<td>Hypothetical protein; putative signal peptide</td>
<td>Transposase, IS111A/IS128/IS1533:Transposase IS116/IS170/IS5902</td>
</tr>
<tr>
<td>Putative non-ribosomal peptide synthetase</td>
<td>Transposase (probable), IS81/IS136/IS1341:Transposase, IS605 OrfB</td>
</tr>
<tr>
<td>Branched-chain amino acid transport protein (ABC superfamily)</td>
<td>Putative IS630 family transposase</td>
</tr>
<tr>
<td>Conserved hypothetical protein; putative amidohydrolase domain</td>
<td>Hydantoinase/oxopeptidase</td>
</tr>
<tr>
<td>Putative GlnK-family transcriptional regulator</td>
<td>ATP-binding region, ATPase-like</td>
</tr>
<tr>
<td>Putative TetR-family transcriptional regulator</td>
<td>Putative O-methyltransferase</td>
</tr>
<tr>
<td>Hypothetical protein; putative dihydrolipoamide desulfurization protein</td>
<td>Putative plasmid replication initiator protein</td>
</tr>
<tr>
<td>Cytochrome P450</td>
<td>Transposase, IS54</td>
</tr>
<tr>
<td>Hypothetical protein</td>
<td>Hypothetical protein</td>
</tr>
<tr>
<td>Putative monooxygenase</td>
<td>Amino acid adenylation</td>
</tr>
<tr>
<td>Hypothetical protein</td>
<td>Putative DNA-binding protein</td>
</tr>
<tr>
<td>Putative aldolase dehydrogenase</td>
<td>Hypothetical protein</td>
</tr>
<tr>
<td>Putative aldehyde dehydrogenase</td>
<td>Hypothetical protein</td>
</tr>
</tbody>
</table>

The top 20 categories of duplicated genes were defined by a BlastClust analysis of a data set comprised of all duplicates. Duplicates were defined as having the best BLAST score of another gene within the same genome relative to genes within NR plus the other Frankia strains, Acidothermus, Kineococcus, and Streptomyces spp. with a minimum cutoff of 10^-4. Finally, a large portion of all duplicates in CC23 (33% overall) is of transposases. In all strains, the majority of duplicates appeared as two copies of a single gene.

Stain-specific genes (SSGs) include genes lost by two of the three Frankia strains plus genes that have no hits in databases (ORFans). Such genes could also have been horizontally transferred from other bacteria. Using a permissive threshold (E-value ≤ 10^-3) between the genomes, and allowing self-genome hits to eliminate duplication, we found that 23% (1,563) of the genes in ACN, 12.8% (578) in CC23, and 17.7% (1,289) in EAN...
have no clear homologs in the other two genomes. Of those, 854 (12.5%) in ACN, 158 (3.5%) in Cc13, and 355 (8.9%) in EAN were ORFans with no hits in NR, or the related Acidiphilus, Kinecoccus, or Streptomyces spp. genomes. The higher number and percentage of ORFans in ACN may reflect a lower evolutionary pressure to eliminate non-essential genes, a characteristic also reflected in its having the fewest deleted genes overall (Table 3).

Conclusions
We have shown that the unusual size divergence displayed by the *Frankia* genomes has arisen by the processes of deletion, duplication, and retention/ acquisition operating in all strains but to different extents (Table 3). These processes have driven the genomes in different directions, reducing that of Cc13, expanding that of EAN, and keeping ACN relatively stable. The results of these broad comparisons lead us to propose a link between the biogeographic history of the actinorhizal plants and the genome evolution of the bacterial symbionts.

Evidence from ecological (Zinsmeister et al. 1997), molecular ecological (Simonet et al. 1999), physiological (Selseth 1995), and now genomic studies indicates that Casuaria strains represented by Cc13 have evolved to become specialists with reduced genomes. Unlike Alnius and Elaeagnus strains, they have not been detected by trapping experiments in soils outside the native ranges of their host plants (Zinsmeister et al. 1997; Simonet et al. 1999), and they infect a narrow spectrum of hosts (Fig. 2A; Table 1).

Genome reduction is well documented in obligate pathogens and obligate symbionts in plants and animals (Mira et al. 2001; Ochman and Moran 2001; Moran 2003; Batut et al. 2004), and in some free-living cyanobacterial Prochlorococcus sp. (Duarte et al. 2005). Genome reduction has not been described in bacterial facultative symbionts that also exist free-living in the soil; indeed, this is a most unexpected finding.

We suggest that a likely explanation for genome reduction in Cc13 is its geographic and symbiotic isolation in Australia and the Pacific islands, paralleling its host plant's isolation beginning 100-65 Ma. Casuariaceae species emerged as part of the flora of Gondwana as evidenced by fossils in New Zealand and South America today that are outside the native range (Campbell and Holden 1984). These plants, and their bacterial symbionts, adapted to a hotter, drier climate as Australia split from Antarctica and moved north toward the equator. Present-day Casuaria strains live in locales where the soil biotic capacity is reduced and actinorhizal host diversity is limited.

In contrast, plants infected by Elaeagnus strains have a global distribution (Fig. 2C) with ancestral origins in both Gondwana (Gymnostoma in the Casuariaceae in Western Oceania, actinorhizal Colletiae in the Rhamnaceae) and Laurasia (Elaeagnaceae, Myricaceae). Such plants occupy a wide range of soil types and climates. Genome expansion by gene duplication and divergence is a mechanism used by soil bacteria to exploit new niches and new substrates (Francisco 2005; Konstantinidis and Tiedje 2005), and may be inferred to have occurred in the ancestors of EAN as they and their hosts coadapted to new and diverse soils. Indeed, the types of genes duplicated are largely involved in introducing substrates into central metabolic pathways.

The genome of ACN appears more stable than those of Cc13 and EAN, in the sense that it has few transposases and integrases; it also has lost the fewest genes by deletion, has the lowest proportion of duplicated genes, and retains the most strain-specific genes, including ORFans. Its stability may reflect its host range focused on the ancient lineages in the Betulaceae and Myricaceae leading to high soil abundance and relatively strong genome homogenization. Its host plants have the longest fossil record of the N2-fixing clade (Magallon et al. 1999), and have inhabited similar and milder environments in northern latitudes since appearing in Laurasia during the late Cretaceous (Crane 1989).

Taken together, the gene content of these actinorhizal strains appear to reflect the biogeographic history of the host plants they infect, and as such may provide the first example of differential genome contraction and expansion occurring in closely related facultatively symbiotic soil bacteria that may be linked to the evolutionary history of their hosts on a global scale.

Methods

Strains
Cc13 was isolated from *Casuarina cunninghamiana* plants growing in a greenhouse at Harvard Forest in Petersham, MA (Zhang et al. 1984) on soils coming from its original provenance. ACN was isolated initially from *Alnus viridis* subsp. *crispa* plants in Tadousac, Quebec (Normand and Lalonde 1982; Benson et al. 2004). Strain EAN was isolated from field nodules of *E. angustifolia* growing in Ohio (Lalonde et al. 1981).

Genome sequencing, assembly, and finishing—Cc13 and EANpec
We sequenced the three genomes of *Frankia* strains ACN14a, Cc13, and EAN1pec using a shotgun approach. The genomes of *Frankia* strains Cc13 and EAN1pec were sequenced at the Joint Genome Institute (JGI) using a combination of 34-bp, 84-bp, and 40-kb (fosmid) DNA libraries for each strain. Draft assemblies were based on 82,561 total reads for Cc13 and 125,615 total reads for EAN1pec. The different libraries provided 4.6× (34-bp), 4.1× (84-bp), and 0.5× (fosmid) coverage of Cc13 and 4.9× (34-bp), 3.4× (84-bp), and 0.6× (fosmid) coverage of EAN1pec. End sequencing and fingerprinting of fosmid clones aided in assembly verification, determination of gap sizes, and ordering and orientation of scaffolds beyond assembly gaps.

Sequencing gaps were closed mainly by primer walking on plasmid and fosmid subclone templates. In cases where no acceptable template was available,PCR products were made and sequenced using customized primers. Gaps resulting from hard-to-sequence DNA structures had to be covered using special chemistries and protocols developed in-house. Mis-assemblies were identified and corrected by means of clone pairing; these primarily occurred due to long repeats (rRNA, IS elements). Over-collapsing of repeat copies often resulted in pseudo-gaps in the assembly, which could not be closed by routine primer walking. Each one of those had to be filled in using one of the following two methods. Small pseudo-gaps were closed using the editing features of CONSED (Gooden et al. 1998), by locating and placing appropriate reads individually into their proper repeat copy. Long pseudo-gaps and long misassembled repeats (>2 kb) had to be isolated and separately assembled. Only consistent, partially unique clone-mates would be allowed in those subassemblies. After verifying the subassembly's integrity and primer-walking over the poorly covered regions, the isolated contigs were reintroduced into the main assembly as "fake reads," that is, single continuous long sequences reflecting the correctly assembled repeat copy.

All other general aspects of library construction, sequencing, and automated annotation were carried out as previously de-
scribed for bacterial genomes sequenced at the JGI (Chain et al. 2003). In addition, predicted coding sequences are subject to manual analysis using the Integrated Microbial Genomes (IMG) annotation pipeline. Detailed information about genome annotation and other genome properties can be obtained at https://img.jgi.doe.gov (Markowitz et al. 2006).

Genome sequencing, assembly, and annotation—ACNI4a

For ACNI4a, four libraries were made: Two plasmid libraries of 3 kb and 10 kb, obtained by mechanical shearing, were constructed at Genoscope (Evry, France) into pCIS2A.1 (Invitrogen) and into the pCNS home vector (pSU18 modified, Bartolome et al. 1991), respectively. Two BAC libraries of an average insert size of 104 kb were constructed at Clemson University Genomics Institute (CUGI) by enzymatic digestion (EcoRI and HindIII) into pCUGI8AC1 (Luo and Wing 2003). Plasmid and BAC DNAs were purified and end-sequenced using dye-terminator chemistry on ABI3730xl DNA Analyzer sequencers. We generated 150,890 sequences from both ends of genomic clones from the four libraries.

The Phred/Phrap/Consed software package (www.phrap.com) was used for sequence assembly and quality assessment (Ewing and Green 1998; Ewing et al. 1998; Gordon et al. 1998). A total of 8956 additional reactions were necessary to close gaps, generally with the transposing method (Entrancceposon, Finnyzmos, Espoo, Finland) and to raise the quality of the finished sequence. The validity of the final sequence was assessed by comparing the restriction enzyme pattern deduced from the sequenced and experimentally observed restriction pattern obtained by digestion of genomic DNA. After a first round of annotation, regions of lower quality as well as regions with putative frame-shifts were resequenced from PCR amplification of the dubious regions.

Using the AMIGene software (Bocs et al. 2003), a total of 5279 CDSs were predicted and submitted to automatic functional annotation (Vallenet et al. 2006). Each predicted gene was assigned a unique identifier prefixed with "FRAA," Sequence data for comparative analyses were obtained from the NCBI databank (RefSeq section). Putative orthologs and syntenic groups (conservation of the chromosomal co-localization between pairs of orthologous genes from different genomes) were computed between ACN and the 265 other complete genomes using the procedure described in Vallenet et al. (2006). Manual validation of the automatic annotation was performed using the Magne (Magnifying Genomes) interface, which allows graphic visualization of the ACN annotations enhanced by a synchronized representation of synteny groups in other genomes chosen for comparisons. The ACN nucleotide sequence and annotation data have been deposited at EMBL, databank under accession number CT573213. In addition, all the data (i.e., syntactic and functional annotations, and results of comparative analysis) were stored in a relational database, called Frankiakode (Vallenet et al. 2006). This database is publicly available via the Magne interface at http://www.genoscope.cns.fr/agc/mage/frankia/Login/log.php.

Methods used for determining deleted, duplicated, strain-specific, and ORFan genes

Genes deleted from one strain were identified by using reciprocal BLAST hits from each pair of genomes. That is, each pair of orthology identified each other as the lowest BLAST score. Genes were scored as deleted if they did not have a reciprocal hit in another Frankia genome. The data set included three Frankia strains.

Gene duplications were assessed as having the best BLAST hits within the same genome (duplicates) using an E-value cut-off of 10^-4 and a data set consisting of NR (minus Frankia sequences) + Kluveromyces lactis + Streptomyces coelicolor + S. avermitilis + ACN + CeC1 + EAN. To cluster duplicates, the program BlastClust (NCBI) was used with settings reported in the text. Strain-specific genes (SSGs) include genes found in one but not another Frankia strain at an E-value cutoff of 10^-4, plus genes that have no hits in databases. The latter are referred to as ORFans.

Acknowledgments

This work was supported by the National Science Foundation Microbial Genome sequencing program to D.B., L.S.T., and M.P.F. The work on CeC1 and EAN was performed under the auspices of the U.S. Department of Energy's Office of Science, Biological, and Environmental Research Program, and by the University of California, Lawrence Livermore National Laboratory, and Los Alamos National Laboratory. The work on ACNI4a was performed at Genoscope, Evry, France and was supported by CNRS/ACI Microbiologie and MRT/ACI IMBRe2004. Funds to Anita Selbstedt are from FORMAS and NFR, Sweden. Funds for the construction of the ACN BAC library were provided in part by the Genome Science and Technology Program, The University of Tennessee. Correspondence and requests for materials should be addressed to D.R.B. (david.benson@ncmrr.gov) for CeC1, to P.N. (normand@biohem.univ-lyon1.fr) for ACN, and to L.S.T. (tib@hypha.unih-saarland.de) for EAN.

References


Frankie genome evolution


Received August 1, 2006; accepted in revised form October 18, 2006.
CHAPTER IV

TRANSPOSON MUTAGENESIS OF FRANKIA ISOLATE EUI1C BY THE CONJUGATIVE TRANSPOSON Tn916 OF ENTEROCOCCUS FAECALIS

Abstract

Genetic analysis of Frankia is in its infancy and is a major obstacle to both the applied and basic aspects of the field. Although three Frankia genomes have been sequenced, very little is understood about gene expression. There are no known systems for gene transfer and reliable standardized mutagenesis protocols have not yet been firmly established. The development of transposon mutagenesis protocols to generate Frankia "knock-out" mutants would provide a powerful tool for genetic analysis. The conjugative transposon Tn916 was introduced into Frankia strains EUI1c and EAN1pec by mating with the donor Enterococcus faecalis GC110. Isolated putative transconjugants were stable and have been maintained in culture. The generation of stable tetracycline and novobiocin resistant transconjugants suggests that the transposon inserted into the genome. Several methods were used to verify the insertion of Tn916 into the genome. First, primer sets for the tetM gene and 1343 bp region of the Tn916 were used to confirm the presence of the transposon in the transconjugants. Amplicons for each primer set were generated with gDNA from the donor cells (E. faecalis) and the transconjugants, but were not
amplified with gDNA from the wild type *Frankia* control. A second primer set for the *Frankia*-specific glutamine synthetase II gene (*glnII*) was used to confirm that all of the transconjugants were *Frankia*. Amplicons were generated with gDNA from the wild type *Frankia* control (recipient cells) and the transconjugants, but were not amplified with gDNA from the donor cells (*E. faecalis*). Two more experiments were performed to determine if these transconjugants were *Enterococcus* cells on the surface of the *Frankia* or "true" transconjugants. The use of primer sets for *E. faecalis*-specific genes only resulted in the generation of amplicons with gDNA from the donor cells (*E. faecalis*) and not from either the transconjugants or recipient *Frankia* controls. Variable region 3 (V3) of the 16S rRNA gene from *Frankia* and *E. faecalis* was amplified and analyzed by denaturing gradient gel electrophoresis (DGGE). Only a single DGGE band was observed for all of the samples and the G+C rich *Frankia* bands migrated much further into the denaturant than the *E. faecalis* band. All of these results indicate that the Tn916 was present in the *Frankia* transconjugants. The presence of Tn916 was also confirmed by Southern hybridization experiments. A 1343 bp probe was developed and used on Scal-digested gDNA. A single band was observed in the donor (*E. faecalis*) and some transconjugants. The observed differences in the position of the bands suggest different locations for the transposon in the chromosome. The probe did not hybridize to gDNA from the wild type recipient *Frankia* cells. These data are consistent with the transposition of Tn916 into the *Frankia* genome. The transposon insertion sites have also been mapped to a physical map of the *Frankia*
chromosome that was developed by macrorestriction analysis. This is the first successful transfer and expression of foreign DNA into *Frankia*.

**Introduction**

*Frankia*, a member of the bacterial family actinomycetes, are distinguished by their unique ability to induce nitrogen fixation in root nodules of certain nonleguminous plants (Torrey *et al.*, 1979). These bacteria form symbiotic nitrogen-fixing associations with over 200 different species of actinorhizal plants. This symbiotic relationship contributes a significant amount of nitrogen to the planet. The ecological range of actinorhizal plants is quite diverse and varies from dry sandy, desert areas to wetlands and forests (Benson and Silvester, 1993). Due to their ability to survive in such diverse and nutrient-poor conditions, they are often used in many areas of ecological growth as a pioneer species to promote community development. As a result, they possess tremendous economic significance with respect to land reclamation, soil stabilization, landscaping, nurse cropping, and commercial lumber (Benoit and Berry, 1990). *Frankia* also have the ability to bind and sequester several toxic heavy metals (Richards *et al.*, 2002). This property gives it the potential for bioremediation and phytoremediation applications especially on heavy metal contaminated lands. In Western Europe, actinorhizal plants are used extensively for reclamation of industrial wastelands and mine spoils (Wheeler *et al.*, 1990).

In order to take full advantage of its economic and ecological potential, it is essential that the physiology and genetics of *Frankia* be thoroughly investigated.
Since the first isolation of *Frankia* in 1978 (Callaham *et al.*, 1978), there have been many advances in the knowledge and understanding of actinorhizal nitrogen fixation for the physiology, morphology, and biochemistry of these bacteria (Benson and Schultz, 1990, Wall 2000). The genome sequencing of three *Frankia* isolates has provided an enormous amount of information to elucidate the complex genetic system of this bacterium however, a mutagenesis protocol is vital to understand gene expression and to identify vital gene function.

Our current knowledge about nitrogen fixation has been greatly advanced because of the development of sophisticated methods of genetic analysis. The genetics of *Frankia* is in its infancy (for review see Lavire and Cournoyer, 2003). Due to the slow growth rate of *Frankia* the genetic tools are only beginning to be developed. Genetic analysis of *Frankia* has been restricted to gene cloning via hybridization to genes from other organisms, phylogenetic analyses based on selected gene sequences, and isolation and characterization of plasmids (Lavire and Cournoyer, 2003, Mullin and An 1990, Normand and Lalonde 1986). Currently, there is no known system for gene transfer for *Frankia*, there are no phages, R-plasmids, transposable elements, or conjugative plasmids and standardized mutagenesis protocols have not been firmly established.

Previous studies in the Tisa lab were initiated to test the use of the conjugative transposon Tn916 with *Frankia* (Myers *et al*. 2002). Transposons are an important and useful genetic tool. These genetic elements are able to enter a host and integrate into its chromosome (Salyers *et al.*, 1995, Burrus *et al.*, 2002). The conjugative transposon Tn916 is highly promiscuous and has the ability to transfer
into and between gram-positive and gram-negative bacteria (Bertram et al., 1991). Tn916 was introduced into *Frankia* by mating strain Eu1c with *Enterococcus faecalis* GC110. Tn916 is one of the most intensively studied conjugative transposons; it is 18.5kb in size and encodes the antimicrobial resistance determinant *tetM* which produces resistance to tetracycline and minocycline by protection of the ribosome (Flannagan et al., 1994). Since Tn916 transfers readily to a wide variety of Gram-positive and Gram-negative bacteria (Bertram et al., 1991), it is a good candidate to test as a genetic tool for *Frankia*. Several potential *Frankia* Eu1c transconjugants were generated that were resistant to both novobiocin and tetracycline. Strain Eu1c is novobiocin-resistant and the donor (*E. faecalis*) is sensitive. The purity of the transconjugant cultures with respect to contamination by the donor was checked (i) microscopically, (ii) by repeated streaking, and (iii) by counterselection in the presence of novobiocin. A concentration of 25 ug/ml was sufficient to completely inhibit growth of *E. faecalis*, whereas *Frankia* Eu1c was unaffected.

Filter and liquid matings produced transconjugants at $1.5 \times 10^{-4}$ and $5.3 \times 10^{-5}$ transconjugants/recipients, respectively. Novobiocin- and tetracycline-resistant colonies were never obtained from the control recipient cells without mating. Several of these putative transconjugants were isolated for further characterization. These transconjugants were stable and have been maintained in culture. These experiments were also repeated with another *Frankia* strain (EAN1pec) and several transconjugants were isolated (Dwyer, unpublished). The physiological properties of these putative *Frankia* Eu1c transconjugants were investigated (Myers et al. 2002).
Although conjugation generated stable-tetracycline-resistant *Frankia*, molecular methods are required to verify the insertion of Tn916 into the *Frankia* genome. The work described in this chapter was performed to confirm the transposition of Tn916 into the *Frankia* genome.

**Materials and Methods**

**Frankia Strains and Growth Conditions.**

All strains of *Frankia* and *Enterococcus faecalis* used in this study are listed in Table 1. *Frankia* strain Eu1c (Baker *et al.*, 1980) was grown and maintained in basal medium with NH₄Cl as the nitrogen source and 20mM glucose as carbon and energy source, as described previously (Tisa *et al.* 1983, Tisa *et al.* 1999). *Frankia* transconjugants were grown and maintained in basal growth medium supplemented with 0.05% casamine acids, 0.05% yeast extract, 20µg/ml novobiocin, 20µg/ml tetracycline and with NH₄Cl as the nitrogen source and 20mM glucose as a carbon source. *E. faecalis* GC110 was grown and maintained on LB medium supplemented with 20µg/ml tetracycline and incubated at 37°C with aeration.

**PCR Assays.**

PCR was performed with the Qiagen HotStarTaq™ (Qiagen, CA, USA) polymerase according to the manufacturer's recommendations. Amplification was performed in volumes of 50µl containing 25µl of the HotStarTaq™ Master Mix (supplied by the manufacturer), 30 pMol of each primer, and 250ng of template DNA. Product was amplified under the following parameters: 95°C for 15 min (initial activation), 30 cycles of 95°C for 1 min (denaturation), 61°C for 45 sec (annealing),
72°C for 1 min (extension) followed by a 5 min final extension at 72°C.

Oligonucleotide primers were purchased from Integrated DNA Technologies Inc. and are listed in Table 2.

**Genomic DNA Isolation and Purification**

The genomic DNA of the wildtype strains and mutant strains were extracted following the method described in Current Protocols in Molecular Biology Vol.1. Large volumes, concentrated solutions of genomic DNA (gDNA) were isolated from *Frankia* by the CTAB method (Ausubel et al. 1988). Hyphae were harvested by centrifugation at 12,000 x g for 15 min at room temperature and washed with MP buffer (50mM MOPS, 10mM KPO₄ pH 6.8). The hyphae were resuspended in TE buffer (10mM Tris, 1mM EDTA, pH 8.0) and a final concentration of 0.5% SDS, 100µg/ml of proteinase K and 50µg/ml of lysozyme was added. The suspension was mixed thoroughly and incubated for 1 h at 37°C. After incubation, 100µl of 5M NaCl was added and mixed thoroughly. Then 80µl of CTAB/NaCl solution was added and incubated at 65°C for 10 min. A chloroform/isoamyl alcohol extraction was performed and the DNA was precipitated with isopropanol and washed with 70% ethanol. After the DNA pellet was dry the pellet was resuspended in 100µl of TE buffer. DNA concentrations were determined by UV spectrophotometer (Applied Biosystems. CA, USA).

**Construction of a DIG-labeled Probe**

DNA was labeled by random priming with a non-radioactive dioxigenin (DIG) labeling kit (Boehringer Mannheim GmbH). A 1343bp probe was created by using primers specific for the transposon Tn916 (Table 2) and generated by PCR. E.
faecalis gDNA was used as template DNA and DIG-labeled dUTPs were added to
the PCR according to manufacturer's recommendations (Amersham). The DIG-
labeled PCR product was purified by using Qiagen Qiaquick gel extraction kit
according to manufacturer's recommendations (Qiagen, CA, USA). The purified
probe was stored in elution buffer at 4°C until used.

Restriction Enzyme Digests

Genomic DNA isolated from Frankia strain Eu1c, E. faecalis, and the
transconjugants were digested with restriction enzyme Scal. Restriction
endonuclease reactions were carried out as recommended by manufacturer (New
England Biolabs, Inc.).

Southern Blot Hybridization

After digestion with Scal, genomic DNA was separated on two identical 1%
agarose gels. One gel was stained with ethidium bromide and the other was used
for electro-transfer of the DNA to a nylon membrane (Zeta-Probe GT membrane,
Bio-Rad) using the Bio-Rad Transblot SD cell (Bio-Rad). Before transfer, the gel
was equilibrated for 15 minutes in 0.5 X TBE (44.5 mM Tris-base, 44.5 mM boric
acid, 1 mM EDTA at pH 8.0). A piece of extra thick block paper pre-soaked in 0.5 X
TBE was placed on the platinum anode of the transblot cell. A Zeta Probe GT
membrane (Bio-Rad) which was equilibrated for 10 min in 0.5 X TBE buffer was then
placed on top of the blot paper and the equilibrated gel was placed well-side up on
top of the membrane. Another pre-soaked extra thick blot paper covered the gel and
the cathode was placed on top and locked into place. The transfer was run at
15volts (3.55mA/cm) for 20 min. At the end of the run, the membrane was rinsed
briefly in 2 X SSC. The DNA was fixed on the membrane by placing it between two pieces of dry blot paper and baking at 80°C for 30 min. The membrane was stored at 23-25°C until further use.

Prehybridization and hybridization of the DIG-labeled probe was performed with the DIG-High Prime DNA labeling and detection starter kit (Boehringer Mannheim GmbH) according to manufacturer’s recommendations. The hybridization was carried out at 45°C overnight with gentle agitation. After hybridization, the membrane was washed in 2 X SSC, 0.1% SDS at room temperature twice for 5 min and then washed in 0.5 X SSC, 0.1% SDS at 68°C twice for 15 min each.

The probe was immunologically detected following manufacturer’s recommendations and the membrane was exposed to Hyperfilm ECL (Amersham) for 20 min. The film was developed as suggested by the manufacturer.

**Denaturant Gradient Gel Electrophoresis**

Denaturing Gradient Gel Electrophoresis was performed with the BioRad DCode™ Universal Mutation Detection System, as described previously (Naser 2003, Muyzer et al. 1993, Teske et al. 1996). The V3 region of the 16S rRNA gene was amplified by PCR with 338F and 907R primers (Table 2; Lane et al. 1985, Amann et al. 1990) using 250 ng of DNA template. A GC-clamp (Muyzer et al. 1998) was attached to the 5' end of the forward primer 338F for DGGE analysis. The PCR was performed in 25μl reaction volumes using 0.5μM of each primer and Qiagen HotStarTaq Master Mix (Qiagen, CA. USA). Thermal cycling parameters were as follows: i) an activation step at 95°C for 15 min was performed to activate HotstarTaq DNA polymerase, ii) DNA denaturation at 94°C for 2.5 min was followed
by 11 cycles of denaturation at 94°C for 30 s, primer annealing at 56°C for 45 s, and primer extension at 72°C for 1 min, iii) 11 cycles of denaturation at 94°C 30 s, primer annealing at 56°C for 1 min and primer extension at 72°C for 90 s followed, iv) finally 14 cycles of 30 sec denaturation at 94°C for 30 s, primer annealing at 56°C for 75 s, primer extension at 72°C for 2 min 15 s and an extension step of 7.5 min completed the procedure. PCR products were purified using QIAquick PCR Purification Kit (Qiagen, CA, USA). The samples were loaded on an 8% acrylamide gel with a 40%-60% denaturing gradient. The gel was run with 1XTAE buffer (2M Tris base, 1M Acetic Acid, glacial, 0.5M EDTA, pH 8.0 for 10 X TAE) at 70V for 16 hours and were stained for 20 min in 1x TAE buffer (Sambrook et al. 1989) containing ethidium bromide (50µg/ml). Gels were destained for 30 min in 1x TAE buffer and photographed on an UV transilluminator (FVSTI-88).

Pulse Field Gel Electrophoresis

Pulse Field Gel Electrophoresis was performed using the BioRad CHEF™ mapper system as described previously (Rawnsley and Tisa, 2007). All agarose gels (0.5%, 0.8% or 1.0% SeaPlaque GTG) were prepared in 0.5 X TBE (45mM Tris-HCl, 45mM Boric Acid and 0.5mM EDTA, pH 8) with 100µM thiourea. DNA samples were separated by the use of contour-clamped homogenous electric field (CHEF™). Pulse times and gel concentrations were optimized with respect to the length of DNA to be separated. A midrange and low range PFGE markers (New England Biolabs, MA, USA), as well as lambda DNA digested with HindIII, and Saccharomyces pombe yeast chromosomes (Bio-Rad) were used as molecular size standards.
Results

Introduction of the Conjugative Transposon

*Enterococcus faecalis* GC110 with Tn916 was chosen as a donor. *Frankia* strain Eu11c was used as a recipient. The results of conjugation experiments performed by Myers (2001) are presented in Table 3. In all cases, transfer of tetracycline resistance was observed. The purity of the transconjugant cultures with respect to contamination by the donor was checked (i) microscopically, (ii) by repeated streaking, and (iii) by counterselection in the presence of novobiocin. A concentration of 25 µg/ml novobiocin was sufficient to completely inhibit the growth of *E. faecalis*, whereas *Frankia* Eu11c was unaffected.

The conjugative transposon Tn916 was introduced into *Frankia* by mating strain Eu11c with *E. faecalis* GC110. Several *Frankia* Eu11c transconjugants were generated that were resistant to both novobiocin and tetracycline (Table 1). Strain Eu11c is novobiocin-resistant and the donor (*E. faecalis*) is sensitive. Both filter and liquid matings produced transconjugants (Table 3). Novobiocin- and tetracycline-resistant colonies were never obtained from the control recipient cells without mating. Several of these transconjugants were isolated for further characterization. These transconjugants were stable and maintained in broth cultures (Figure 1).

Detection and Evidence of the Conjugative Transposon

Conjugation with *E. faecalis* generated stable tetracycline-resistant *Frankia* which suggested that the transposon was inserted into the *Frankia* genome. Several methods were used to verify the insertion of Tn916. For the first approach, the
presence of the Tn916 in the transconjugants was confirmed by PCR. A 1343 bp product was amplified with gDNA from the donor cells (*E. faecalis*) and the transconjugants, but was not amplified with gDNA from the wild type control (Figure 2). These results indicate that transposon Tn916 was present in the DNA samples.

A second primer set for the *Frankia*-specific glutamine synthetase II gene (*glnW*) was used to confirm that all of the transconjugants were *Frankia*. The presence of the glutamine synthetase gene confirmed the identity of the *Frankia* strains. The 400 bp product was amplified with gDNA of the wild type recipient cells and transconjugants but was not amplified with gDNA of the donor cells (Figure 3).

Two more experiments were performed to determine if these PCR results were the result of *Enterococcus* cells on the surface of the *Frankia* or “true” transconjugants. The use of primer sets to detect a *E. faecalis*-specific gene (the conjugal transfer protein) resulted in the generation of the 1300 bp amplicon with gDNA from the donor cells (*E. faecalis*), but a similar product was not amplified with gDNA of the transconjugants and the wild type recipient cells (Figure 4).

Variable region 3 (V3) of the 16S rRNA gene of these samples was amplified and analyzed by denaturing gradient gel electrophoresis (DGGE). DGGE resolves these samples based on their nucleotide sequence and has been extensively used in community profiling. PCR was performed with universal primers to amplify the 16S rRNA with gDNA of the wild type recipient cells, transconjugants and the donor cells. The primers used in this PCR included a GC-clamp so that the products could then be run on a denaturant gradient gel. Only a single DGGE band was observed for all of these samples and the G+C rich *Frankia* bands migrated much further into the

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
denaturant than the *E. faecalis* band (Fig. 5). The products of the GC-rich wild type recipient cell and the transconjugants ran to very similar locations on the denaturant gradient gel, whereas the product of the AT-rich donor cell ran to a distinctly different location on the denaturant gradient gel (Figure 5).

Pulse field gel electrophoresis (PFGE) was used to investigate noticeable changes in the restriction patterns of the *Frankia* physical map. Genomic DNA of the recipient wild type and transconjugant cells were isolated within agarose plugs and digested with various restriction endonucleases. After the DNA fragments were separated on an agarose gel by PFGE, the banding patterns were compared. The restriction patterns of the wild type recipient cells and transconjugant cells were similar (Figure 6). This would confirm the above PCR results indicating that the transconjugants were not maintaining surface contamination by the donor strain. A closer analysis of the *NdeI* restriction patterns showed that the patterns of two transconjugants differed from the wild type recipient by the absence of a band each: a 290 kb fragment in transconjugant *Eu1c/EET2* and a 300 kb fragment in transconjugant *Eu1c/EET4*. With *AseI* digest DNA, a 70 kb fragment found in the wild-type was absent in transconjugant *Eu1c/EET2* (Figure 6). These data also support the hypothesis that Tn916 was inserted into the genome of the transconjugants.

As a final line of evidence, the presence of the Tn916 was confirmed by Southern hybridization experiments (Figure 7). A 1343 bp probe was developed and used on *ScaI* digested gDNA. A single band was observed in the transconjugant EET1 and EET3 and three bands where observed with transconjugants EET6 and EET8, and a
positive result was seen with donor \((E. \text{faecalis})\) lanes. The probe did not hybridize to gDNA from the wild type recipient cells. These results indicate that the transposon was introduced into the \textit{Frankia} genome and also suggests multiple insertions of the transposon into the genome.

**Discussion**

The development of a transposon mutagenesis protocol to generate \textit{Frankia} "knock-out" mutants is essential for identifying the function of vital genes and will greatly facilitate the studying of \textit{Frankia} genetics. \textit{Enterococcus faecalis} GC110, containing the conjugative transposon Tn916 (tetM), was mated with \textit{Frankia} Eu11c and EAN1pec by both filter and liquid matings. This transposon readily inserts itself into foreign DNA and encodes for the antimicrobial resistance determinant tetM, which produces resistance to tetracycline and minocycline by protection of the ribosome. The donor \((E. \text{faecalis})\) is novobiocin-sensitive and tetracycline-resistant and the recipient \((Frankia)\) is novobiocin-resistant and tetracycline-sensitive. The transconjugants generated in this study are novobiocin-resistant and tetracycline-resistant and have been maintained stably in culture (Figure 1). Additional experiments, including PCR, DGGE, and PFGE have been completed to provide evidence for the successful transformation of \textit{Frankia} (Figures 2-6). Although, I tried to be consistent throughout the experimentation and perform all experiments on all the potential transconjugants, however due to the slow growth rate of \textit{Frankia} the cultures are easily contaminated by faster growing microorganisms and therefore not all transconjugants were available at the time of each experiments. This is the reason for the inconsistency among transconjugants in each experiment. PCR
confirmed the presence of the transposon Tn916 in both the donor and mutant isolates EET1, EET2, EET4, EET6, EET7, EET8 and EET10, and also confirmed Frankia specific genes in the wildtype recipient isolate and the mutant isolates (Figure 2-4). DGGE confirmed the genus of the mutant isolates as Frankia by using PCR to amplify the 16S rRNA hypervariable region and running the amplified product through a denaturing gradient which is able to separate DNA fragments according to single base pair differences (Figure 5). Lastly, Southern blot hybridization also confirmed multiple insertions into the genome in some transconjugants including a single insertion in EET1 and EET5, and multiple insertions in EET6 and EET10 (Figure 7). These results are of significant since this is the first time that foreign DNA has been shown to be expressed in Frankia.

This study shows that transposon mutagenesis is an effective method for constructing Frankia mutants and for identifying vital genes involved in sustaining this complex symbiosis. The experiments described in this study are imperative for the advancement in knowledge and understanding of this ecologically and economically important plant symbiont. The generation of vital genetic tools will provide an opportunity to further investigate all aspects of this symbiosis including infectivity, vesicle development, sporulation, and nitrogen fixation, as well as provide a means for comparative studies among other nitrogen fixing plant symbionts.
<table>
<thead>
<tr>
<th>Frankia Strain</th>
<th>Relevant Phenotype</th>
<th>Reference or Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>EuI1c</td>
<td>Parental wild type Novobiocin-resistant Tetracycline-sensitive</td>
<td>Baker et al., 1980; Tisa et al., 1999</td>
</tr>
<tr>
<td>EuI1c/EET1</td>
<td><em>E. faecalis</em> mating transconjugant Tetracycline-resistant (20µg/ml)</td>
<td>Myers 2001</td>
</tr>
<tr>
<td>EuI1c/EET2</td>
<td><em>E. faecalis</em> mating transconjugant Tetracycline-resistant (20µg/ml)</td>
<td>Myers 2001</td>
</tr>
<tr>
<td>EuI1c/EET3</td>
<td><em>E. faecalis</em> mating transconjugant Tetracycline-resistant (20µg/ml)</td>
<td>Myers 2001</td>
</tr>
<tr>
<td>EuI1c/EET4</td>
<td><em>E. faecalis</em> mating transconjugant Tetracycline-resistant (20µg/ml)</td>
<td>Myers 2001</td>
</tr>
<tr>
<td>EuI1c/EET6</td>
<td><em>E. faecalis</em> mating transconjugant Tetracycline-resistant (20µg/ml)</td>
<td>Myers 2001</td>
</tr>
<tr>
<td>EuI1c/EET7</td>
<td><em>E. faecalis</em> mating transconjugant Tetracycline-resistant (20µg/ml)</td>
<td>Myers 2001</td>
</tr>
<tr>
<td>EuI1c/EET8</td>
<td><em>E. faecalis</em> mating transconjugant Tetracycline-resistant (20µg/ml)</td>
<td>Myers 2001</td>
</tr>
<tr>
<td>EuI1c/EET10</td>
<td><em>E. faecalis</em> mating transconjugant Tetracycline-resistant (20µg/ml)</td>
<td>Myers 2001</td>
</tr>
<tr>
<td>EAN1pec</td>
<td>Parental wild type Novobiocin-resistant Tetracycline-sensitive</td>
<td>Lalonde et al., 1981; Tisa et al., 1999</td>
</tr>
<tr>
<td>EAN1pec/AD1</td>
<td><em>E. faecalis</em> mating transconjugant Tetracycline-resistant (20µg/ml)</td>
<td>Dwyer, unpublished</td>
</tr>
<tr>
<td>EAN1pec/AD2</td>
<td><em>E. faecalis</em> mating transconjugant Tetracycline-resistant (20µg/ml)</td>
<td>Dwyer, unpublished</td>
</tr>
<tr>
<td>Enterococcus faecalis GC110</td>
<td>Donor Novobiocin-sensitive Tetracycline-resistant</td>
<td>Gawron-Burke et al., 1982</td>
</tr>
</tbody>
</table>
### Table 2. Primers used in this study.

<table>
<thead>
<tr>
<th>Target</th>
<th>Primer</th>
<th>Sequence 5'-3'</th>
<th>Target Organism</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tn916</td>
<td>Tn1343-F</td>
<td>GGTGTCTTCGGATGGTG</td>
<td>E. faecalis</td>
<td>This Study</td>
</tr>
<tr>
<td></td>
<td>Tn1343-R</td>
<td>CGTCTTTTCGTTGGCTCTCT</td>
<td>E. faecalis</td>
<td>This Study</td>
</tr>
<tr>
<td>GlnII</td>
<td>FGgs19</td>
<td>TACATCTGGATCCACGGCAC</td>
<td>Frankia</td>
<td>Hahn et al. 1999</td>
</tr>
<tr>
<td></td>
<td>FGgs417</td>
<td>GCCGACCGCCGCACTAGTA</td>
<td>Frankia</td>
<td>Hahn et al. 1999</td>
</tr>
<tr>
<td>Conjugal</td>
<td>Ef-con</td>
<td>TATCGGGAGATACAGGTAACGGGA</td>
<td>E. faecalis</td>
<td>This Study</td>
</tr>
<tr>
<td>Transfer</td>
<td>transfer-F</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>Ef-con</td>
<td>CCACTTCATCGAAAACACCACGG</td>
<td>E. faecalis</td>
<td>This Study</td>
</tr>
<tr>
<td></td>
<td>transfer-R</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16srRNA</td>
<td>338F-B (GC-clamp)</td>
<td>ACTCTACGGAGGAGGAGGCAG</td>
<td>Hypervariable Region</td>
<td>Amann et al. 1990</td>
</tr>
<tr>
<td></td>
<td>907R</td>
<td>CGTCAATTCCTTTRAGTTT</td>
<td>Hypervariable Region</td>
<td>Lane et al. 1985</td>
</tr>
</tbody>
</table>

### Table 3. Transfer of Tn916 to *Frankia* Eul1c<sup>a</sup>. From Myers 2001.

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Transconjugant/donor</th>
<th>Transconjugant/recipients</th>
</tr>
</thead>
<tbody>
<tr>
<td>Liquid Mating</td>
<td>$4.2 \times 10^9$</td>
<td>$5.3 \times 10^5$</td>
</tr>
<tr>
<td>Filter Mating</td>
<td>$7.5 \times 10^9$</td>
<td>$1.5 \times 10^4$</td>
</tr>
</tbody>
</table>

<sup>a</sup> *Frankia* was mated.  
<sup>b</sup> Ratio of donor to recipient cells.
**Figure 1.** Photograph of recipient and transconjugant colonies on medium with NH₄Cl as the nitrogen source. Cultures were maintained at 30°C with appropriate carbon source for days. Panel A. Recipient wild type *Frankia* Eu11c in basal medium supplemented with 20 μg/ml novobiocin. Panel B. *Frankia* Eu11c transconjugant EET2 in basal medium supplemented with 20μg/ml tetracycline. Panel C. *Frankia* Eu11c transconjugant EET8 in basal medium supplemented with 20 μg/ml tetracycline.

**Figure 2.** PCR amplification of the transposon Tn916. Lanes: (1) Lambda-HindIII size marker, (2) donor *E. faecalis*, (3) *Frankia* transconjugant EET1, (4) *Frankia* transconjugant EET2, (5) *Frankia* transconjugant EET4, (6) *Frankia* transconjugant EET6, (7) *Frankia* transconjugant EET7, (8) *Frankia* transconjugant EET8, (9) *Frankia* transconjugant EET10, (10) recipient *Frankia* Eu11c (parental wild type), (11) negative control. The primers used in the PCR were Tn1343-F and Tn1343-R.
Figure 3. PCR amplification of the glutamine synthetase gene found in *Frankia* strains. Lanes: (1) 100bp ladder, (2) recipient *Frankia* Eu1c, (3) *Frankia* transconjugant EET1, (4) *Frankia* transconjugant EET2, (5) *Frankia* transconjugant EET4, (6) *Frankia* transconjugant EET6, (7) *Frankia* transconjugant EET7, (8) *Frankia* transconjugant EET8, (9) recipient *Frankia* EAN1pec, (10) *Frankia* transconjugant AD1, (11) *Frankia* transconjugant AD2, (12) donor *E. faecalis*, and (13) negative control. The primers used in PCR were FGgs19 and FGgs417.

Figure 4. PCR amplification of the conjugal transfer protein found in *E. faecalis*. Lanes: (1,11) 100bp ladder, (2) recipient *Frankia* Eu1c, (3) *Frankia* transconjugant EET1, (4) *Frankia* transconjugant EET2, (5) *Frankia* transconjugant EET4, (6) *Frankia* transconjugant EET6, (7) *Frankia* transconjugant EET7, (8) *Frankia* transconjugant EET8, (9) donor *E. faecalis*, and (10) negative control. The primers used in PCR were EF-contransfer-F and EF-contransfer-R.
Figure 5. DGGE of the 16S rRNA for recipient Frankia Eu1c and EAN1pec, transconjugants and E. faecalis. Lanes: (1) recipient Frankia Eu1c, (2) Frankia transconjugant EET1, (3) Frankia transconjugant EET2, (4) Frankia transconjugant EET3, (5) Frankia transconjugant EET6, (6) donor E. faecalis, (7) recipient Frankia EAN1pec, and (8) Frankia transconjugant AD1. The primers used in the PCR were 338F-B and 907-R.

Figure 6. PFGE of Frankia and transconjugants digested with restriction endonucleases. A. Digested with NdeI. Lane 1 Low Range PFGE marker, Lane 2 Eu1c, Lanes 3 EET2, Lane 4 EET8. B. Digested with Asel. Lane 1 Low Range PFGE marker, Lane 2 Eu1c and Lane 3 EET2.
Figure 7. Southern blot hybridization of *Frankia* and transconjugants digested with Scal. A. Gel Electrophoresis of genomic DNA digested with restriction enzyme Scal. Lanes (1) Lambda HindIII marker, (2) recipient *Frankia* Eu1c (parental wild-type), (3) *Frankia* transconjugant EET1, (4) *Frankia* transconjugant EET2, (5) *Frankia* transconjugant EET3, (6) *Frankia* transconjugant EET6, (7) *Frankia* transconjugant EET8, (8) *Frankia* transconjugant EET10, (9) donor *E. faecalis*, and (10) dig-labeled Lambda HindIII marker. B. Southern Blot Hybridization of genomic DNA digested with restriction enzyme Scal. Tn916 DIG-labeled probe was hybridized to membrane. Lanes are identical in order to the lanes in Figure 3A.
CHAPTER V

SUMMARY AND DISCUSSION

The completion of the experiments described in this study is imperative for the advancement in knowledge and understanding of this ecologically and economically important plant symbiont. It is an understatement to describe the latency of the genetics of *Frankia*, however with the completion of the sequencing of the genomes for three *Frankia* isolates the genetics of *Frankia* will now be able to be revealed. The creation of the physical map is a vital tool that has enabled me to understand the size and structure of this genome while the generation of a protocol for transposon mutagenesis has allowed for further genetic analysis and investigation of gene expression. The generation of these vital genetic tools provides an opportunity to further investigate all aspects of this symbiosis including infectivity, vesicle development, sporulation, and nitrogen fixation, as well as provide a means for comparative studies among other nitrogen fixing plant symbionts.

Due to the lack of genetic knowledge of the bacterium *Frankia* very little was known about the size, structure and organization of the genome. I was interested in developing a physical map of the *Frankia* chromosome by the use of macrorestriction analysis and pulsed-field gel electrophoresis (PFGE). The physical mapping in this study was completed for genomes of three *Frankia* strains (Eu1c, EAN1pec and Ccl3). Two of these strains (Eu1c and EAN1pec) represent one of the three major lineages of *Frankia*, while strain Ccl3 represents another lineage.
The host compatibilities of these strains differ greatly. Strain Ccl3 is restricted to *Casuarina* and *Allocasuarina* sp.; the host range of the other two strains differs greatly. PFGE analysis has allowed me to estimate the sizes of the genomes of *Frankia* strains Ccl3, EANIpec and Eu1c to be 5.4, 9.1, and 8.5 Mb, respectively. The extraordinary size discrepancy among the three strains was unexpected, however, it may be explained by their differences in host-plant specificities. Strain Ccl3 had the smallest genome; and it has a narrow host-range, infecting only members of the Casuarinaceae family. The two larger genomes were associated with the broad-host range *Frankia* strains. These two *Elaeagnus* isolates differed in size by only 1 Mb and may reflect a loss of genes involved in nodule function. Strain Eu1c infects and nodulates its host plant but results in an ineffective nodule that is unable to fix N₂. Through a collaboration project, we were able to sequence and compare the genomes of three isolates, including the narrow host range Casuarina strain, a medium host range Alnus strain, and the broad host range Elaeagnus strain. The results of the sequencing validated the results obtained through the physical mapping and again the most striking difference between the three genomes is their sizes, 5.4 Mb for Ccl3, to 7.5 Mb for ACN, to 9.04 Mb for EAN1pec. This size correlation raises the hypothesis that genome size and content is driven by the host range and biogeography of the symbiosis. The comparative analysis studies revealed that much of the size difference can be accounted for by gene rearrangement, gene duplication, and gene deletion.

Another surprising finding of the physical mapping study was that the *Frankia* genome has a circular topology, unlike many other actinomycetes that have large
genome sizes that have shown to have linear topology, such as the closely related
**Streptomyces**. Several lines of evidence support the circular topology of the **Frankia**
chromosome including PFGE. First, undigested **Frankia** DNA was unable to enter
the agarose gels while linear genomes readily migrated into the agarose gel.
Second, the absence of the terminal proteins covalently bound to the 5' ends of
linear chromosomes was confirmed by excluding the proteinase K treatment prior to
electrophoresis of the DNA and by attempting to visualize the terminal protein bound
DNA fragments that have been shown to exist on linear chromosomes. The
sequencing results of the three genomes confirmed reported circular topologies.

The genetic mapping of strain Eu11c in this study will aid in the efforts to
understand the genetic basis of actinorhizal symbiosis. Both I-Ceu1 digestion
results and Southern hybridization experiments indicate that **Frankia** Eu11c
chromosome harbors three **rrn** operons. The data obtained through macrorestriction
analysis and the sequencing data also suggests that **Frankia** EAN1pec and Ccl3
possess three and two **rrn** operons, respectively. The two **Elaeagnus** isolates harbor
three **rrn** operons compared with the two **rrn** operons found with the alder and
casuarina isolates. The additional **rrn** genes may be a reflection of the size
differences among isolates or provide a selective advantage to the organism under
various growth conditions.

The final approach used to understand the genetics of **Frankia** was
developing a transposon mutagenesis protocol to elucidate gene expression as well
as to help identify vital gene function. The transconjugants generated in this study
are stable and have been maintained in culture. Several lines of experimentation

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
were pursued to ensure the successful transformation of *Frankia* including PCR, PFGE, DGGE, as well as Southern blot hybridization. This study showed that transposon mutagenesis is an effective method for constructing *Frankia* mutants and for potentially identifying vital genes involved in sustaining this complex symbiosis.

My approach has focused on developing tools to allow a better understanding of both the genetics of this bacterium along with generating new knowledge about the genomics of this organism. This has been achieved through a number of approaches. First, the development of a physical map for *Frankia* isolates Eu11c, EAN1pec and Ccl3 and a partial genetic map of the *Frankia* chromosome Eu11c provided tremendous insight on the organization and structure of this complex genome and allowed for the preliminary data needed to advance the sequencing project of the three *Frankia* isolates. Second, the genome sequencing of *Frankia* isolates Ccl3, EAN1pec and ACN14a has provided an enormous amount of information to elucidate the complex genetic system of this bacterium. Lastly, confirming the feasibility of a transposon mutagenesis protocol is vital for future genetic analysis. The generation of these vital genetic tools presents an opportunity to further investigate all aspects of this symbiosis including infectivity, vesicle development, sporulation, and nitrogen fixation, as well as offer a means for comparative studies among other nitrogen fixing plant symbionts.

Prior to this study, the genetic and genomics of *Frankia* was largely unknown. The absence of standard genetic tools, such as no known system for gene transfer or a mutagenesis protocol, as well as no known phages, R-plasmids or transposable elements, hindered the study of this actinorhizal symbiosis. However, the advent of
genomic approaches has provided an alternative way to understand the symbiosis. Since, this study there has been funding available for the sequencing of additional \textit{Frankia} strains which will enable a comprehensive metagenomic study on a broad range of isolates in all host specificity groups. In addition, other '-omics' approaches are rapidly gaining acceptance and should be used to gain an overall understanding of this complex bacterium and its important symbiotic relationship.
LITERATURE CITED


model system for analysis of biological process. Horizon Scientific Press, Wymondham, UK.


Hopwood, D.A., Bibb, M.J., Chater, K.F., Kieser, T., Bruton, C.J., Kieser, H.M.,
manipulation of Streptomyces: a laboratory manual. John Innes Foundation, 
Norwich, England.


Ikeda, H., Ishikawa, J., Hanamoto, A., Shinose, M., Kikuchi, H., Shiba, T.,
Sakaki, Y., Hattori, M., and Omura, S. 2003. Complete genome sequence and 
comparative analysis of the industrial microorganism Streptomyces avermitilis. 

Kimura, M. 1980. A simple method for estimating evolutionary rates of base 
16: 111-120.

Konstantinidis, K.T. and Tiedje, J.M. 2005. Genomic insights that advance the 

Practical Streptomyces Genetics, Norwich, United Kingdom: John Innes 
Foundation.

map of the Streptomyces coelicolor A3(2) chromosome. J. Bacteriol. 174:5496-
5507.


Lalonde M, Knowles R. 1975. Ultrastructure, composition, and biogenesis of the 
encapsulation material surrounding the endophyte in Alnus crispa var. mollis root 


Electrophoresis, Ribosomal DNA Sequencing and Real Time PCR. PhD. Dissertation. Durham, NH. University of New Hampshire.


APPENDIX

FRANKIA BAC LIBRARY

My original thesis proposal stated that the physical map would be used to facilitate the ordering of a *Frankia* Bacterial Artificial Chromosome (BAC) library. The construction of a BAC library for *Frankia* Ccl3 was initiated; this was to be achieved by partially digesting high molecular weight DNA with a restriction endonuclease. After separation by PFGE, DNA fragments in the 50kb to 100kb range were to be purified and cloned into a BAC vector. The BAC vector is a low copy number plasmid, which provides stability and reduces the potential for recombination between cloned fragments. This vector has the potential to maintain up to 300kb fragments of cloned DNA, therefore allowing a relatively small number of clones to provide complete coverage of the genome. The BAC vector (pBeloBAC11) has three unique cloning sites, HindIII, BamHI, and SphI within the lacZ gene (Figure 1), allowing lacZ-based positive color selection of the clones containing inserted DNA. Since the physical map determined *Frankia* isolate Ccl3 to have a genome size of 5.4Mb, 1000 clones with an average BAC insert size of 80kb would provide 15X genome coverage.

In order to construct the BAC library, DNA agarose plugs were generated as previously described (Chapter 2). The DNA agarose plugs were partially digested with a restriction enzyme and the fragments were separated by PFGE. The
restriction enzyme Sau3AI was used to digest the DNA and several incubation times were tested (Figure 2). To obtain DNA fragments in the size range of 50 kb to 100 kb, the DNA plugs were digested with 4 units of enzyme for 10 min. The initial partial digest experiments were successful and provided digested DNA within the size range for cloning. However, there was numerous problems during the excision step. One common problem with performing molecular techniques with Frankia is the difficulty in obtaining sufficient quantities of the high-molecular-weight DNA. The slow growth rate (30-80h doubling time) and low biomass affect the yield of high quality genomic DNA. I was successful in obtaining sufficient yields of high-quality high-molecular-weight gDNA for restriction digest, but the amount of digested DNA excised out of the gel was insufficient for cloning into the BAC vector. These studies were continued in collaboration with the Hubbard Genomic Center at the University of New Hampshire, but were also unsuccessful. To aid in the sequencing of the Frankia Ccl3 and EAN1pec genomes, a fosmid library (40kb inserts) were generated to scaffold the sequence information from the shot-gun cloned small (2-3kb) and large (8-10kb) insert libraries (Chapter 3). Thus, this portion of the thesis project was terminated and is presented to guide future work.
Figure A1. Cloning vector pBelOBAc11.
Source: www.genome.clemson.edu/groups/bac.html.

Figure A2. Partial digest of Frankia isolate Ccl3 by restriction enzyme Sau3AI run on a 1.0% gel, 5.5 V/cm, 5-20sec switch time for 16.5 h. Lane (1) Ccl3 digested by 4U of Sau3AI for 5 min. Lane (2) digested for 10 min. Lane (3) digested for 15 min. Lane (4) digested for 20 min. Lane (5) digested for 30 min. Lane (6) marker.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.