Phylogenomic study and specific diversity depiction of frankia genus : special focus on non-cultivable strains and ecological implications

Hector Hugo Bautista Guerrero

N° d'ordre Année 2010

THESE DE L'UNIVERSITE DE LYON

Presentée devant

L'UNIVERSITE CLAUDE BERNARD LYON 1

Pour l'obtention du

DIPLOME DE DOCTORAT

(Arrêté du 7 août 2006)

Soutenue publiquement le 1er juillet 2010

par

M. Hector Hugo BAUTISTA GUERRERO

PHYLOGENOMIC STUDY AND SPECIFIC DIVERSITY DEPICTION OF FRANKIA GENUS. SPECIAL FOCUS ON NON-CULTIVABLE STRAINS AND ECOLOGICAL IMPLICATIONS

Directeurs de thèse : Pr. Maria P. FERNANDEZ & Pr. Maria VALDES

JURY :

M. D Bogusz	Directeur de Recherches, IRD, Montpellier	Rapporteur
M. P. DELAJUDIE,	Directeur de Recherches, IRD, Montpellier	Rapporteur
Mme M. TRUJILLO,	Directeur de Recherches, Universidad de Salamanca	Rapporteur
M. V. DAUBIN,	Directeur de Recherches, CNRS, Université Lyon1	Examinateur
Mme M. P. FERNANDEZ,	Professeur Université Lyon 1	Examinateur
Mme M. VALDES,	Professeur ENCB – IPN	Examinateur

UNIVERSITE CLAUDE BERNARD - LYON 1

Président de l'Université	**M. le Professeur L. Collet**
Vice-président du Conseil Scientifique	M. le Professeur J-F. Mornex
Vice-président du Conseil d'Administration	M. le Professeur G. Annat
Vice-président du Conseil des Etudes et de la Vie Universitaire	M. le Professeur D. Simon
Secrétaire Général	M. G. Gay

COMPOSANTES SANTE

Faculté de Médecine Lyon Est – Claude Bernard	Directeur : M. le Professeur J. Etienne
Faculté de Médecine Lyon Sud – Charles Mérieux	Directeur : M. le Professeur F-N. Gilly
UFR d'Odontologie	Directeur : M. le Professeur D. Bourgeois
Institut des Sciences Pharmaceutiques et Biologiques	Directeur : M. le Professeur F. Locher
Institut des Sciences et Techniques de Réadaptation	Directeur : M. le Professeur Y. Matillon
Département de Biologie Humaine	Directeur : M. le Professeur P. Farge

COMPOSANTES ET DEPARTEMENTS DE SCIENCES ET TECHNOLOGIE

Faculté des Sciences et Technologies	Directeur : M. le Professeur F. Gieres
Département Biologie	Directeur : M. le Professeur C. Gautier
Département Chimie Biochimie	Directeur : Mme le Professeur H. Parrot
Département GEP	Directeur : M. N. Siauve
Département Informatique	Directeur : M. le Professeur S. Akkouche
Département Mathématiques	Directeur : M. le Professeur A. Goldman
Département Mécanique	Directeur : M. le Professeur H. Ben Hadid
Département Physique	Directeur : Mme S. Fleck
Département Sciences de la Terre	Directeur : M. le Professeur P. Hantzpergue
UFR Sciences et Techniques des Activités Physiques et Sportives	Directeur : M. C. Collignon
Observatoire de Lyon	Directeur : M. B. Guiderdoni
Ecole Polytechnique Universitaire de Lyon 1	Directeur : M. le Professeur J. Lieto
Institut Universitaire de Technologie de Lyon 1	Directeur : M. le Professeur C. Coulet
Institut de Science Financière et d'Assurance	Directeur : M. le Professeur J-C. Augros
Institut Universitaire de Formation des Maîtres	Directeur : M R. Bernard

ABSTRACT

The depiction of the phylogenetic structure of the genus *Frankia* is still troublesome and the evolutionary forces guiding the speciation, dispersion and diversity are not well documented. The current phylogeny has been defined on the basis of the comparative analysis of the 16S rRNA gene sequence while de genomospecies definition is still subjected to DNA-DNA hybridization trials.

Aiming to bring to light the genomic variability of the genus and its translation into the ecological and specific diversity, our studies consisted in, firstly, evaluating the specific diversity within the genus and the ability of the Amplified Fragment Length Polymorphism technique (AFLP) to describe *Frankia* genomospecies and their phylogenetic liaisons. Moreover this technique was also tested for the study of the non isolated *Frankia* directly in the actinorhizal nodules. Secondly, we defined a MLSA (Multilocus Sequence analysis) scheme which allowed us to establish a phylogeny of the genus by using a hundred of strains and for the first time to describe the phylogenetic divergence of a group of non culturable strains exhibiting the particular ability (phenotype) of sporulating *in planta* (Sp+). The Sp+ strains are distributed into two divergent clades whose structure is highly correlated to the host genotype.

The importance of genetic markers having impact over ecology of the strains has been revised. In this regard we have studied the phylogenetic analysis and the occurrence of the genetic components for the siderophore production and of the *sodF* gene in *Frankia*.

RESUME

La définition de la structure phylogénétique du genre *Frankia* est encore problématique, les forces évolutives guidant son spéciation, dispersion et donc la génération de sa diversité ne sont pas complètement documentées. La phylogénie actuelle du genre a été définie par l'analyse comparative de la séquence du 16S rRNA. Par ailleurs, la définition des espèces génomiques a été gênée par la faible applicabilité de la technique d'hybridation ADN-ADN.

Dans le cadre de cette thèse nos travaux on consisté à étudier la variabilité génomique dans le genre et sa conséquente traduction en variabilité spécifique et écologique. Dans un premier temps, nous avons évalué la diversité spécifique du genre ainsi que l'utilité de la technique AFLP (Amplified Fragment Length Polymorphism) pour la définition des espèces génomiques. De plus, notre protocole fut aussi utilisé pour analyser souches non isolées en appliquant le protocole directement sur des nodosités actinorhiziennes. Dans un deuxième temps, un schéma MLSA (Multilocus Sequence Analysis) nous a permit de redéfinir la phylogénie du genre sur une centaine de souches, et pour la première fois de décrire la divergence phylogénétique d'un groupe de souches non-isolées présentant un phénotype unique de sporulation *in planta* (Sp+). Les souches Sp+ sont distribuées dans deux clades très divergents dont la structuration est fortement corrélée au génotype de la plante hôte et au phénotype Sp+/Sp- de la souche.

L'intérêt de marqueurs génétiques présentant un intérêt pour l'écologie des souches a été révisé. Dans ce but nous avons étudié la présence, distribution et phylogénie de *sodF* et des différents composants génétiques impliquées dans la production des siderophores chez *Frankia*.

RESUMEN

La estructura filogenética del genero *Frankia* aun no es clara, las fuerzas evolutivas que han guiado los procesos de especiación, dispersión y en consecuencia la generación de la diversidad dentro del género no ha podido ser clarificados. La filogenia actual del genero ha sido definida con base en el análisis comparativo de la secuencia del gen 16S rRNA. Por otro lado, la definición de especies genómicas ha sido obstaculizada por las diversas restricciones que impone la técnica de hibridación ADN-ADN. Los trabajos comprendidos en esta tesis doctoral se focalizaron en el estudio de la variabilidad genómica del género y su consecuente traducción en variabilidad específica y ecológica. En primer lugar, evaluamos la diversidad específica del género y la aplicabilidad de la técnica AFLP (Amplified Fragment Length Polymorphism) para definir especies genómicas. En seguida, el protocolo establecido fue utilizado para el análisis de cepas no cultivables directamente en nódulos actinorrizicos. Posteriormente, un análisis multilocus, MLSA (Multilocus Sequence Analysis) nos permitió reconstruir la filogenia del genero, para tal efecto, una centena de cepas fueron analizadas. Dicho estudio nos permitió por la primera vez describir la divergencia filogenética de un grupo de cepas no cultivables, que presentan como fenotipo característico, la producción de esporangios in planta (llamadas en consecuencia "Sp+"). Las cepas Sp+ aparecen como dos grupos genómicos divergentes entre si y de las cepas Sp-, la estructuración de cada grupo está relacionada al genotipo de la planta hospedero.

Finalmente, genes con relevancia ecológica fueron elegidos y empleados como potenciales marcadores filogenéticos de *Frankia*. Con ese objetivo fijado, estudiamos la presencia y la distribución del gene de la superoxido dismutasa (*sodF*) asi como los diferentes elementos genéticos implicados en la producción de sideroforos dentro del género.

Acknowledgements

It is a pleasure to thank the many people who made this thesis possible.

It is difficult to overstate my gratitude to my Ph. D supervisors, Mrs. Maria Valdés and Mrs. Maria Fernandez. They first taught me the rigourousness and the meaning of a scientific work. Throughout my thesis, they provided me good teaching, a lot of good ideas, soud advice and their invaluable expertise. They also provided me encouragement and all the support I have needed to carry out my work.

I also like to express my gratitude to Mr. René Bally and Mr. Philippe Normand that kindly received me at the research unit "UMR5557 Ecologie Microbienne" and in the Group 1 "Actinorhizal Symbiosis", respectively.

I am deeply grateful to the National Council of Science and Technology in Mexico (CONACYT) for the trust and support they gave me in order to make my Ph. D studies.

I would like to thank to Mrs. Martha Trujillo, Mr. Philippe DeLajudie, and Mr. Didier Bogusz who kindly accepted to make the report of my dissertation. In addition I would like to thank to Mr. Vincent Daubin who examinate this work and contributed to its improvement with his assertive advices.

I wish to thank to Mr. Philippe Normand, Mr Xavier Nesme and to all of the people whose thoughtful advice often served to give a sense of direction to my research.

I am indebted to my many laboratory student and collegues for providing a stimulating and fun environment in wich to learn, work and grow. Thanks to Group 1 people especially to Aude Herrera who with humour and kindness gave me support and encouraged me. I would thank also our collaborators from the laboratory of Vegetal Biology 1 of the "Centro de Biotecnologia Genomica", Reynosa Mex, for their invaluable technical support.

It is time to thank to all of my friends, for helping me get through the difficult times, for all emotional support, comraderie, entertainment and caring they provided. I will not tempt to make a list of them but because of fear to leave someone out, I am just going to say THANK YOU VERY MUCH TO YOU ALL.

Someone of you are, anyway, lucky…Thanks to Daniela, Nuno, Jan, Jesus, Heber, Jean Baptiste, Yannick, Leonardo, Claudia…for being a cornerstone in my life.

I wish to thank my entire extended family for providing a loving environment for me. Lastly, and most importantly, I wish to thank my mother, Cristina because she always

represented the strenghtness, the support and the love for me. I am thankful to this amazing woman and to her I dedicate this work.

TABLE OF CONTENTS

GENERAL INTRODUCTION

The actinorhizal symbiosis is a nitrogen-fixing interaction established between a group of plants collectively called Actinorhizal Plants (AP) and the actinomycetes from the genus *Frankia* (Benson & Silvester, 1993). The actinorhizal symbiosis contributes importantly in the fixation of atmospheric nitrogen in terrestrial ecosystems (Paschke *et al.*, 1989). Thanks to their association with *Frankia*, the actinorhizal plants are able to grow under adverse environmental conditions (high salinity, acidic soils…) thus acting as pioneers in disturbed soils and enhancing vegetal succession in these ecosystems (Dawson, 2007a).

Mutualistic interactions between species underpin much of nature's biodiversity and the actinorhizal symbiosis is not the exception. In the past years, several studies have been conducted to increase the understanding of the processes implicated in the symbiosis establishment and the specificity and efficiency of host-microsymbiont interactions. In this regard the actinorhizal symbiosis exhibits a high range of diversity, adaptation and specificity. Different actinorhizal models can vary in: a) their specificity, ranging from very specific, i.e. *Casuarina*- effective *Frankia* symbiosis (Simonet *et al.*, 1999), to promiscuous interactions i.e. *Myrica*- *Frankia* symbionts (Huguet *et al.*, 2005a), b) their ecology, because the actinorhizal plants are widely distributed inhabiting a wide range of different ecosystems (Dawson, 2007a).

The genus *Frankia* comprises a group of morphologically homogeneous diazotrophic actinobacteria capable to establish a symbiotic association with the roots of the actinorhizal plants. Also classified within the group of fastidious organisms, *Frankia* exhibits very long duplication times, low yields of biomass production and attempts for its genetic transformation have failed systematically (Hahn, 2007). Moreover, *Frankia* infective of some actinorhizal plant-genus are still refractory to *in vitro* culturing. Consequently a global picture of genetic and ecological diversity of *Frankia* remains elusive. Nevertheless its depiction is critical for the understanding of strain adaptation to different environments and hosts as well as the overall evolution of the genus.

The current study of microbial diversity implies the inclusion of the microorganisms into arbitrarily defined clusters, species (Stackebrandt *et al.*, 2002b). However, bacterial species definition is a difficult task, and we still do not have a clear definition of the concept. The experts are struggling to unify the different theories in a single species concept, thus the bacterial taxonomists have made propositions without achieve the consensus (Cohan & Perry,

2007; Fraser *et al.*, 2010; Gevers *et al.*, 2005a; Staley, 2006). Shortly, if bacterial species do exist they have to exhibit (at minimum) the following properties : a) Each species must be a cohesive group, whose diversity is limited by an evolutionary force b) the different species are irreversibly separated from the others c) they are ecologically distinct and d) they are found only once in the nature (Cohan & Koeppel, 2008; de Queiroz, 2005).

Microbial species are currently defined in a pragmatic way, their depiction being based on phenotypic, genotypic and more recently, genomic traits. Actually, an arbitrary cut off level of 70% of pair wise genomic DNA-DNA hybridization (DDH) together with ΔTm values < 3°C between two individual strains is the main parameter that determines they belong to the same genomospecies (Wayne *et al.*, 1987). Such cut off value is not based on particular theoretical justification but on statistical consistencies between published DDH and ΔTm values and phenotypic traits (physiological and morphological) used for defining species at that time. Consequently the groups to date reported are not consistent with a theoretical species concept and nor always in accordance with phylogeny; instead their definition is guided by the methodologies applied to identify them (Achtman & Wagner, 2008).

Nevertheless the DNA-DNA hybridization technique is still considered as the standard for bacterial genomospecies definition since it permits a comparison of the relatedness of all the genetic features among strains (Stackebrandt *et al.*, 2002b). The applicability of this technique is highly restrictive though, the pair wise comparison of individual strains is not necessarily symmetrical and the differences in genome size can affect the results. Also, it is difficult to standardize, it is time consuming, the number of pair wise combinations increases exponentially with respect to the number of strains, the data cannot be accumulated and finally non cultivated strains are excluded from the analysis (Achtman & Wagner, 2008; Gevers *et al.*, 2005a).

A limited number of studies have conducted the genomospecies identification within the genus *Frankia*. To date about twenty genomospecies are reliably delineated among the collections of *Frankia* isolates, in addition to several others to be confirmed (Akimov & Dobritsa, 1992a; An *et al.*, 1985a; Fernandez *et al.*, 1989a). In most cases, a low number of strains were considered for the experimental hybridization schemes with minimal or no correspondence between the different sets of strains studied, reflecting a lack of agreement between the experts in the management of the *Frankia* genomospecies problematic. As a consequence, no consensus exists on *Frankia* taxonomy, i.e. the limits and name of the hitherto described genomospecies.

The study of *Frankia*'s specific diversity by DNA-DNA hybridization is challenging either because of the drawbacks inherent to the technique itself and by *Frankia'*s culture and biomass production difficulties. In general, the applicability of DDH is restricted to isolated specimens, thus avoiding the genomospecies description within the group of non cultivable *Frankia* strains. Alternative phenotypic and molecular techniques have let to refine and expand the knowledge about *Frankia* phylogeny and diversity at different taxonomic levels. Nonetheless, any of them seems able to replace the DDH.

Since genomospecies definition is a problem concerning the microbiologist en general, the committee for the re-evaluation of species definitions in bacteriology made a reconciliation of the bacterial systematics. Some of the methods listed as promising alternatives are the DNA profiling approaches, the sequencing of housekeeping genes and DNA arrays. In all cases, sufficient degree of congruence with DNA-DNA reassociation must be demonstrated, whatever the alternative technique chosen (Stackebrandt *et al.*, 2002b).

In this work, we wanted to revisit the *Frankia* genetic diversity and genus phylogeny in such a way to give strong evolutionary arguments for a further base of *Frankia* strain classification and taxonomy. We aimed also to give novel insights on the evolutionary frame for better understanding of *Frankia* ecologic distribution and symbiotic specificity.

For this purpose, an extensive sampling of representative reference strains and field collected nodules were studied by means of two promising approaches proposed by the ad hoc committee: 1) The AFLP (Amplified Fragment Length Polymorphism) and 2) the MLSA (Multilocus Sequence Analysis). The success of such techniques in the phylogenetic redefinition of other bacterial genus encouraged us to hypothesize such techniques can reliably replace the DDH in *Frankia* genomospecies definition. The results of the experimental work are spread in three chapters (Chapter I) which analyses the current state of *Frankia* phylogeny and genomospecies definition. (Chapters II, III and IV) are described as follows:

Chapter II: *Frankia* PHYLOGENY DEPICTION AND SPECIES DEFINITION BY AMPLIFIED FRAGMENT LENGTH POLYMORPHISM (AFLP) AND MULTILOCUS SEQUENCE ANALYSIS (MLSA)

The first part of this chapter comprises the study of the genomic variability within a collection of *Frankia* reference strains by means of the AFLP genomic approach. We stated

its reliability and robustness in strain discrimination and moreover in genomospecies definition. The applicability of AFLP to study non-culturable strains was also tested and debated; that is to our knowledge the first time to be reported. In a second part we selected five housekeeping genes spread throughout *Frankia*'s genome to set up a multigenic phylogenetic approach (MLSA). We tested the capacity of this technique to provide a criterion for defining the extent of a microbial species. Moreover we also seek to establish the phylogenetic liaisons between the different clusters and its organization within the genus.

Chapter III THE *Alnus*-EFFECTIVE Sp+ *Frankia*e FORM A DIVERGENT PHYLOGENETIC GROUP WITH CLONAL STRUCTURE

In this chapter we studied the phylogenetic, taxonomical placement and ecological significance of a particular group of non cultivated strains which possesses the original capacity of sporulating *in planta* (Sp+ phenotype) by using three species within *Alnus* genus as models and a MLSA approach.

Chapter IV PHYLOGENETIC ANALYSIS OF ECOLOGICALLY RELEVANT MARKERS

In this chapter instead of analyzing neutral genes our study targeted genes potentially involved in the symbiosis (adaptation to the plant) or to different environments and their potential utility as phylogenetic and ecological markers. Firstly, the *sod*F gene, which plays a key role in the oxidative stress defense and that is likely involved at the initiation stage of actinorhizal symbiosis was used to construct a phylogeny of the genus *Frankia*. Secondly, we determined the occurrence of some genetic determinants for siderophore biosynthesis among the genus, their phylogeny and their expression. The latter work completed the study about siderophore biosynthesis occurrence in *Frankia* started by E. Bagnarol (PhD Thesis, 2007).

CHAPTER I: CURRENT TAXONOMIC CLASSIFICATION AND SPECIES DEFINITION IN *Frankia*

Chapter I: Current taxonomic classification and species definition in *Frankia*

Introduction

The occurrence of *Frankia* and its interaction with woody plants was reported at the end of the XIX century. The first successful *in vitro* isolation was achieved in 1978 (Callaham *et al.*, 1978), since then, empirical synthetic or semi-rich media has been employed in the isolation of these actinomycetes. No general protocols for isolation have been established and often only a small percentage of attempts succeed. After a successful isolation, the culture maintenance is still a hard task, and high yield of biomass are not always obtained. Such difficulties in *Frankia* isolation and *in vitro* culture in addition to the very slow growth rates have hindered the studies of host-specificity, strain diversity and *Frankia* taxonomic classification. To date *Frankia* isolates have been obtained from a small portion of the almost 280 angiosperm nodulated species known, and therefore a high proportion of the richness and genetic diversity of these microsymbionts is not known (Benson & Dawson, 2007a).

Frankia genus comprises a group of actinomycetes exhibiting the formation of septate, branching hyphae, the production of multilocular sporangia, the presence of non-motile spores in multilocular sporangia and the production of thick-walled, lipid-encapsulated structures called diazovesicles, which are the site of the nitrogenase (Fig. 1) (Lechevalier, 1984a). The cells are characterized by a type-III cell wall, a type PI phospholipids pattern, large amounts of hopanoid lipids relative to the total lipid content and 2-O-methyl-D-mannose as a diagnostic sugar (Lechevalier, 1994). The high G+C genomic (66-77% mol) as well as the ability to fix nitrogen *in vitro* and in symbiosis are together additional criteria to define the genus. The current taxonomical classification of *Frankia* is based upon a polyphasic approach and is thus the result of the compilation of morphological, cytochemical, physiological and genetic traits.

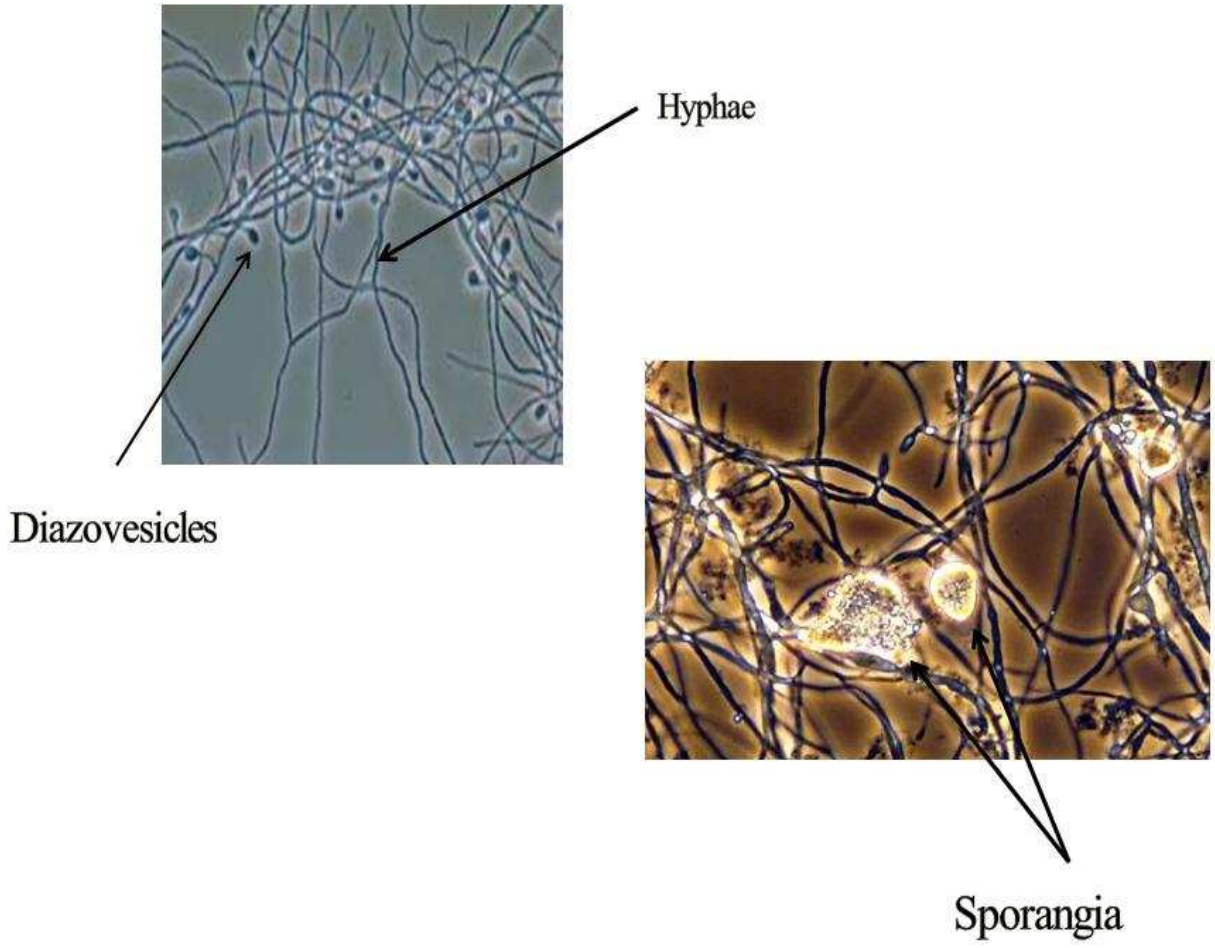

Figure1. *Frankia* characteristic morphology: a) Vegetative hyphae (DAPI /Sytox Green) modified from: http://web.uconn.edu/.../benson/Frankia/FrankiaHome.htm

b) *Frankia* sporangia (phase contrast image modified) from: http://web.uconn.edu/mcbstaff/benson/Frankia/FrankiaTaxonomy.htm

I.1. Host Specificity Groups

In a pioneer study, crushed nodules collected from various AP were employed in a cross inoculation experiment. The specific-host plant infection groups obtained allowed the definition of ten *Frankia* species (Becking, 1970a). The proposed classification was rapidly refused by Baker (1987), who performed another cross inoculation scheme but this time using pure cultures and a larger range of host plants. The yielded results conduct him to gather the strains into discrete clusters, called Host Specificity Groups (HSG) (Fig. 2). The host specificity is a criteria widely determined for other symbiotic bacteria i.e. Rhizobiaceae and can give insights into the strain specialization. In *Frankia*, the degree of host specificity is variable since some strains establish high specific interactions while others are versatile and infect a large spectrum of host plants (Bosco *et al.*, 1992). Conversely, there are host plants which can be infected by different groups of strains, as is the case of the promiscuous host *Myrica* (Huguet *et al.*, 2005a). Hence cross-inoculation groups do not always correlate with the origin of the plants and a classification based on this character can be misleading. Moreover all these studies were likely biased because only a reduced group of host plants and *Frankia* specimens were employed. To summarize, the HSG cannot be considered as a taxonomic criterion itself, nonetheless it gives insights of the versatility, ecology and genomic expression of *Frankia* strains; therefore it is strongly recommended to determine it whenever it is possible.

I.2. Current phylogeny of Frankia genus

I.2.1. Phenotypic traits

Several studies focused to unveil the diversity, within *Frankia* strain collections by analyzing phenotypic traits as. In this regard, total proteins patterns, sugar content, fatty acid composition, isoenzyme patterns and antibiotic susceptibility or resistance are some of the phenotypic traits already analyzed in *Franki*a strains (Benson *et al.*, 1984; Dobritsa, 1998b; Gardes *et al.*, 1987; Lechevalier *et al.*, 1981; St-Laurent *et al.*, 1987). An additional paramenter -the nodulation speed- was also proposed as taxonomic criterion since it was positively correlated with the *Frankia* strain phenotype (Nesme *et al.*, 1985).

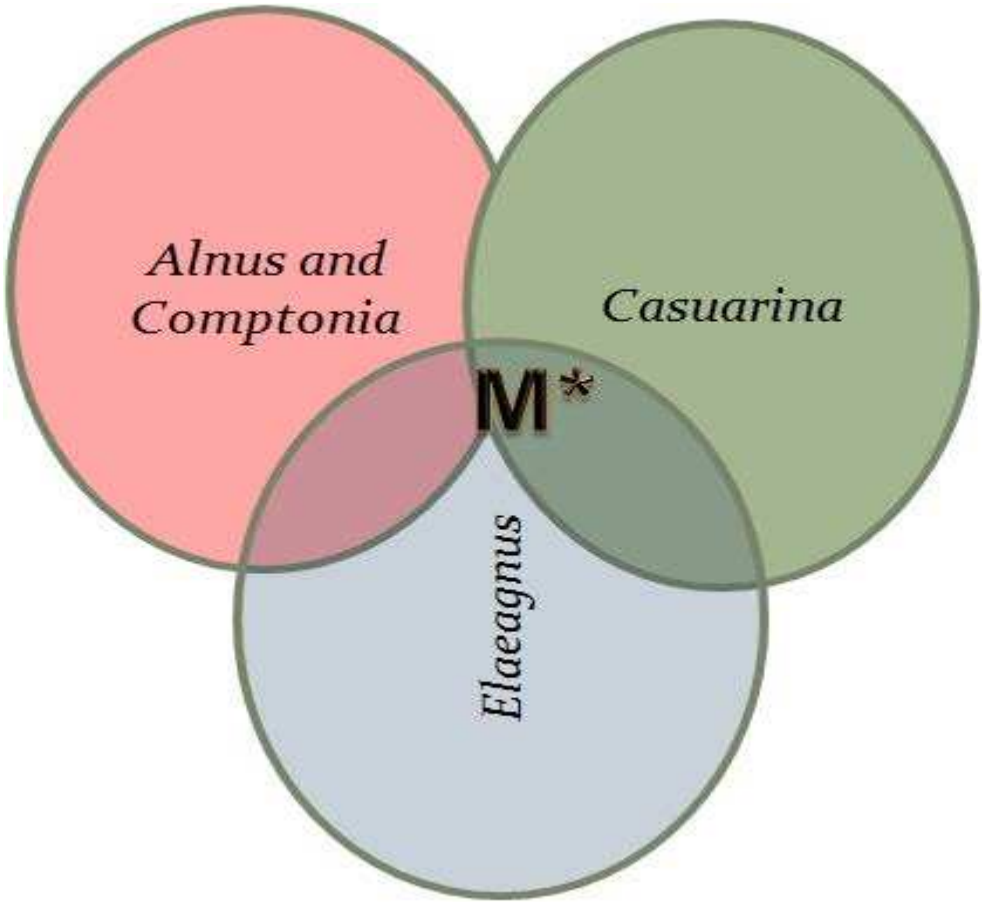

Fig 2. Classification of *Frankia* strains based on the Host Specificity Groups. Scheme, as described by Baker (1987). * Indicates the promiscuous group of *Myrica* strains.

Most of the time, *Frankia* strains grouping by means of phenotypic data was in agreement with the HSG groups. The different methods discriminate the strains at variable extent. Consequently, the usefulness of morphological and physiological features for the discrimination among *Frankia* strains is limited (Hahn, 2008)

I.2.2. Molecular classification

The PCR technique has proven to be a powerful tool to analyze several genetic markers; there is an extensive list of oligonucleotides which permit to target partial and complete sequences of different genes (genetic markers) in isolated strain but also in the uncultured frankiae inhabiting actinorhizal nodules (Hahn *et al.*, 1999). The amount of evolutionary information contained in a single genetic region is variable i.e. the protein-coding genes are conserved sequences in contrast to variable intergenic regions (IGS) where mutation accumulation is likely frequent. The choice of the marker to target will depend on the degree of taxonomic discrimination desired in each study.

The characteristics of the 16S rRNA gene such as its essential function, its ubiquity and its evolutionary information content, have made this region to become the most commonly used molecular marker in bacterial taxonomy and for phylogeny constructions (Stackebrandt & Goebel, 1994). By utilizing the total 16S rRNA sequences, Normand *et al* (1996) were able to propose the emendation of the family *Frankiaceae* and to depict the molecular phylogeny for the genus *Frankia*. Such emendation implied the expulsion of the related taxa *Geodermatophilus* and *Blastococcus*, leaving the genus *Frankia* as the sole member in this family. Conversely, it was found that the closest neighbour of this family is *Acidothermus cellulolyticus,* a bacteria isolated from 55-60°C acidic water and mud samples collected in Yellowstone National Park. The figure 3 shows the current phylogenetic depiction of the genus *Frankia* based upon the comparative analysis of the 16S-rRNA sequence. Four main clusters can be clearly delimited within the genus:

Cluster 1: Comprises the strains infective on the *Myricaceae* (*Comptonia*, *Myrica* and *Morella* genera) as well as the *Casuarinaceae*-compatible strain and those compatible with *Alnus* sp.

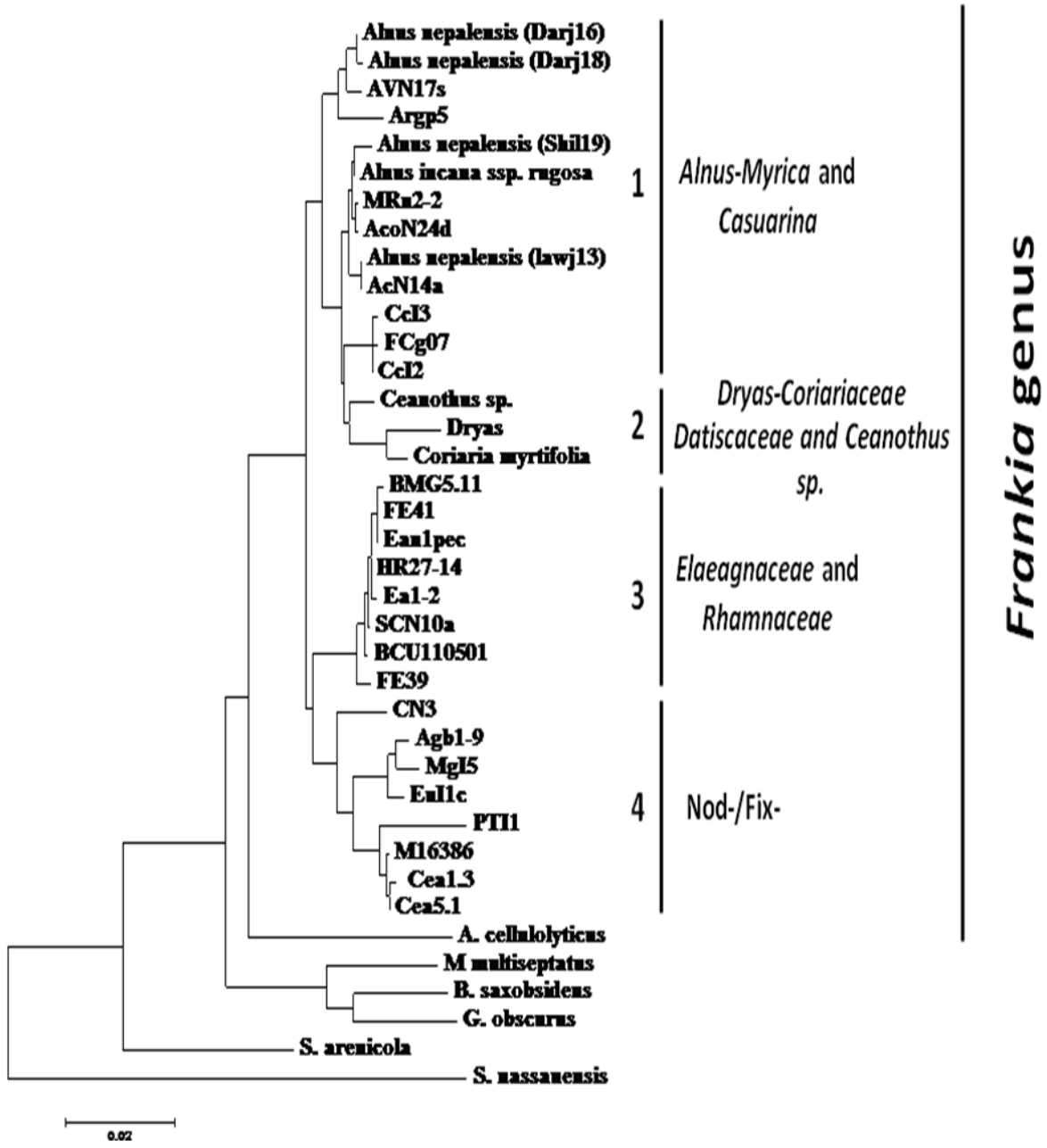

Fig.3 *Frankia* genus phylogeny based on almost complete 16S rRNA sequences. The dendrogram was constructed with the Neighbour-joining method. Numbers indicates the designation of clusters

Cluster 2: Encompasses the uncultured endophytes of *Dryas*, *Coriaria* and *Datisca* nodules

Cluster 3: Includes typical strains of *Elaeagnaceae* (*Elaeagnus*, *Hippophae* and *Shepherdia*) and seven genera of *Rhamnaceae*.

Cluster 4: Comprises strain isolated from different AP which shared the incapability to fix nitrogen and to renodulate their original hosts; they are collectively called "Nod^{-} Fix^{-}" strains.

Subsequent studies based in the 16S rRNA sequence and applied to different strain and nodule sets have contributed to the depiction of a more complete phylogenetic frame of *Frankia* genomic diversity. In most cases there is a relative correspondence in terms of strains and the HSG. In this regard some exceptions have been founded for instance, the strains infecting *Gymnostoma* (a *Casuarinaceae* member) grouped into the cluster 3 (Clawson *et al.*, 1998; Navarro *et al.*, 1997). The second striking exception is the placement of the uncultured endophytes of *Ceanothus americanus* and *C. thyrsiflorus* (*Rhamnaceae* members) in the cluster 2, close to the *Dryas drummondi* microsymbionts rather than with *Frankia* isolated from *Rhamnaceae* (Benson *et al.*, 1996).

The topology depicted by 16S suggests that Cluster 1 and 3 are sister groups while clusters 2 and 4 take more divergent positions. Indeed the relative position of the latter groups is not well defined and it can varies according to outgroup used to construct the tree (data not shown). Moreover when the 16S rRNA sequences of novel isolates are added and the phylogenetic analysis is performed with more sophisticated methods, the major clusters are still found but their placement within the topology changes (data not shown) leading to the conclusion that sampling bias may conduct to erroneous depiction of the phylogeny.

The *Frankia* phylogeny have also been studied by the comparative analysis of others protein-coding genes i.e. two symbiosis-related markers, *nifH* (Baker & Mullin, 1994; Normand *et al.*, 1992; Welsh *et al.*, 2009a) *glnII* (Cournoyer & Lavire, 1999), and the housekeeping *recA* (Marechal *et al.*, 2000) and *glnA* (Clawson *et al.*, 2004a). The retrieved phylogenies supported the clustering of strains into major groups; however they failed to reliable support the clustering of strains below the genus level, the figure 4 represents the topological comparison between the phylogenies depicted with three different genetic markers.

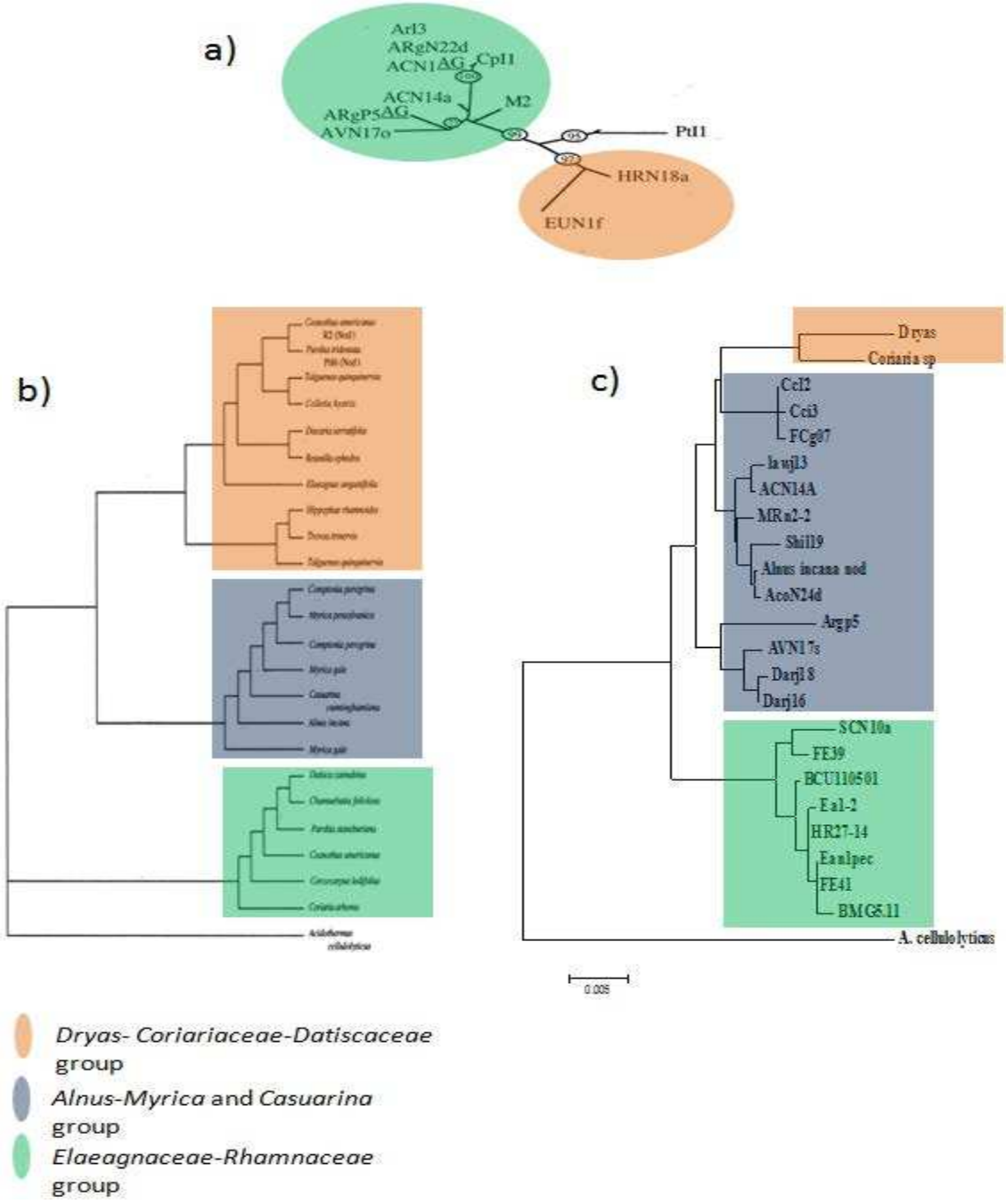

Fig. 4 Comparison between the topologies of *Frankia* phylogeny based on three different genetic markers: a) Radial dendrogram based on glnII sequence (modified from Cournoyer, B and Lavire C, 1999) b) Neighbour-joining dendrogram based on the glnA sequence (modified from Clawson, M and Benson D, 2004) c) Neighbour-joinin dendrogram based on the 16S rRNA complete sequence.

I.3. How many species exist within the Frankia genus?

Since the first *in vitro* isolation a large collection of *Frankia* strains have been obtained from different host plants, without their classification into species. The numerical phenetic and chemotaxonomic tests are no consistently useful criteria for proper classification of *Frankia* strains because the analysis of such characters is exceptionally laborious and time consuming, mainly if we take into account the slow growth rate of *Frankia* and its lack of aerial mycelium (Dobritsa, 1998b). *Frankia* growth on agar plates is limited, and usually characterized by the formation of microcolonies that are difficult to observe to the unaided eye.

Moreover, the filamentous morphology hinders the objective estimation of biomass production, which is the first parameter quantified when testing carbon sources utilization or antibiotic sensibility. Attempts to overcome the difficulties in determining the biomass production of *Frankia* cultures have consisted in the quantification of total protein content and other indirect parameters. However, the results are not always accurate (Dobritsa, 1998b; Harriott & Bourret, 2003).

Since polyphasic taxonomy approaches integrate phenotypic, genotypic and phylogenetic information, the lack of some of these characters avoid the designation of species in prokaryotes, as is the case for *Frankia.*

I.4. Frankia species delineation by DDH

DNA hybridization is recognized as the standard method for the elucidation of relationships between closely related taxa and species delimitation (Stackebrandt & Goebel, 1994). The phylogenetic definition of species would generally include strains with approximately 70% or greater DNA-DNA relatedness and 5°C or less ΔTm; phenotypic characteristics should agree with this definition. The discriminatory power of this method relies on the fact that for two isolates the real sequence similarity has to be at least 80-85% to allow formation of the heterologous hybrids (Stackebrandt & Goebel, 1994).

Thus the DNA-DNA hybridization technique was employed to determine the existence of genomic species within the genus *Frankia*. In a first analysis (An *et al.*, 1985a) one genogroup, consisting of isolates from *Alnus*, *Comptonia* and *Myrica* host plants, was clearly

delineated with homology levels of 67.4-94.1% to the ArI4 strain. Another group consisted in two *Casuarina equisetifolia* isolates, G2 and D11, which displayed a high homology level, 97%, and no correlation with other strains.

By employing the same technique and a larger spectrum of *Frankia* isolates from various geographical areas, Fernandez *et al* (1989), defined at least 9 genomic species within the genus. Posterior DDH trials have described additional genomospecies (Akimov & Dobritsa, 1992a; Bloom *et al.*, 1989b), however because of the lack of consistency to keep the same reference strains and strain collections, minimal or no correlation can be stated between the different genomospecies reported.

In order to bring sense to the *Frankia* genomospecies to date described, a summary is here presented (Table 1). The more representative species and those from which we posses more information are described in more detail. First are described genomospecies reported by Fernandez et al. (1989) and from other studies which kept the same nomenclature system, the following species were reported by Akimov and Dobritsa (1992).

Genomospecies 1 (GS1)

It corresponds to genogroup 1 described by An *et al*., 1985, and to the originally defined *Frankia alni* (Becking, 1974). This clade corresponds also to the **S2** genomic species described by Shi and Ruan (1992) and to the **P1** described by Akimov and Dobritsa (1992). Bloom *et al*., (1989) have described the genospecies **B1** which includes *Frankia* AvcI1.

These strains nodulate members of the genus belong to HSG (they nodulate *Alnus, Myrica* and *Comptonia*) and has been grouped by other methods i.e. RFLP (Nittayajarn *et al.*, 1990), isoenzyme electromorphs, whole cell sugar content and total protein electrophoretic pattern (Gardes *et al.*, 1987) (Benson *et al.*, 1984; St-Laurent *et al.*, 1987)

Genomospecies 2 (GS2)

It encompasses specimens isolated from *A. viridis*, *A. incana* and *A. cordata,* all of them inhabiting French soils. This group is quite divergent from *F. alni* since in average 3% of DNA relatedness was found between them. The geographic distribution of these strains may be restricted because until now no isolates belonging to this group have been isolated abroad (Fernandez *et al.*, 1989a).

Table 1. Genomospecies defined by DNA-DNA relatedness

Strains analyzed	Genomospecies	Reference
Alnus-Myrica and Casuarina HSG		
ArI_4, ArbN4b, ACN1 ; AGNIg, CpI_1, $AGN1_{EXO}{}^{AG}$, MPI_1, AirI1, AvcI1	A1	(An *et al.*, 1985)
ACoN24d, Ar24H3, Ar24O2, Ag24251, ArgN22d, ACN1, CpI_1, A2J	1	(Fernandez *et al.*, 1989)
AvcI1, RBR162013, RBR162008, RBR162010, RBR162014	B1	(Bloom *et al.*, 1989)
CpI_1, ArI_4, ArI_3, AvcI1, AvcI1.R1, AirI1, An2.1, An2.24, AgPM2.8, A7, A43, A51, A153	P1	(Akimov & Dobritsa, 1992)
ArI4	S2	(Shi & Ruan, 1992)
Av22c, AVN17o, $Ac23_{40}$, $AI43_1$	2	(Fernandez *et al.*, 1989)
$ArgP_5{}^{AG}$	3	(Fernandez *et al.*, 1989)
Ai1c, Ai15a, Ag9b	P2	(Akimov & Dobritsa, 1992)
Ai6b	P3	(Akimov & Dobritsa, 1992)
AirI2	P5	(Akimov & Dobritsa, 1992)
CeD, CcI3, CeF, AllI1, ORS020608, ORS020609, Cj1-82, ORS022602	9	(Fernandez *et al.*, 1989)
C1, C2, C3, C4, C5	S1	(Shi & Ruan, 1992)
Elaeagnaceae HSG		
$Ea1_{12}$, $Ea1_2$, $Ea2_6$, $Ea3_3$, $EaCm5_1$, $Ea29_1$, $Ea291_4$, $Ea47_3$, $Ea47_4$, $Ea49_1$, $Ea49_2$, $Ea49_{11}$, $Ea30_5$, $Ea30_7$, $Ea30_9$, $Ea36_7$,$Ea36_9$, $Ea32_1$, $Ea35_2$, $Ea35_4$, $Ea48_2$, $Ea48_4$, $Ea48_8$, $Ea7_4$, $Ea7_1$, $Ea7_3$, $Hr27_{14}$	4	(Fernandez *et al.*, 1989) and (Jamann *et al.*, 1992)
$TX31e^{HR}$, $EAN1_{pec}$, HRX401a	5	(Fernandez *et al.*, 1989)
EUN1f	6	(Fernandez *et al.*, 1989)
$Ea8_4$	ND*	(Jamann *et al.*, 1992)
HRN18a	7	(Fernandez *et al.*, 1989)
$Ea50_1$	8	(Fernandez *et al.*, 1989)
EaI10, EaI11, EaI8	P6	(Akimov & Dobritsa, 1992)
CcI.17	P7	(Akimov & Dobritsa, 1992)
S15, S14, S13, Hr5-o, H109	P8	(Akimov & Dobritsa, 1992)
E1, E13, E15	10	(Lumini *et al.*, 1996)
2.1.7, HrI1	11	(Lumini *et al.*, 1996)
EuI1b	P9	(Akimov & Dobritsa, 1992)
G2, D11	A2	(An *et al.*, 1985)
Atypical Nod-/Fix- strains		
AgI5	P4	(Akimov & Dobritsa, 1992)
Pti1	S3	(Shi & Ruan, 1992)

Genomospecies 3 (GS3)

The strain ArgP5, isolated from a nodule of *A. rugosa* plant in Quebec is the unique member of this genoespecies. It harbors specific phenotypic traits i.e. a specific isoenzyme pattern (Gardes *et al.*, 1987) and high level of sporangia synthesis. Lalonde *et al.* (1988) proposed that this strain represents a different *F. alni* subspecies, *F. alni* subsp. v*andijkii.* However, only 2% of DNA relatedness was found between these two groups, thus representing different genomospecies.

Genomospecies 9 (GS9)

Comprises strains isolated from *Casuarinaceae* inhabiting different ecosystems and geographic sites worldwide, as introduced species. Characterization studies unveiled high genomic homogeneity among these strains: their 16S rRNA sequences are almost identical (99.9%) and genotyping analyses, such as the Rep-PCR and *nif*D-K RFLP (Rouvier *et al.*, 1996), always grouped tightly these strains. Moreover it has been reported that these strains do not possess the ability to grow in soils devoid of their host plants (Zimpfer *et al.*, 1999).

A recent study revealed that CcI3 genome is 3.6 Mb smaller than genome of the strain EAN1pec. The evolutionary implication of this finding might be expressed as high symbiotic specificity toward its host plant and a tendency to become obligate symbiont (Normand *et al.*, 2007a).

Genomospecies 4 (GS4)

This genospecies contains several French *Eleagnaceae*-typical strains for which DNA reassociation with type strain Ea1-12 ranged from 69 to 100%. Jamman *et al*., (1992) have studied the distribution and the genetic diversity of this group in France. From 23 isolates tested 21 were undoubtedly included into the genomospecies; moreover they exhibited low genetic diversity, suggesting this group originated from a recent diversification event of the *Elaeagnaceae* strains carried out in France.

Genomospecies 5 (GS5)

This group also includes *Elaeagnaceae* compatible strains isolated from Canada and United States. The type strain is *Frankia* EAN1pec which has been recently sequenced (Normand *et al.*, 2007a). Even if phylogenetic relationships between two different

genospecies cannot be elucidated by the DNA reassociation values, it is likely to think the GS4 and GS5 are closely related because 49% of homology was found between specimens of these two genomospecies (Fernandez *et al.*, 1989a).

Genomospecies 6, 7 and 8 (GS6-8)

All these three genospecies are represented by unique *Elaeagnaceae* typical strains, EUN1f, HRN18a and $Ea50_1$, respectively. The three specimens yielded low DNA reassociation rates with GS4 and GS5 species.

Genomospecies 10 and 11 (GS 10 and GS11)

Comprises typical *Elaeagnaceae* strains displaying the major phenotypic feature of nodulate both *Alnus* and *Elaeagnus* plants (Bosco *et al.*, 1992). It means they are versatile in their mode of infection. The genomospecies 10 is a homogeneous group (96% to 99% of DNA relatedness) that comprises *Elaeagnus* infective strains. Conversely the genomospecies 11 grouped 2 *H. rhamnoides* strains. By *nifD-K* typing these two genospecies where proved to be closest neighbours in addition to another broad host range isolate HRN18a (Genomospecies 7) (Bosco *et al.*, 1992).

Genomospecies described by Akimov and Dobritsa (1992):

Genomospecies P2, P3 and P5

P2 and P3 contain *A. incana* and *A. glutinosa* infective *Frankia* strains, isolated in Finland. These two clades are phylogenetically related according with Dobritsa (1998). In contrast, the P5 genomic species represented by the AirI2 isolate only seems to be distantly positioned in the phylogeny (Dobritsa, 1998b).

Genomospecies P6

Contains three strains isolated from Russian *E. argentea* that belong to the group of broad host specificity spectrum because they are capable to infect *Alnus*. They are closely related to the GS 8 described by Fernandez et al (1989).

Genomospecies P8

Encompasses typical *S. argentea* and *H. rhamnoides* strains. These strains produce soluble pigment and (Dobritsa, 1998b). This genomospecies displayed the same *nifD-K* and *rrs* RFLP profile than genomospecies 5, likely suggesting the occurrence of a unique genomic group.

In addition to the genomospecies above described, some other are represented by orphan strains (Table 1) which acquired this taxonomical status because of the low DNA reassociation rate exhibited with other genomospecies. In summarize, 12 genomic species have been identified within the *Elaeagnaceae*- infectivity while only 8 have been reported in the *Alnus-Myrica-Casuarina*- infectivity group. It is not likely to affirm that these results are related with the overall diversity existing in each cluster, because bias originated by the strain isolation and sampling cannot be discarded.

In spite of its discriminatory power, the DNA-DNA hybridization exhibits some drawbacks which render it difficult to perform and sometimes inapplicable. The main drawbacks imposed by the DDH technique are the following:

- The respective references have to be included in every individual experiment and a large set of strains implies a disproportionate effort and cost.
- The results depend on the experimental parameters, consequently low reproducibility is achieved. Moreover the retrieved data cannot be accumulated.
- Differences in genome size and in DNA concentration could have a drastic influence on the obtained results.

The technique is useful for isolated specimens only, in the case of obligate symbionts such as the group 2 *Frankia* strains, the method is not applicable.

I.5. Frankia diversity below species level

The subspecies level has been studied in *Frankia* by means of several fingerprinting techniques:

- The PCR-RFLP of some hypervariable genetic regions has been widely employed. The intergenic spacer (IGS) localized between the 16S and 23S rRNA genes is highly informative and considerable variation can occur between species in both length and nucleotide sequence (Gurtler & Stanisich,

1996). For instance, this region allowed to discriminate among endemic *Frankia* populations infecting *C. equisetifolia* in its native geographic area (Rouvier *et al.*, 1992). Additionally, the genetic divergence between the European and North American *M. gale*-infective strains was also stated by means of this IGS analysis (Huguet *et al.*, 2004a).

- The IGS region between two structural genes of nitrogenase *nifD* and *nifK* (Nalin *et al.*, 1995) is commonly combined with the IGS 16S-23S analysis and is suitable for an accurate *Frankia* diversity detection and the estimated relationships deduced from these genotypic data correlate well with established *Frankia* taxonomic schemes. Lumini and Bosco (1999) used both IGS to find the phylogenetic relationships within a large *Frankia* strains set and succeeded in the genospecies identification. Moreover Jamman et al (1992) performed a fine typing of genomospecies 4 members.
- Rep-PCR genomic fingerprinting covers protocols consisting on the amplification of genomic regions that are located between the repetitive extragenic palindromic (REP) sequence (Stern *et al.*, 1984), the enterobacterial repetitive intergenic consensus (ERIC) (Hulton *et al.*, 1991) or the BOX element (Martin *et al.*, 1992) present in *Frankia* genomic DNA. The Rep-PCR has been applied to the *Frankia* diversity study. With the Rep-PCR genetically the homogeneous group of *Ceanothus* endophytes was splitted in 12 subgroups (Murry *et al.*, 1997).This techniques allowed to study the genetic diversity of frankiae nodulating *C. equisetifolia* introduced in Mexicain soils (Pérez *et al.*, 1999).

The discrimination level reached is dependent of the technique employed and the genomic group of *Frankia* studied. However the fingerprinting techniques such as the RAPD (randomly amplified polymorphic DNA) and Rep-PCR have many drawbacks, basically the lack of reproducibility and robustness. In addition these fingerprinting methods can hardly be used to robustly evaluate pairwise genomic distances because they provide only a too limited number of fragments in agarose gel electrophoresis (Meudt & Clarke, 2007b).

I.6. Impact of the Host Plant over the Frankia evolution and diversification

The delineation of host infection groups lead to the hypothesis that evolutionary processes are shared between the host plant and *Frankia* strains, thus there is likely an influence of the evolutionary traits of the host-plant on the *Frankia* evolution and distribution.

I.6.1. Molecular phylogeny of the actinorhizal plants

The AP are a group of angiosperms distributed in 8 botanic families (Table 2), highlighting an important difference with respect to the Rhizobia-Legumes symbiosis in which only one family of plants (*Fabaceae*) and an extra genus (*Parasponia* sp.) are implicated in the association. AP diversity and taxonomy have been studied from diverse perspectives. First taxonomic studies consisted in phenotypic analyses and pollen and fossil records (Cronquist, 1981).

The initial molecular phylogenetic analyses, based on *rbcL* gene, which encodes the large subunit of the photosynthetic enzyme ribulose-1, 5-biphosphate, have shown that the plant families implicated in the symbiotic processes occurred together as part of a coherent genomic clade, later designed as Eurosid I (Soltis *et al.*, 1995) (Fig. 5). This clade comprises both Fabaceae and Actinorhizal Plants along with several families lacking symbiotic properties.

A multigenic approach (including 18S rRNA+*rbc*L+*atp*B sequences) also gathered these plants, thus providing strong evidence for the existence of a nitrogen-fixing clade within the Eurosid I group, confirming the underlying genetic predisposition toward nodulation within this clade of angiosperms (Soltis *et al.*, 2000). The AP appear dispersed in four paraphyletic clades within the Eurosid I clade; (i) the *Rosaceae*, (ii) *Elaeagnaceae* and *Rhamnaceae*, (iii) *Datiscacea*e and *Coriariaceae* and (iv) the higher Hamamelidae, which includes the actinorhizal *Betulaceae*, *Myricaceae* and *Casuarinaceae*. The tree topology suggests that symbioses arose in each group as a result of a separate gain in the ability to be nodulated, leading to the extant AP (Swensen, 1996). Conversely non-symbiotic plant families are interspersed with AP (Fig. 5); supporting the idea that loss of symbiosis may be a common feature of this evolutionary process. An analysis of the correlation between the evolutionary history of the host plant and its symbionts have shown that AP that have

Table 2. Actinorhizal plant taxa, species number and native distribution. Adapted from (Dawson, 2007a; Schwintzer, 1990b; Valdés, 2007b)

Family	Genus	No. Of Species	Ecosystem or soil
Betulaceae	*Alnus*	47	Riparian, temperate forest; moutain valleys, circumpolar soils.
Casuarinaceae	*Allocasuarina*	54	Extremely poor soils.
	Casuarina	16	Sand dunes, subtropical forest.
	Ceuthostoma	2	
	Gymnostoma	18	Secondary forest, ultramafic soils.
Coriariaceae	*Coriaria*	5	Temperate mountian forest, dry woods, rocky places.
Datiscaceae	*Datisca*	2	Temperate and cold forests, stream banks.
Elaeagnaceae	*Elaeagnus*	38	Sand dunes, disturbed agriculture lands.
	Hippophaë	2	Seashores, sand dunes,cliffs, dry river beds, montains valleys.
	Shepherdia	2	Temperate forest.
Myricaceae	*Comptonia*	1	Dry open forest stands.
	Myrica	28	The widest geographic distribution in the world.
Rhamnaceae	*Adolphia*	2	Temperate to subhumid climates, eroded slopes.
	Ceanothus	31	Chaparral from mid-to-low elevations.
	Colletia	4	Xerophytic matorral.
	Discaria	5	Matorral.
	Kentrothamnus	1	Xerophytic matorral.
	Retanilla	2	Xerophytic matorral.
	Talguenea	1	Xerophytic matorral.
	Trevoa	2	Xerophytic matorral.
Rosaceae	*Cercocarpus*	4	Disturbed moutain forest, forest ecosystems of semiarid regions.
	Chamaebatia	1	
	Cowania	1	Xeric chaparral.
	Dryas	1	Tundra.
	Purshia	2	Forest ecosystems of semiarid lands.

diverged more recently were susceptible to a narrower spectrum of *Frankia*, thus suggesting that evolution has proceeded toward narrower promiscuity and greater specialization and has conducted the loss of the symbiosis in more recent lineages (Maggia & Bousquet, 1994).

The last affirmation is supported by findings in the *Myricaceae* which is likely the most ancient actinorhizal family according to the pollen fossil data. Most members of this family have been considered promiscuous because in greenhouse trials they are nodulated by a large spectrum of *Frankia* strains, as an exception *M. gale* showed low levels of promiscuity in natural conditions (Huguet *et al.*, 2001a). Nonetheless, it is likely to affirm that *Myricaceae* possess the genetic background to establish symbiosis with a variety of *Frankia* strains.

In contrast, the genus *Allocasuarina* ssp. shows a high specificity level. It is nodulated by a unique *Frankia* genomic group, which is not capable to nodulate other *Casuarinaceae*. (Simonet *et al.*, 1999)

I.6.2. Coevolution

Several co-evolutionary processes have been widely studied in the symbiosis Rhizobia-Legumes (Alvarez-Matínez *et al.*, 2009). The coevolution is a controversial concept within evolutionary biology. Basically in a symbiotic interaction it refers that both species are evolving in parallel as a result of reciprocal selection. When comparing the phylogeny of both partners a reflected correlation must be found not only in the branching pattern, but also in the branch lengths (Laine, 2009). The rough correlation between the plant and *Frankia* evolutionary pathways opens the question about the evidence of co-evolution of the two partners.

In order to find cladistic correlation Jeong and Myrold (1999) used the *rbc*L gene and the 16S rRNA sequence for the respective AP and *Frankia* phylogeny delineation. The authors did not find evidence of co-evolution because the plant phylogeny did not reflect the *Frankia* tree and shows the presence of cross-infection events (Fig. 6). These authors concluded that if a co-evolutionary event existed it must be above family level of plants. Similar results were obtained by using a *nifH*-based *Frankia* phylogeny, thus suggesting there is no evidence of co-evolution processes in the Actinorhizal symbiosis (Jeong *et al.*, 1999).

Besides, in order to detect coevolution signals four symbiotic traits were studied among several actinorhizal models: i) the shape and septation of *Frankia* vesicles within the nodules (Newcomb *et al.*, 1979), ii) the location of vesicles within the infected cells (Mirza *et al.*, 1994), iii) the location of infected cells within the nodule and iv) the mode of infection (Berry & Torrey, 1983; Liu & Berry, 1991).

Frankia enters the roots of its host plant either by root hair infection (RHI) or by intercellular penetration. 16S rRNA cluster 1 strains enter through the RHI pathway, while the strains of cluster 2 use the direct intercellular penetration (IP) (Berry & Torrey, 1983; Liu & Berry, 1991). Cluster 3 comprises flexible strains, which are able to enter into the plant by the two infection mechanisms (Bosco *et al.*, 1992). The relevance of the infection mechanism in the evolutionary scenario of the symbioses is still unknown, but it suggests that the IP pathway is the ancestral one and it was retained by cluster 2 strains, likely the basal group in the *Frankia* phylogeny (Fig. 6). According to the Fig. 6, the Cluster 1 and 3 diverged from a common ancestor, the latter keep a versatile mode of infection (RHI/IP) while the cluster 1 strains have specialized in the root hair infection, in any case this correlation indicates coevolution existence though.

If particular *Frankia*-Host plants interactions are analyzed, some degree of co-evolution may be found i.e. the high specialized relationship between *Allocasuarina* species with its natural nodulating strains. Besides, *Casuarina* species are nodulated by different *Frankia* genotypes in their native ecosystem (Australia and western pacific islands). However in other places around the world where these plants were introduced, only a predominant *Frankia* genotype has been described, leading to the hypothesis such strain was exported with *Casuarina* seeds (Simonet *et al.*, 1999). The fact no other strain may nodulate the *Casuarina* plants inhabiting foreign soils gives evidence of a co-evolutionary process between the two actinorhizal symbionts.

The above mentioned example support the hypothesis that co-evolution might be present at genus level, however data presented does not support the idea that this process occurs at familial level. Thus, the complex divergence patterns in *Frankia* phylogeny may best be explained in a geographic mosaic theory of co-evolution in which multiple confounding factors like geographic isolation, plant host preferences, and environmental factors converge to shape the evolutionary patterns of *Frankia* (Thompson, 1999; Welsh *et al.*, 2009c).

I.7. Evolution and speciation of Frankia is the result of diverse factors

Frankia strains inhabiting soils devoid of their host plants have been repeatedly detected, suggesting that saprophytic subsistence in soil is likely widespread in the genus (Huss-Danell, 1997; Maunuksela *et al.*, 1999). Nevertheless, the adjective cosmopolitan is not completely applicable to these actinomycetes since some strains might be restricted to a particular geographic area (i.e. *Casuarina* strains) while some others (i.e. *Alnus* strains) are globally distributed regardless of the presence of a suitable host. The biogeographic distribution requires the study of *Frankia* populations variation, at different levels and spatial scales between distantly located sampling sites (Ramette & Tiedje, 2007). The processes designing the biogeographic patterns in prokaryotes are speciation, extinction and dispersal, in *Frankia* these processes should not be different, and likely occur in the following manner:

Vertical speciation

The **vertical speciation** in prokaryotes is affected by the existence homologous recombination and lateral gene transfer (Lawrence, 2002). Such genetic events may not alter the integrity of distinct lineages and its geographical adaptation. In *Frankia* the genetic processes of lateral transfer have not been well studied, primarily because of the lack of transformation and mutagenesis systems. Plasmids, transposons and insertion sequences likely play an important role in the plasticity of *Frankia* genome (Normand *et al.*, 2007c; Simonet *et al.*, 1984). Nonetheless little is known about the promiscuity of gene transfers, the extent of these processes in environmental conditions and the impact on *Frankia* speciation.

Allopatry

The **allopatric** divergence has emerged as a prerequisite for population speciation and physical isolation has been advocated as an additional evolutionary force shaping bacterial populations (Whitaker, 2006). This process results from the divergent evolution of geographically isolated populations. A first example of *Frankia* strains allopatric speciation comprises the case of the genus *Ceanothus* (a *Rhamnaceae* member). The *Ceanothus* – *Frankia* endophyte relationship was likely guided by geographic isolation, because they are

confined to North America while the other *Rhamnaceae* (nodulated by cluster 3 strains) inhabit the rest of the continent. Other example is the allopatric speciation of some strains nodulating relict *Myrica* species (Huguet *et al.*, 2005b). Geographical isolation may permit speciation to become the result of neutral genetic drifts converging in the accumulation of mutations.

Soil conditions seem to be important, in the successful nodulation of appropriate hosts by *Frankia*. This conclusion is supported by direct demonstrations of dominance in alder and *Myrica* stands (Clawson *et al.*, 1999; van Dijk, 1984). In some studies *Frankia* structure and soil conditions exhibited similar patterns of variation among habitats. The soil organic matter, exchangeable cations and pH seem to have a synergistic effect on the *Frankia* diversity and distribution (Anderson *et al.*, 2009). Therefore when a genotype is subjected to physical isolation and the two different places exhibit different edaphic conditions it is likely that an allopatric divergence event will occur for this *Frankia* genotype.

Dispersion

The physical **dispersal** of *Frankia* strains has been reported and both wind and birds have been proposed as *Frankia* propagules vectors (Burleigh & Dawson, 1994; Paschke & Dawson, 1993), however these mechanisms are valid only for the saprophytic-stage; for more specialized symbionts the transport might be guided by the host dispersion itself. It is important to note that the presence of resistance structures in *Frankia*, the spores, might have a positive impact in *Frankia* dispersion.

Genome dynamics

Comparative genomics analyses have demonstrated that *Frankia* genomes exhibit differences in mobile elements content. The number of insertion sequences (IS) and prophage found in the AcN14a genome was low in contrast to the CcI3 and EaN1pec genomes which posses a higher average density of such an elements (Normand *et al.*, 2007a). Thus, AcN14a

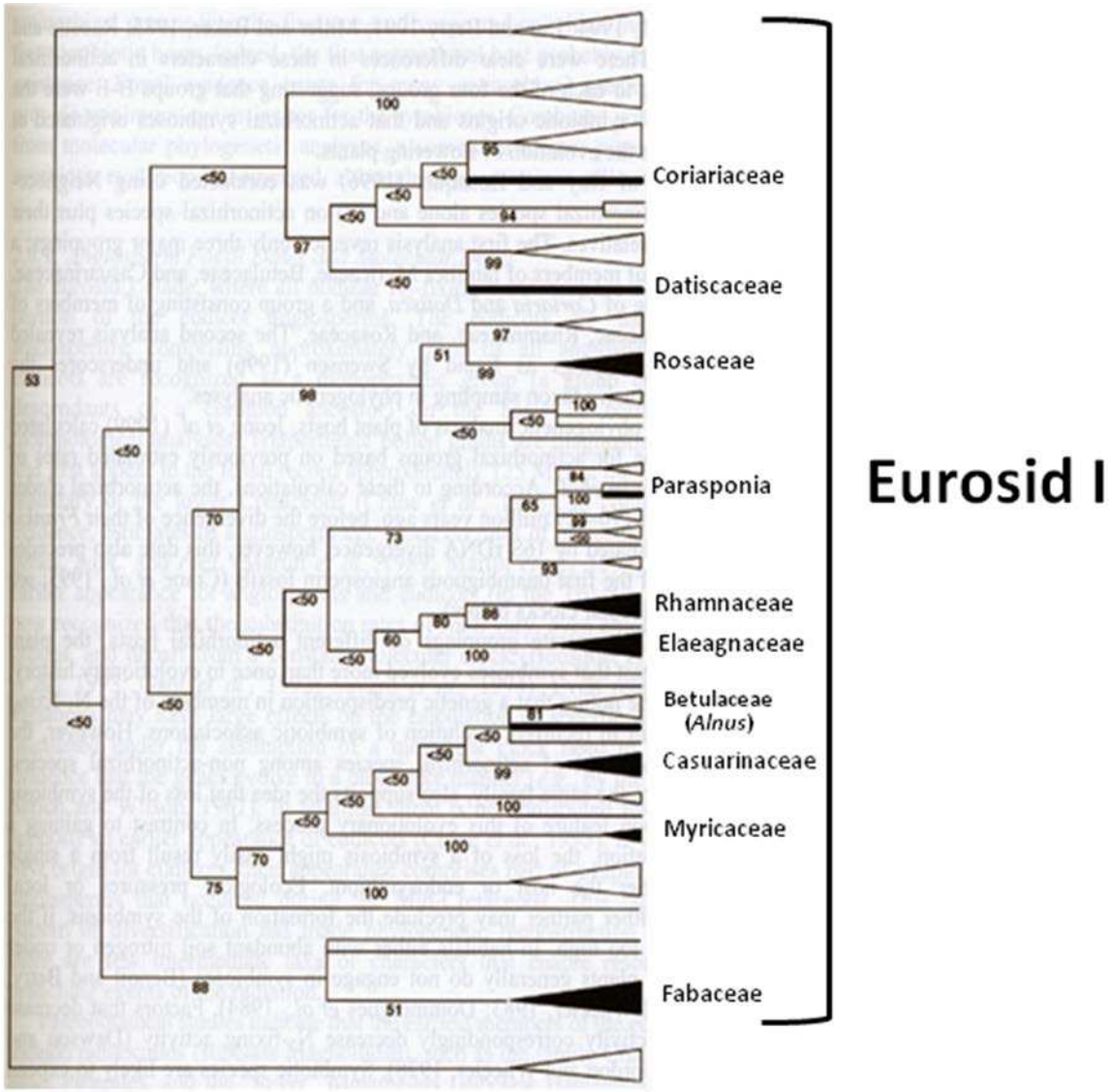

Fig 5. Summarized Maximun parsimony tree representing a part of the Eurosid I clade, which contains the Actinorhizal Plant families, based on the ***rbcL*** sequence.Modified from (Swensen, SM and Benson D 2007).

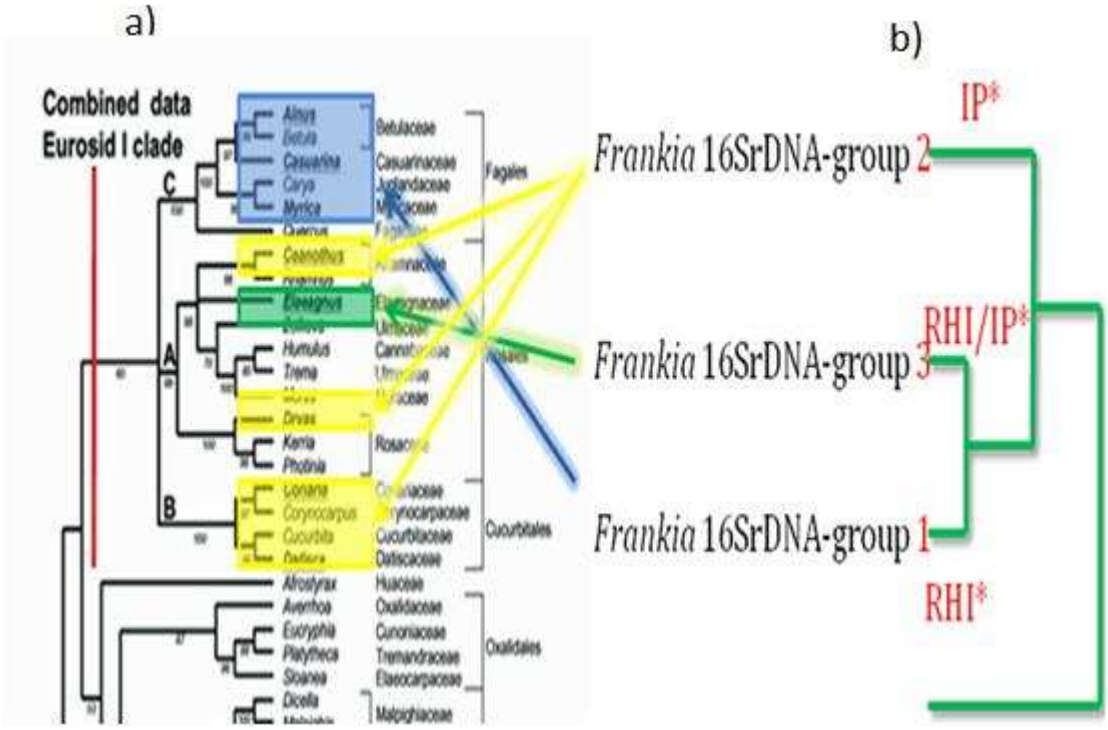

Fig. 6 a) Partial Maximum Parsimony Tree of the Eurosid I clade. The three actinorhizal clades are A-C. Bootstrap values below the branches calculated using 1000 replicates. Adapted from (Benson *et al.*, 2004) and based in the data from (Soltis *et al.*, 2000) and of (Swensen & Mullin, 1997). The red line demark the Actinorhizal clade and in yellow are circled the host plants infected by the group2 *Frankia* strains, in green the host nodulated by group 3 *Frankia* and in Blue the group 1 *Frankia* compatible host plants. b) Schematic tree representing the phylogenetic relationships between three major *Frankia* genomic groups. *Indicates the mode of root infection, the RHI (Root hair infection) and the IP (Intercellular penetration).

genome might be more stable and less subjected to rearrangements than the other two. Besides, an important difference in genome size among the three reference strains has been found, duplication and new genes acquisition (with the respective retention) as well as gene have been proposed as the responsible of this phenomenon (Normand *et al.*, 2007a).

The EaN1pec genome has been enlarged by gene duplication processes, genes whose products are associated with metabolic processes as well as mobile genetic elements. The gene duplication is a major process by which bacteria are capable to adapt to changing environments (Gevers *et al.*, 2004).

Besides, the genome of the CcI3 is 3.6 Mb smaller than EAN one, the loss of genes is the first responsible process by which this genome have lost the size. It is not as obvious that the gene loss could serve as a means of bacterial adaptation, however in *Shigella* the deletion of DNA was crucial to the development of virulent strains (Nakata *et al.*, 1993). A large set of symbionts and pathogens have undergone more massive gene loss and one consequence of genome reduction is that cannot revert to host-independent life-style (Ochman & Moran, 2001). Thus, the ongoing evolutionary processes of the CcI3 strain might be lead to the establishment of an obligatory symbiosis.

I.8. Conclusions

In this work a general review of the current taxonomic classification of the genus *Frankia* was presented. It is not likely to affirm that this classification manages to be clear and genomic, ecologic and phenotypic data are still lacking. To date, the grouping of *Frankia* strains has been based on the information retrieved by cross inoculation trials as well as phenotypic analysis. Moreover, similar grouping patterns have been yielded when comparative sequence analysis of single genetic markers (16S rRNA, *nifH* etc…) has been carried out.

In summarize, all the methods hitherto applied, allowed to unveil the diversity existing within the genus. Such diversity is exhibited in a variety of manners: 1) functionality: nitrogen fixation efficiency, sporulation rate, substrate utilization 2) symbiotic properties: host specificity spectrum, nodulation speed, root-hair infection mechanism, *in planta* sporulation 3) Genetic: 16S rRNA groups, other markers based clustering 4) genomic: genome size and

consequently gene content. Hence, it is sensible to affirm that diversification processes are indeed happening, though the frequency and the mechanisms responsible are not yet clear.

Besides, DNA-DNA technique allows reaching the species level, and so genomospecies definition. However, the low applicability of this method is reflected in the scarce number of *Frankia* genomospecies described. Moreover the speciation within the non-cultivated strains group will never be determined by means of this technique because of the obstacles it imposes. Thus, it is imperative to find alternative techniques suitable to replace the DDH technique. Such techniques must be easily applicable, methodologically flexible and allow non cultivable strains characterization; they also must lead the grouping of strains with a cutoff values equivalent to the arbitrary 70% of DNA relatedness. Finally, they must provide enough information to delineate coherent clusters within an evolutionary framework.

Finally, it s important to keep in mind that *Frankia* diversification and geographic distribution is the result of the convergence of many factors (Host plant- dispersion, dispersion, allopatry, environmental conditions etc.) and all they must all be considered for the accurate phylogenetic depiction of the genus and for species definition.

CHAPTER II: *Frankia* PHYLOGENY DEPICTION AND SPECIES DEFINITION BY AMPLIFIED FRAGMENT LENGTH POLYMORPHISM (AFLP) AND MULTILOCUS SEQUENCE ANALYSIS (MLSA)

Chapter II: *Frankia* phylogeny depiction and species definition by Amplified Fragment Length Polymorphism (AFLP) and Multilocus Sequence Analysis (MLSA)

Introduction

The current phylogeny and taxonomic classification of the genus *Frankia* is based on polyphasic approaches and currently divides the genus into four major clusters (16S- clusters) matching with the host compatibility groups (Hahn, 2007). However, the hurdles imposed by *in vitro* culturing and by non cultivability of some strains, have restricted extensive taxonomical studies and therefore, accurate *Frankia* classification and species definition have not been achieved.

DNA-DNA hybridization (DDH) strategy is regarded as the gold standard for species definition and description (Gevers *et al.*, 2005a; Stackebrandt *et al.*, 2002b). Studies focusing the species identification within the genus *Frankia* were performed by several laboratories and led to the establishment of about twenty coherent genospecies distributed among three 16S-clusters (Akimov & Dobritsa, 1992a; An *et al.*, 1985a; Bloom *et al.*, 1989b; Fernandez *et al.*, 1989a). Some others orphan strains representing groups apart were detected; nevertheless they cannot be equated with previously identified species because a different set of strains was used in each analysis and thus it is uncertain whether they represent novel species or belong to those already described (Fernandez *et al.*, 1989a). In all cases the species status has not been granted because of the scarce phenotypic information available.

The DDH display some drawbacks and limitations, because it requires high amounts of excellent quality DNA, is non-cumulative and time consuming. Moreover it is not applicable to analyze non isolated bacteria (Gevers *et al.*, 2005a). *In vitro* culturing of *Frankia* represents itself a challenge. Indeed, around 50% of actinorhizal symbionts are considered non culturable and the slow growing rates and the poor biomass production of the available isolates, render the DDH application a cumbersome. For this reasons the ad hoc committee for the re-evaluation of the species definition in bacteriology have proposed different techniques and methods that can fairly replace DDH technique (Stackebrandt *et al.*, 2002b). Two of these promising alternative methods have been chosen to carry out our study and are here described.

Chapter II: *Frankia* phylogeny depiction and species definition by Amplified Fragment Length Polymorphism (AFLP) and Multilocus Sequence Analysis (MLSA)

II.1. AFLP: Amplified Fragment Length Polymorphism

The AFLP technique originally described in 1995 by Vos et al. in 1995 is included in the category of selective restriction fragment analyses fingerprinting techniques. Until now it has been widely employed in epidemiology and taxonomy even if it was initially conceived for the construction of high density linkage maps and molecular breeding (Vaneechoutte, 1996). This method is based on the enzymatic restriction of the DNA under study. There are four major steps comprised in this method (Fig. 1).

II.1.1. AFLP protocol

The protocol originally designed by Vos et al (1995) is here described, a schematic representation is found in the Fig. 1. Total DNA is extracted, purified and then subjected to digestion with a pair of restriction endonucleases, a rare cutter (recognition site of 6 nt) and a frequent cutter (recognition site of 4 nt) (Vos *et al.*, 1995a). The pair of endonucleases most commonly used for bacterial and plant genomes analyses is EcoRI/MseI, however the combination can vary depending on the genome G+C content (Vos *et al.*, 1995a). Double linkers or adaptors with complementary sticky ends then ligated to the restriction fragments.

The PCR primers used for a preselective amplification contain DNA sequences homologous to the linker ligated, in this step a subset of all restriction fragments generated is amplified, the name preselective is given because a preselection of the markers (fragments) is made, though at the end of this step the quantity of fragments is still too large to be resolved by electrophoresis. In a second round of PCR one or two selective bases are added at the 3' end of the PCR primers, thus reducing the number of amplified fragments to a suitable number to be visualized by electrophoresis. One of the primers is radioactive or fluorescently labeled, so that all the strands synthesized by from this primer will be fluorescently labeled. The PCR products are analyzed by capillary fluorescent electrophoresis in an automatic sequencer. The capillary instruments detect fragments present in the spectrum of each fluorophore, producing an electronic profile of relative fluorescence units (RFU) versus fragment size (Mortimer & Arnold, 2001b). The classical profile of an AFLP analysis is represented in the figure 2.

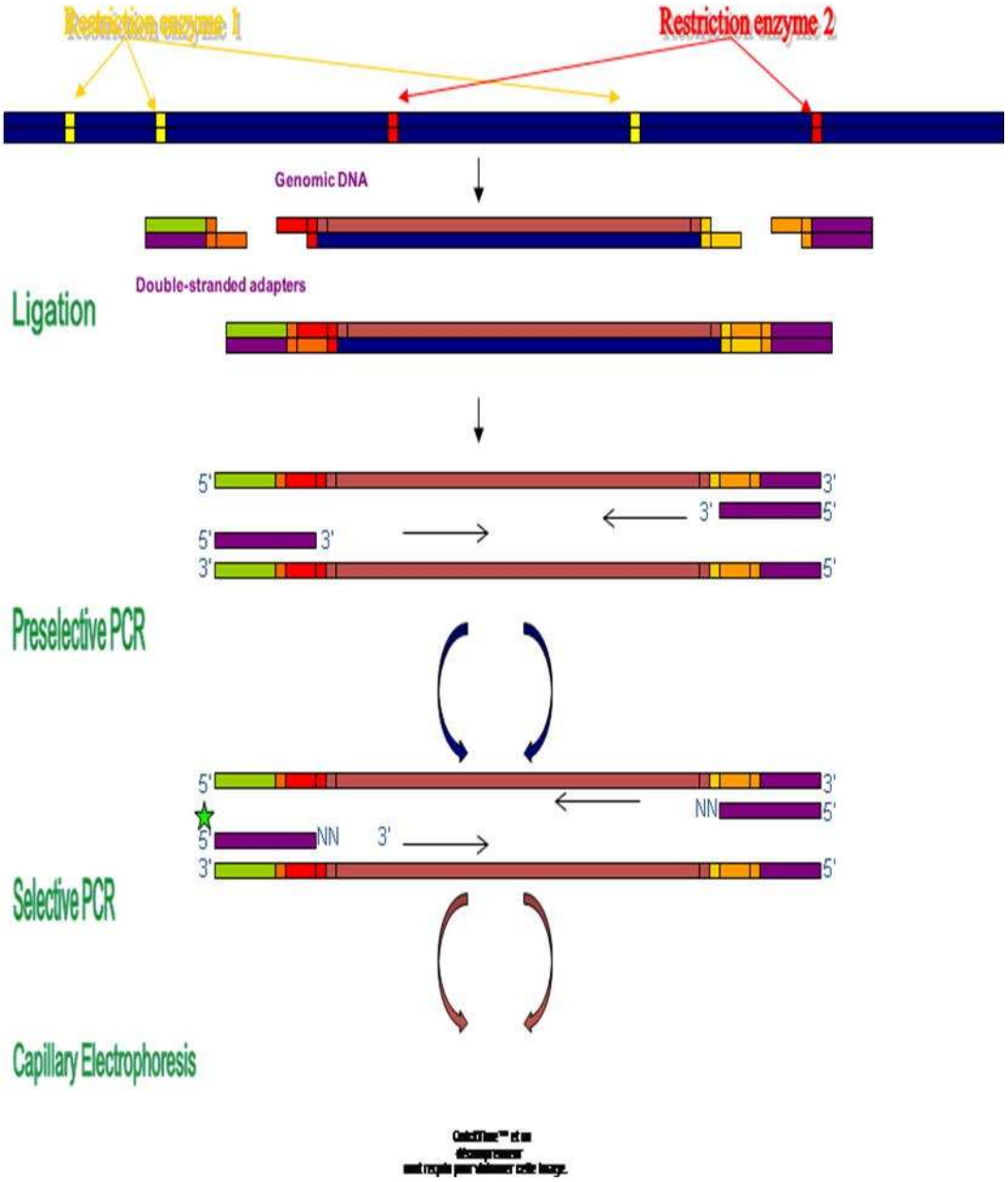

Fig. 1 Schematic representation of the AFLP protocol

For the data scoring, profiles from multiple individuals are aligned and scored. A binary matrix is constructed based in the presence (1) or absence (0) of same-length fragments. The analytical treatment of the data will depend on the aim of the study; statistical traits as well as phylogenetic methods are employed in the AFLP data analysis (Meudt & Clarke, 2007b).

II.1.2. Advantages of the AFLP technique over other fingerprinting methods

The AFLP technique exhibit different advantages compared with other fingerptinting techniques as the RAPD (Random Amplified polymorphic DNA) and RFLP (Restriction Fragment Length Polymorphism) on the basis of **reproducibility and robustness** (Blears *et al.*, 1998; Vaneechoutte, 1996). The high reproducibility level obtained with the AFLP is due to the possibility of using stringent PCR annealing conditions and the selective amplification of fragments.

Besides, genotyping methods differ in their **discriminative power**, depending on the taxonomic level a category(Gurtler & Mayall, 2001). In bacteriology, discrimination level of fingerprinting methods ranged from species to strain level. For bacteria, AFLP analysis seems to have the same taxonomic range as other fingerprinting techniques (Fig. 3) like RAPD analysis, pulsed field electrophoresis (PFGE) and RFLP analysis, repetitive DNA-sequenced based PCR (rep-PCR) (Olive & Bean, 1999). Thanks to the amount of informative fragments generated the discriminative power of AFLP results in most cases to be higher than this from other methods. The AFLP is protocol adaptable, thus the number of samples and enzyme-selective nucleotides combination can be easily scaled up, in order to increase the **number of polymorphic markers generated** (Olive & Bean, 1999; Vuylsteke *et al.*, 2007). There are the so called high-throughput analyses in which an enormous number of data (>1000) are generated in a minimal period of time and moderate cost (van den Braak *et al.*, 2003). Making this technique well suited for comparison of alleles frequencies within populations (Krauss, 2000).

One of the major problems of using fingerprinting techniques is the weak phylogenetic signal obtained, thus threshold values to delimit groups are variable and depend on the group

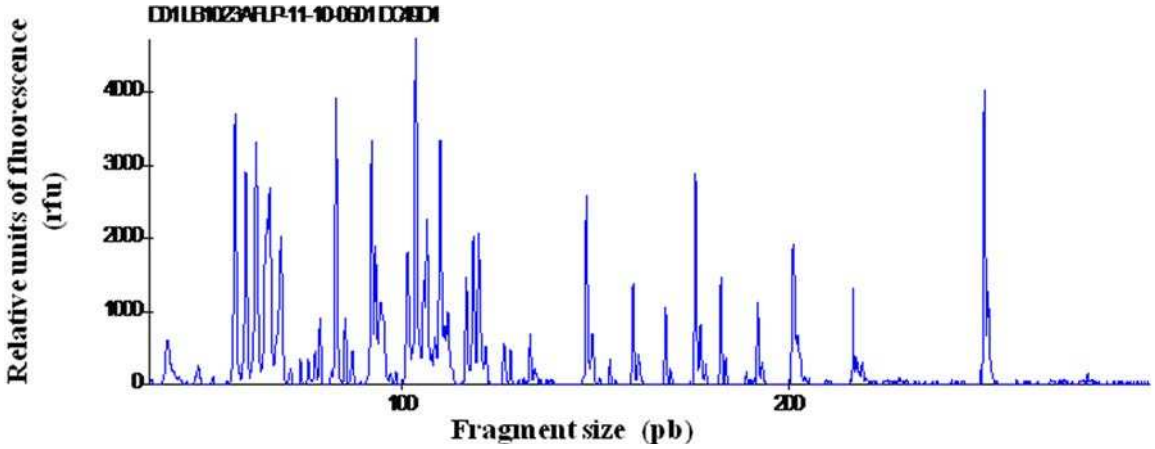

Fig. 2 Electrophoretic AFLP profile.

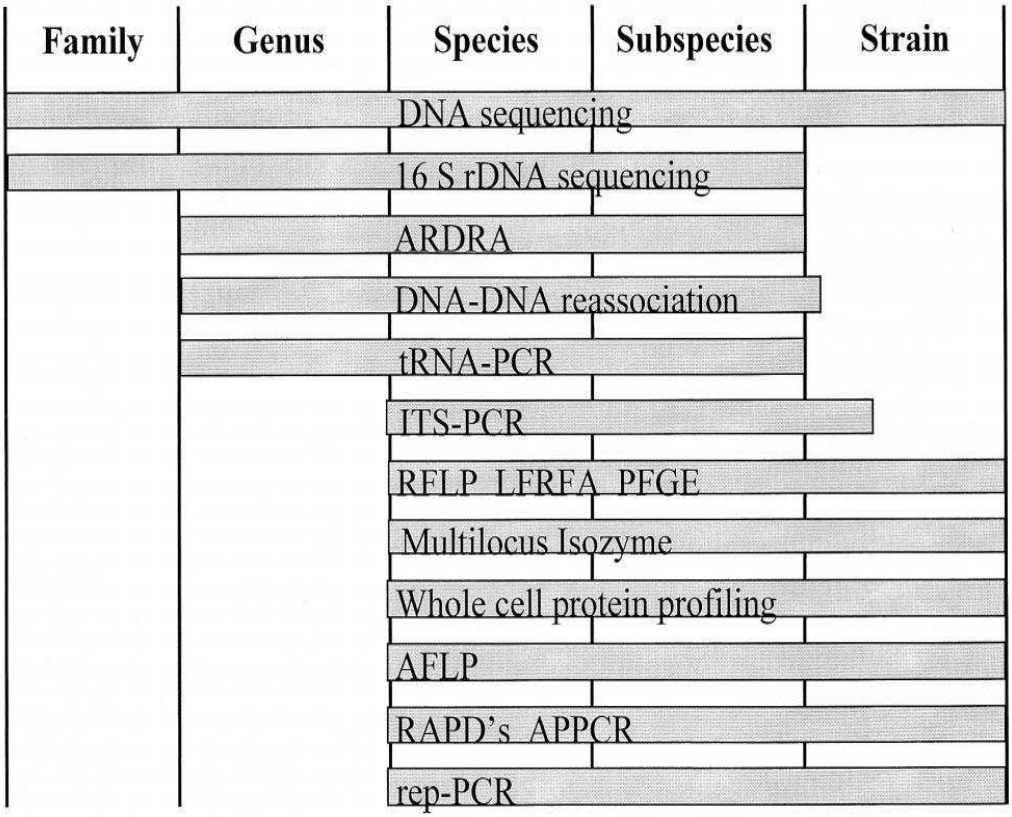

Fig. 3 Relative applicability of various fingerprinting and DNA techniques at different levels of taxonomic resolution. Taken from (Savelkoul *et al.*, 1999)

of bacteria analyzed. In addition, the larger the phylogenetic distance between the organisms studied is, the higher the probability two fragments of same size represent different alleles (phenomenon called Homoplasy) (Pompanon *et al.*, 2005). The anonymous nature of the generated fragments could be result in the scoring of fragments representing different genomic regions as the same allele generating clustering artifacts (Meudt & Clarke, 2007b).

It has been argued that AFLP implies the organism typed must be isolated, since DNA extracted from complex samples such as stool, soil or root nodules can disturb the electrophoretic patterns (Meudt & Clarke, 2007b; Olive & Bean, 1999). Therefore, intracellular organisms obligate symbionts have not been submitted to this typing analysis.

II.1.3. Bacterial species definition by means of AFLP

The hurdles imposed by the DDH technique have been reviewed elsewhere in this manuscript, however it is worth to note that in order to search alternative methods to replace the so called gold standard in bacterial species definition Stackebrandt and col. 2002 have proposed the AFLP and other fingerprinting techniques as alternative.

In some studies, phylogenetic signals have been detected within the AFLP data. By employing this technique coherent genomic clades corresponding to genomic species have been delineated in some bacterial taxa, giving confidence this technique can be applied in taxonomic studies (Clerc *et al.*, 1998a; Janssen *et al.*, 1996; Portier *et al.*, 2006b; Rademaker *et al.*, 2000b). AFLP analysis of Agrobacteria showed that members of the same genomospecies clustered consistently together in groups supported by bootstrap resampling (Mougel *et al.*, 2002b). However it has been suggested to always standardize the genomospecies definition cut off value, by employing a set of reference strain for which DDH values have been determined (Mougel *et al.*, 2002b; Portier *et al.*, 2006b; Rademaker *et al.*, 2000b; Stackebrandt *et al.*, 2002b).

II.1.4. AFLP an alternative technique to study Frankia diversity

In summarize the AFLP technique exhibits characteristics that make it a reliable alternative to analyze *Frankia* genetic diversity. In addition to the advantages already enunciated, this technique seems suitable for the *Frankia* biodiversity analysis because:

The quantities of DNA required for the start up of the study are minimal and so a little biomass production is required.

- *In silico* analysis can be employed to initially standardize the optimal conditions of enzymatic digestion and PCR parameters. Consequently the high resolved profiles can be generated with the less of homoplasy.

- Cut off values for species definition can be stated, because of the availability of DDH values for a number reference strains.

The major challenge *Frankia* genomospecies definition imposes is the non cultivability of some strains. In this regard, the versatility and technical flexibility of AFLP has to be tested.

II.2. MLSA (Multilocus Sequence Analysis)

The nucleic acid sequences have high potential value since they contain more evolutionary information than traditionally used phenotypic traits and because it is precisely defined and relatively simple to determine. Moreover sequencing has become a routine laboratory technique, it is easy to obtain at relative low cost and allows storage of sequences in data bases.

Evolutionary information can be retrieved from the analysis of nucleotide sequences accounting for the following features: 1) must be universal in its distribution and 2) it must contain sufficient variation to distinguish between two different genetic groups. On these basis the 16SrRNA has been proposed as suitable marker for strain identification and taxonomical classification (Stackebrandt & Goebel, 1994). Thus a phylogenetic classification pionnered by Woese based on the 16SrRNA comparative analysis has been extensively applied, a rapid search on the GenBank databases retrieve 2, 164 279 16S sequences. By comparing the 16S rRNA sequence of an isolate to those of all known prokaryotic species, microbiologists can rapidly identify prokaryotic strains and place them in a universal context.

Chapter II: *Frankia* phylogeny depiction and species definition by Amplified Fragment Length Polymorphism (AFLP) and Multilocus Sequence Analysis (MLSA)

Although the advantages and the enormous utility of 16S rRNA in bacterial identification and classification, 16S rRNA based phylogeny is currently in the middle of the debate because the extent at which this gene phylogeny reflects the evolutionary history of an organism is not known (Cohan & Perry, 2007; Gevers *et al.*, 2005a; Rossello-Mora & Amann, 2001). Moreover it has been demonstrated that any single measure of sequence similarity is subjected both to simple stochastic variation and to the influence of lateral transfer and recombination (Ludwig & Klenk, 2005). In addition, several copies of the ribosomal operon, with variable sequences, have been described in the enterobacteries and variation among the different 16S rRNA copies exists, producing conflicting phylogenetic signals (Cilia *et al.*, 1996).

Another problem is the lack of resolution of this marker when compared with the DNA-DNA hybridization technique in eubacterial and Archea species definition. Whereas isolates that have less than 97% 16S rRNA sequence similarity usually share <70% DDH and belong to different species. However there have been reports where organisms sharing <70% of DNA-DNA relatedness exhibit > 97% of 16S rRNA sequence homology (Gevers *et al.*, 2005a; Stackebrandt & Goebel, 1994). Consequently, near identity of rRNA gene sequences does not eliminate the need to apply other methods to further explore whether isolates are sufficiently similar to be assigned to the same species (Gevers *et al.*, 2005a). This is the case of *Frankia* genus where 98% of similarity was found between strains belonging to two different genomospecies, the CcI3 and EaN1pec, and which differ by 3.6 Mb in their genome size (Normand *et al.*, 2007a).

Consequently novel phylogenetic markers must be found for the genomospecies definition. The *ad hoc* committee considered the house-keeping genes for this purpose as they evolve relatively slowly (though more rapidly that 16S rRNA genes) and most of the variation accumulated in these genes is considered to be selectively neutral (Hanage *et al.*, 2006; Stackebrandt *et al.*, 2002b). Thus, they propose the evaluation of protein-coding gene sequences under stabilizing selection as markers to define phylogenetic relationships between organisms. Such genes should be placed at diverse chromosomal loci and widely distributed among taxa and be represented in one copy (Stackebrandt *et al.*, 2002b). In that way, the recombination events are avoided, and in case of such event the concatenation of all the sequences will buffer this signal (Gevers *et al.*, 2005a).

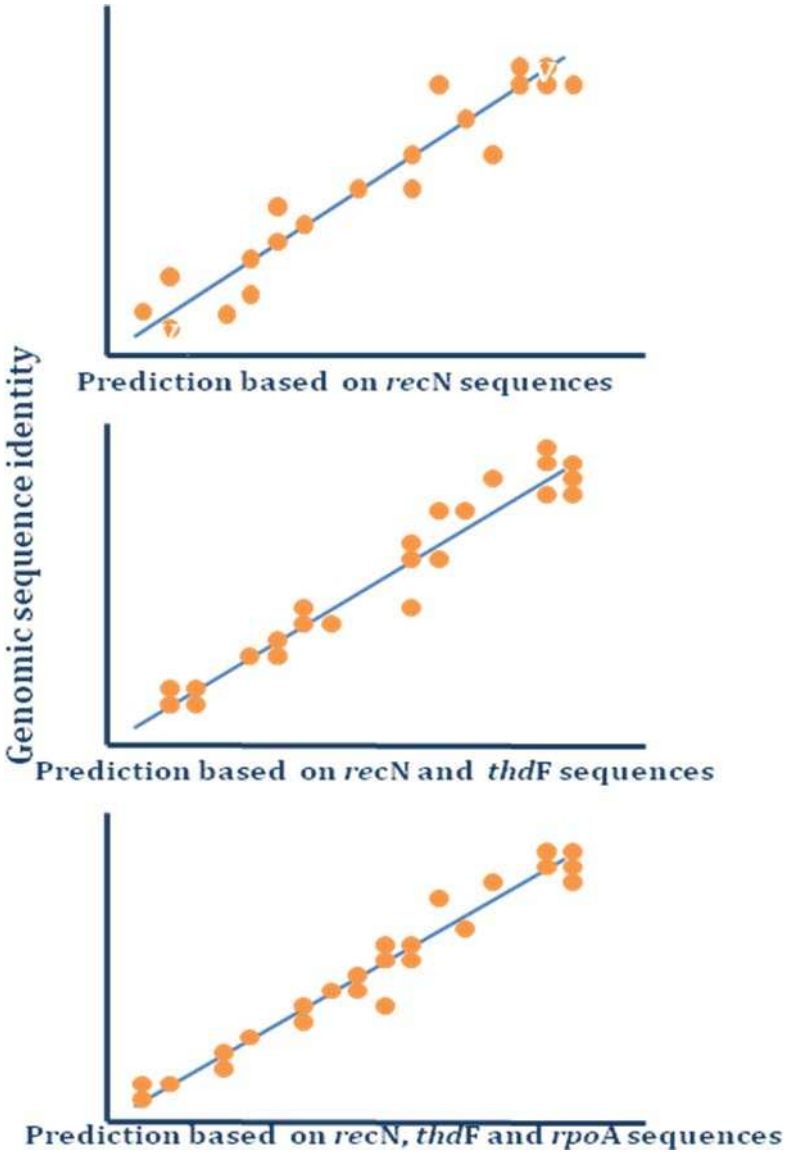

Fig 4. Linear regresion analysis of whole-genome sequence identity score versus scores predicted based on the models described by Zeigler, D. 2003, of three housekeeping genes. The figure modified from that present in the original publication, aim to show that there exist an almost linear relation between the identity calculated by employing all the genome information and the identity scored from housekeeping gene sequence analysis. Moreover such concordance is intensified when several gene sequences (in these case three) are concatenated.

Moreover it has been stated that housekeeping genes sequences consistently reflects the genomic identity and the relation gets almost linear when several sequences are employed (Fig. 4) (Zeigler, 2003).

This approach, called Multilocus Sequence analysis (MLSA) is based on the use of the concatenated sequences of multiple core (housekeeping) genes to assess strain clustering patterns (Bishop *et al.*, 2009b; Gevers *et al.*, 2005a). MLSA aims to depict the distribution of strains of closely related species within an evolutionary framework. The clusters depicted by MLSA are in most cases (Gevers *et al.*, 2005a; Hanage *et al.*, 2006) in agreement with designations obtained by the current polyphasic approaches (Vandamme *et al.*, 1996), therefore corresponding with the species hitherto defined by taxonomists.

The MLSA high resolution power is widely documented and the ease of data accumulation lets easily increase the number of strains and taxa analyzed. Moreover the amplification of the sequences is carried out by a simple PCR step so the non-cultivability is easily overcome. In consequence this method is now considered as the reference in genomospecies definition in prokaryotes (Deletoile *et al.*, ; Gevers *et al.*, 2005a). In this regard two of the most promising approaches were selected to be applied to revisit *Frankia* genus phylogenetic depiction but this time a well defined set of strains representing the ecological, genetic and phenotypic diversity -with the inclusion of non isolated strains- of the genus will be submitted to the analysis. The aim of this study is to depict a robust phylogenetic representation of the genus with a good discrimination of the different taxonomic levels. The subsequent validation of our results by comparison with reference techniques (such as DDH) will allow us to fix the thresholds for species definition and then propose a phylogeny that can be used as reference for future studies.

Chapter II: *Frankia* phylogeny depiction and species definition by Amplified Fragment Length Polymorphism (AFLP) and Multilocus Sequence Analysis (MLSA)

Chapter II: *Frankia* phylogeny depiction and species definition by Amplified Fragment Length Polymorphism (AFLP) and Multilocus Sequence Analysis (MLSA)

(Article 1) **Genomospecies identification and phylogenomic relevance of AFLP analysis of isolated and non-isolated strains of Frankia spp.**

Bautista G.H.H.[1,3], Cruz H.A.[2], Nesme X.[1], Valdés M.[3], Mendoza H.A.[2] and Fernandez M.P.[1]

Running title: Frankia-genomospecies delineation by AFLP

Laboratory addresses

[1] Université de Lyon; Université Lyon1; CNRS; INRA; Ecologie Microbienne, UMR5557, USC 1193, 43 boulevard du 11 novembre 1918, F-69622 Villeurbanne Cedex, France

[2] Centro de Biotecnología Genómica, IPN, col. Narciso Mendoza CP 88710, Reynosa, Tam, Mexico.

[3] Laboratorio de Microbiología Agrícola. Escuela Nacional de Ciencias Biológicas, IPN, México D.F.

Corresponding author :

Maria Fernandez
Ecologie Microbienne, bât. Mendel, 43 boulevard du 11 novembre 1918, F-69622 Villeurbanne Cedex, France
Tel: (33) 472 44 82 00
Fax: (33) 426 23 44 68
E-mail: maria.fernandez@univ-lyon1.fr

Journal: Systematic and Applied Microbiology (Submitted)

Chapter II: *Frankia* phylogeny depiction and species definition by Amplified Fragment Length Polymorphism (AFLP) and Multilocus Sequence Analysis (MLSA)

ABSTRACT

Amplified Fragment Length Polymorphism (AFLP) was tested as an alternative method to DNA-DNA hybridization technique (DDH) to delineate genomospecies and the phylogenetic structure within genus *Frankia*. Forty *Frankia* strains including representative of seven DDH genomospecies were typed to infer current genome mispairing (CGM) and evolutionary genomic distance (EGD). The constructed phylogeny showed the occurrence of three main clusters corresponding to the host-infecting groups as already proposed. In all instances, strains previously assigned to the same genomospecies were grouped in coherent clusters. A highly significant correlation was found between DDH values and CGM computed from AFLP data. The species definition threshold was found to vary from 0.071 to 0.098 mismatches per site according to host-infecting groups, presumably as a result of large genome size differences. Genomic distances allowed us to assign new *Frankia* strains to nine genomospecies previously determined by DDH. The applicability of AFLP for the characterization of uncultured endophytic strains was tested on experimentally inoculated plants and then applied to *Alnus incana* and *A. viridis* field nodules hosting culture recalcitrant spore-positive (Sp+, that nodulate *in planta*) strains. Only 1.3% of the total AFLP fragments were shown to be generated by the contaminant plant DNA and did not interfere with accurate genomospecies identification of strains. Applied to field nodules, the procedure enabled to determine that *Alnus* Sp+ strains are bona fide members of the *Alnus*-*Myrica* host infecting group. They displayed a significant genomic divergence from genomospecies G1 of *Alnus* infecting strains (i.e. *Frankia alni*) and thus may belong to another subspecies or genomospecies.

Key Words: AFLP, evolutionary genomic distance, *Frankia* phylogeny, culture recalcitrant, genomospecies.

.

Chapter II: *Frankia* phylogeny depiction and species definition by Amplified Fragment Length Polymorphism (AFLP) and Multilocus Sequence Analysis (MLSA)

INTRODUCTION

Actinorhizal symbiosis is of major importance in the Earth's ecosystem, and accounts for a percentage of terrestrial nitrogen fixation comparable to that done by legume symbioses (Dawson, 2007b). Thanks to their tight association with the actinomycete *Frankia*, actinorhizal plants are pioneer species in disturbed and marginal soils, and inhabit diverse and widespread ecosystems such as forests, shrublands, coastal dunes, glacial deposits, riparian zones, boglands and deserts (Dawson, 2007b; VinuesaValdés, 2007a). For this reason, they are widely used for revegetalization and reforestation purposes.

Frankia spp. the only known genus of the Frankiaceae family, displays very specific traits that make them easy to distinguish from other actinomycetes (Lechevalier, 1986; Lechevalier, 1984b). On the other hand, there are major morphological, physiological and genetic differences within *Frankia* spp. that led to suspect great diversity and the existence of several species within the genus (Benson & Silvester, 1993; Hahn, 2007). The use of DNA-DNA hybridization (DDH) - still the golden standard for delineating bacterial species (Gevers *et al.*, 2005b; Stackebrandt *et al.*, 2002a) – led to the identification of at least 12 genomospecies within *Frankia* spp. (Akimov & Dobritsa, 1992b; An *et al.*, 1985b; Bloom *et al.*, 1989a; Fernandez *et al.*, 1989b; Jamann *et al.*, 1992b). However, the *bona fide* species status of *Frankia* genomospecies was not granted because no discriminative phenotypic traits have been found to date. In addition, more than 50% of *Frankia* specimens are consistently recalcitrant to *in vitro* culture. The diversity of these uncultured strains at the species level is not known since DDH cannot be used, suggesting that the total number of *Frankia* species is far more than twelve.

Analysis of 16S rRNA gene, either from pure strain DNA or composite DNA extracted from whole nodules, showed *Frankia* spp. to be a monophyletic clade that is roughly divided into three major clusters: cluster 1, which groups strains infective on *Alnus*, *Comptonia, Myrica* and *Casuarina*; cluster 2, which groups the non-isolated symbionts of Rosaceae, Datiscaceae, Coriariaceae and *Ceanothus* sp.; cluster 3, which groups strains infective on Elaeagnaceae and Rhamnaceae (Normand *et al.*, 1996). Cluster 1 is itself subdivided into clusters 1a and 1b containing strains infective on *Alnus, Comptonia* or

Myrica, and on *Casuarina*, respectively. The 16S rRNA gene also showed that culture recalcitrant *Alnus*-infective *Frankia* strains called spore-positive (Sp+) - because of their typical ability to synthesize numerous spores *in planta* - formed a divergent group (Simonet *et al.*, 1994). However, the 16S lacks enough resolution for accurate and secure genomospecies identification, since *Frankia* genomospecies displaying as less as 2 to 7% DNA-DNA homology by DDH can share up to 98% of 16S rRNA sequence similarity (Fernandez *et al.*, 1989b; Welsh *et al.*, 2009b). Besides, comparative analysis based on *glnII* and *nifH* sequences - exhibiting more variability than 16S rRNA - confirmed the major divisions of the genus (Clawson *et al.*, 2004b; Gtari *et al.*, 2007). *nifH* has recently been used to assign uncultured *Frankia* to host infection groups and subgroups (Mirza *et al.*, 2009a; Welsh *et al.*, 2009b). However, a study based on a large geographic sampling showed that the *nifH* clustering of *Frankia* does not consistently describe genomic groups (Welsh *et al.*, 2009b), likely because the *nifH* locus might be subjected to lateral transfers (Zehr *et al.*, 2003). As a general feature, all loci may be subject to lateral transfer, and it is established that a single gene phylogeny does not necessarily reflect species phylogeny to whom multi-locus approach are preferable (Gevers *et al.*, 2005b; Nichols, 2001; Vinuesa *et al.*, 2005).

Fingerprinting methods have also been used to provide insights into strain clustering and diversity. Those include both multi-locus and single locus approaches, such as rep-PCR (repetitive sequence amplification), and PCR-RFLP of intergenic regions 16S-23S rRNA and *nifD-K* respectively (Faten *et al.*, 2010; Huguet *et al.*, 2004b; Jeong & Myrold, 1999; Lumini & Bosco, 1996; Lumini & Bosco, 1999; Murry *et al.*, 1995). Those methods were found to be suitable for strain typing, but as a general feature they deliver too limited number of pattern similarities to allow the inference of robust phylogenesis (Meudt & Clarke, 2007a).

This is not the case of AFLP (amplified fragment length polymorphism), which is a PCR-based fingerprinting method, that randomly target numerous loci everywhere in the genome, allowing the deliverance of a sufficiently high number of pattern similarities for robust phylogenetic purposes (Janssen *et al.*, 1996; Vos *et al.*, 1995b). In addition, the method is versatile and can be suited to target genomes. Actually, even if no knowledge of the genome sequence are required for AFLP, complete genome sequences enables *in silico* simulations allowing to optimally estimate the best parameters to obtain maximum genome

coverage with the least homoplasy (Rombauts *et al.*, 2003). In addition, the method can be adapted to produce as many fragments as required for robust analyses; and by using automatic sequencers and fluorescent adaptors, fluorescent AFLP , automatically provides accurate estimations of fragment size (Lindstedt *et al.*, 2000). AFLP is deemed to be an accurate approach in determining phylogenetic and/or taxonomic relationships as well in the assessment of genetic diversity in bacteria (Mortimer & Arnold, 2001a). Indeed, AFLP trials with several taxa showed that evolutionary genomic divergence (EGD) between strains and their common ancestor deduced from AFLP data is particularly useful for phylogenomic purposes (Janssen *et al.*, 1997; Mougel *et al.*, 2002a; Portier *et al.*, 2006a) and that the members of the same genomospecies consistently clustered in coherent groups. Since there is a strong correlation between DDH values and genomic distances estimated with AFLP data (Clerc *et al.*, 1998b; Mougel *et al.*, 2002a; Rademaker *et al.*, 2000a), the international taxonomic committee proposed that molecular traits including AFLP patterns can now also be used to validly distinguish bacterial species (Stackebrandt *et al.*, 2002a). However, AFLP like DDH was thought to be not applicable to target a specific DNA from a complex source (Mortimer & Arnold, 2001a), thus a priori excluding the study of intracellular organisms such as parasites, obligate symbionts and viruses. This point has however not be investigated thoroughly in the case of microsymbionts such as *Frankia*, that display very different genome compositions than their hosts, a molecular feature monitoring the AFLP specificity.

In the present study, we used the AFLP technique to analyze the taxonomic and the phylogenomic relationships among a set of cultured *Frankia* spp. strains. In a second step, the applicability of this approach for *in planta* characterization of endophytic strains by using composite nodule DNA (i.e. root and bacterial) was tested with nodules obtained from experimentally inoculated plants, and then applied to type endophytic *Frankia* present in *Alnus incana* and *Alnus viridis* Sp+ nodules collected in the field. This allowed us to evaluate the suitability of AFLP for genomospecies assignment of both cultured and uncultured strains of *Frankia* spp.

II.2.1. MATERIALS AND METHODS

Frankia strains and DNA extraction

Frankia strains used in this study are listed in Table 1. They were cultured in BAP medium (Murry *et al.*, 1984b) at 28 °C without stirring prior to DNA extraction. Total DNA from roots and from the pure strains was extracted with the DNeasy plant extraction minikit (Qiagen, Hilden, Germany) following the manufacturer's instructions. Field *Alnus* nodules known as hosting strains with the Sp+ phenotype ((Murry *et al.*, 1984b) and our unpublished results) were collected from *A. viridis* and *A. incana* at La Toussuire and Gevoudaz (Isère, France) in the Alp region, with nodules AvToI and AvToII sampled at different altitudes of La Toussuire site. DNA was extracted from nodules following the classical cethyltrimethylammonium bromide (CTAB; Sigma, St-Louis, MO) method with slight modifications (Huguet *et al.*, 2001b).

In silico predictive AFLP

Predictive AFLPs were performed *in silico* with complete genome sequences of *Frankia* strains CcI3, ACN14a and EAN1pec (Accession: CP000249, CT573213 and AAII00000000), with the In-silico program (http://insilico.ehu.es/AFLP/Amplify.php) to determine the best AFLP parameters to analyze *Frankia* spp. For this purpose, we preferentially focused on endonucleases that recognize high G+C content, and that were expected to provide well resolved patterns in the range of 40-450 pb when combined with relevant selective nucleotides. To limit homoplasy (i.e. fragments of identical length but originating from different genome regions) that would unduly increase similarities, special care was taken to select AFLP conditions that predict the occurrence of well separated fragments with only one predicted fragment for a given molecular length.

Fluorescent AFLP

The AFLP protocol used in this study was adapted from Vos et al. (Vos *et al.*, 1995a) with some modifications, by using the endonuclease combinations PstI-TaqI and PstI-HhaI,

Table 1. Origins of *Frankia* strains used in this study

Trivial designation	Registry no.	Original host	Geographic origin	Ref.*	Genomospecies		AFLP Group
					Original number	This study	
***Alnus* and *Myrica* Cluster**							
32-44(Aid1)	LLR01322	*Alnus incana ssp. rugosa*	Vermont, USA	[4]	1	G1	G1
ACN1^{a,o}	ULQ0102001007	*Alnus crispa*	Tadoussac, Canada	[4, 12]	1	G1	G1
AcoN24d	ULP01010244	*Alnus cordata*	Orleans, France	[12]	1	G1	G1
AcVcl	ULP010102171	*Alnus cordata*	Corsica, France	[12]	1	G1	G1
Ar24H3	ULP0131024083	*Alnus rubra*	Orléans, France	[12]	1	G1	G1
Ar24H5		*Alnus rubra*	Orleans, France	-	1	G1	G1
ARgN22d	ULQ0132022024	*Alnus rugosa*	Québec, Canada	[12]	1	G1	G1
AvcI1	DDB0 10110	*Alnus viridis ssp. crispa*	Ontario, Canada	[4, 6, 12]	1	G1	G1
CpI1	HFP07010701	*Comptonia peregrina*	Massachusetts, USA	[4, 6, 12]	1	G1	G1
MPI1	LLR162001	*Myrica pensylvanica*	Nantucket, VT, USA	[4]	1	G1	G1
M16467	RBR162021	*Myrica pensilvanica*	NewJersey, USA	[6]	1	G1	G1
ACN14a	ULQ010201401	*Alnus crispa*	Tadoussac, Canada	-	ND	G1	G1
Ag24$_{21}$	ULP0107024251	*Alnus glutinosa*	Orléans, France	[12]	ND	G1	G1
ArI3	HFP013103	*Alnus rubra*	Oregon, USA	[12]	ND	G1	G1
I38		*Alnus incana*	Lapalud, France	-	ND	G1	G1
M16477		*Myrica pensilvanica*	NewJersey, USA	-	ND	G1	G1
Ac23$_{a}$	ULP010102340	*Alnus crispa*	Orleans, France	[12]	2	G2	G2
AVL3	ULP014102203	*Alnus viridis*	Lautaret, France	[12]	2	G2	G2
AVN17o	ULP014101715	*Alnus viridis*	LaToussuire, France	[12]	2	G2	G2
ARgP5^{ag}	ULQ0132105009	*Alnus rugosa*	Quebec, Canada	[12]	3	G3	G3
***Casuarina* cluster**							
BR	ORS 020608	*Casuarina equisetifolia*	Brazil	[12]	9	G4	G4
Cd13	HFP020203	*Casuarin cunninghamiana*	Florida,USA	[12]	9	G4	G4
CjI-82	ORS 021001	*Casuarina junghuniana*	Thailand	[12]	9	G4	G4
TA	ORS 022602	*Allocasuarina torulosa*	Australie	[12]	9	G4	G4
Cg70$_{1}$		*Casuarina glauca*	India	-	ND	G4	G4
Cg70$_{5}$		*Casuarina glauca*	India	-	ND	G4	G4
Cg70$_{2}$		*Casuarina glauca*	India	-	ND	ND	G4
CcI2	HFP020202	*Casuarina cunninghamiana*	Florida, USA	[12]	ND	G5	G5
Elaeagnaceae- Rhamnaceae Cluster							
Ea1-12	ULF130100112	*Elaeagnus angustifolia*	Ecully, France	[12]	4	G6	G6
Ea1$_{2}$	ULF130100102	*Elaeagnus angustifolia*	Ecully, France	[12]	4	G6	G6
Ea48$_{6}$		*Elaeagnus angustifolia*	France	[17]	4	G6	G6
Ea7$_{1}$		*Elaeagnus angustifolia*	Toulon, France	[17]	4	G6	G6
EaCm5$_{1}$	ULF130100501	*Elaeagnus angustifolia*	Escarène, France	[12]	4	G6	G6
	ORS060501	*Colletia spinosissima*	Argentina	[12]	ND	G6	G6
Hr75$_{2}$		*Hippophaë rhamnoides*	France	-	ND	G6	G6
CH37	ORS140102	*Hippophaë rhamnoides*	Nogent/Marne, France	-	ND	ND	G9
GFN14a	GFN140101	*Hippophaë rhamnoides*	China	[12]	ND	ND	G10
Atypical strains							
MgI$_{5}$	DDB16110210	*Myrica gale*	NewYork, USA		ND	ND	G11
Pti1	DDB170110	*Purshia tridentata*	Wyoming, USA	[37]	3	G7	G7
G$_{2}$	ORS020604	*Casuarina equisetifolia*	Guadeloupe	[4]	2	G8	G8

*References where the genomospecies here studied were reported for the first time.

adaptors and PCR primers given Table 2. AFLPs were performed with 55 ng of DNA for bacterial pure cultures, and 150 ng for complex DNA extracted from nodules.

For PstI-HhaI conditions, digestion and ligation steps were performed simultaneously by incubating 2 h at 37°C in 11 µL (final volume): the target DNA, PstI (5 U), HhaI (5 U), T4 DNA ligase (1 U), 1X T4 DNA ligase buffer (Boeringer-Manheim, Germany), 0.5 µg of bovine serum albumin, 50 µM NaCl, PstI-specific and HhaI-specific adaptors at 0.18 µM and 1.8 µM, respectively.

For PstI-TaqI conditions, digestions and ligation were performed in two steps because of differences in enzyme optimal temperatures. The digestion with TaqI was performed first at 65°C for 2h, followed by a digestion for 2 h at 37°C with PstI. The adaptor ligation was performed with 55 ng of the digested DNA as described above.

Non-selective and selective PCR amplification

Adapted DNAs (4 µL) were subjected to nonselective PCR performed in a 20 µL (final volume) mixture containing 15 µL of the 'AFLP amplification CoreMix' (Perkin-Elmer Applied Biosystems, Foster City, Calif.), 0.25 µM primer F2758-corePst, and 0.25 µM primer F3881-coreHha for the PstI-HhaI set, or 0.25µM primer F3798-coreTaq, for the PstI-TaqI set. Non-selective PCR reactions were done in a PE-9600 thermocycler (Perkin-Elmer) with: initial denaturation at 94°C for 5 min, and 20 cycles of denaturation at 94°C for 20 s, annealing at 56°C for 30 s, and elongation at 72°C for 2 min. The quality of nonselective PCRs was checked by visual inspection of 1% agarose gel electrophoreses. A good nonselective PCR typically displayed a lot of fragments ("smear") between 200 bp and 600 bp.

Nonselective PCR products were diluted 1/50 before used as template (1.5 µL) in selective PCR mixtures (final volume, 10 µL) with 7.5 µL of 'AFLP amplification CoreMix', 0.25 µM of non labeled primer (F3881-coreHha or F3798-coreTaq), and 0.05 µM of fluorescently labeled primers (Invitrogen). Fluorescently labeled primers were designed with the F2758-corePst sequence plus two discriminating nucleotides at the 3' end, F3804-PstAA-FAM and F3805-PstAC-HEX. Selective PCRs were performed using a touchdown procedure

consisting of denaturation at 94°C for 20 s, annealing at temperatures ranging from 66 to 57°C (the temperature was decreased 1°C per cycle) for 30 s, and elongation at 72°C for 2

Table 2. AFLP oligonucleotides used to construct adaptors and to prime PCRs

Oligonucleotide	Sequence
PstI-specific adaptors	
F1604-adPst+	CTCGTAGACTGCGTACATGCA
F2757-adPst-	CATCTGACGCATGT
TaqI-specific adaptors	
F1365-adTaq+	GACGATGAGTCCTGAG
F3878adTaq-	CGCTCAGGACTCAT
Hha-specific adaptors	
F3879-adHha+	GACGATGAGTCCTGACCG
F3880-adHha-	GTCAGGACTCATC
Core primers	
F2758-corePst	GACTGCGTACATGCAG
F3798-coreTaq	GATGAGTCCTGAGGCA
F3881-coreHha	GATGAGTCCTGACCGC
Selective Primers	
F3804-PstAA-FAM	GACTGCGTACATGCAGAA
F3805-PstAC-HEX	GACTGCGTACATGCAGAC

min for 10 cycles, followed by a conventional PCR consisting of denaturation at 94°C for 20 s, annealing at 56°C for 30 s, and elongation at 72°C for 2 min for 20 cycles.

Total selective PCR products (10 µL) were purified on a Sephadex G-50 column (Amersham Biosciences, Orsay, France) before separation of the AFLP fragments by electrophoresis with a capillary sequencer (MegaBACE 1000; Amersham Pharmacia Biotech Europe, Orsay, France). The device calculated automatically the length of the fluorescent fragments by comparison to the MegaBACE ET-900-R size standard (Amersham). A genetic profiler (model 1.5; Molecular Dynamics Inc., Sunnyvale, Calif.) was used to display the results and to export data in text format. Data were transferred to a spreadsheet with the program Thresholdfilter 1.3 (Yann Legros, Amersham) as previously described (Portier *et al.*, 2006b). A threshold fluorescence value of 100 arbitrary units was used to eliminate the background before subsequent scoring and bioinformatic treatments.

Phylogenomic analysis

The Align2, Lis and LecPCR programs were used to compare duplicated patterns, to optimize fragment assignment in length classes and to transform raw data into tabular binary data, respectively. The DistAFLP program was used to calculate the current genome mispairing (CGM) using Jaccard index and the evolutionary genomic divergence (EGD), expressed as the number of nucleotide substitution per site (nsps), using Dice index, by using 12 as number of nucleotides involved in restriction and selective PCR steps as described by Mougel et al. (Mougel *et al.*, 2002a). The LecPCR, Align2, and DistAFLP programs are accessible at http://pbil.univ-lyon1.fr/ADE-4/microb/. Dendrograms were calculated with the Neighbour/UPGMA module of the PHYLIP package. Bootstrap values were calculated by using the bootstrap option (1000 replicates) of DistAFLP and the Neighbour/UPGMA and Consense modules of the PHYLIP package were used to draw dendrograms with NJ-Plot (http://pbil.univ-lyon1.fr/software/njplot.html).

II.2.2. RESULTS and DISCUSSION

AFLP conditions and reliability. All enzyme-primer combinations tested provided distinct and easily readable AFLP patterns for all *Frankia* strains tested (data not shown). Average numbers of fragments per strain were 60.7 ± 6.5, 58.6 ± 6.9, 61.3 ± 8.4 and 61.4 ± 8.7 for PstI+AA/TaqI+0, PstI+AC/TaqI+0, PstI+AA/Hha+0 and PstI+AC/Hha+0, respectively, for a pooled total of 242.0 ± 21.7. This number varied considerably according to 16S-based clusters and/or host infectivity groups with 249 +/- 10.1, 206 +/- 10.6 and 259 +/- 17.7 for cluster 1a, 1b and 3, respectively, corresponding to host infectivity groups *Alnus-Myrica*, *Casuarina* and Elaeagnaceae, respectively.

Duplicate or triplicate samples produced more than 95% and 93% shared fragments, respectively (data not shown). The AFLP method thus allowed us to reliably found pattern similarities between strains and in turn to confidently calculate genomic distances between different strains. However, the method was found to be not reliable enough to deliver absolutely identical patterns with a given strain (data not shown). This phenomenon has already been observed in other taxa (Clerc *et al.*, 1998b; Costechareyre *et al.*, ; Portier *et al.*, 2006a). Latter authors suggested this unreliability could be related to epigenetic determinants causing differential accessibilities of genome regions to endonucleases (Portier *et al.*, 2008). As a result, this known limitation of the AFLP technology does not allow us to definitively assign *Frankia* isolates to given strains or clone categories.

Phylogenomic analysis of cultured Frankia spp. Percentages of polymorphic fragment (i.e. presence or absence) computed on the whole strain set were 61.2%, 54.7%, 67.8% and 63.4% for PstI+AA/TaqI+0, PstI+AC/TaqI+0, PstI+AA/Hha+0 and PstI+AC/Hha+0 conditions, respectively. This allowed us to calculate pair-wise genomic distances expressed as CGM and EGD (data not shown). Phylogenomic dendrograms drawn with EGD data calculated with each enzyme-primer combination were very similar (data not shown). Thus, EGD data were pooled to infer a more robust phylogeny with UPGMA (Fig. 1), which was almost identical with Neighbor Joining (data not shown).

Assignment of cultured Frankia strains to genomospecies

The phylogeny revealed that isolates found to belong to the same genomospecies by DDH always gathered in significant AFLP clusters (Fig 1). This was the case for

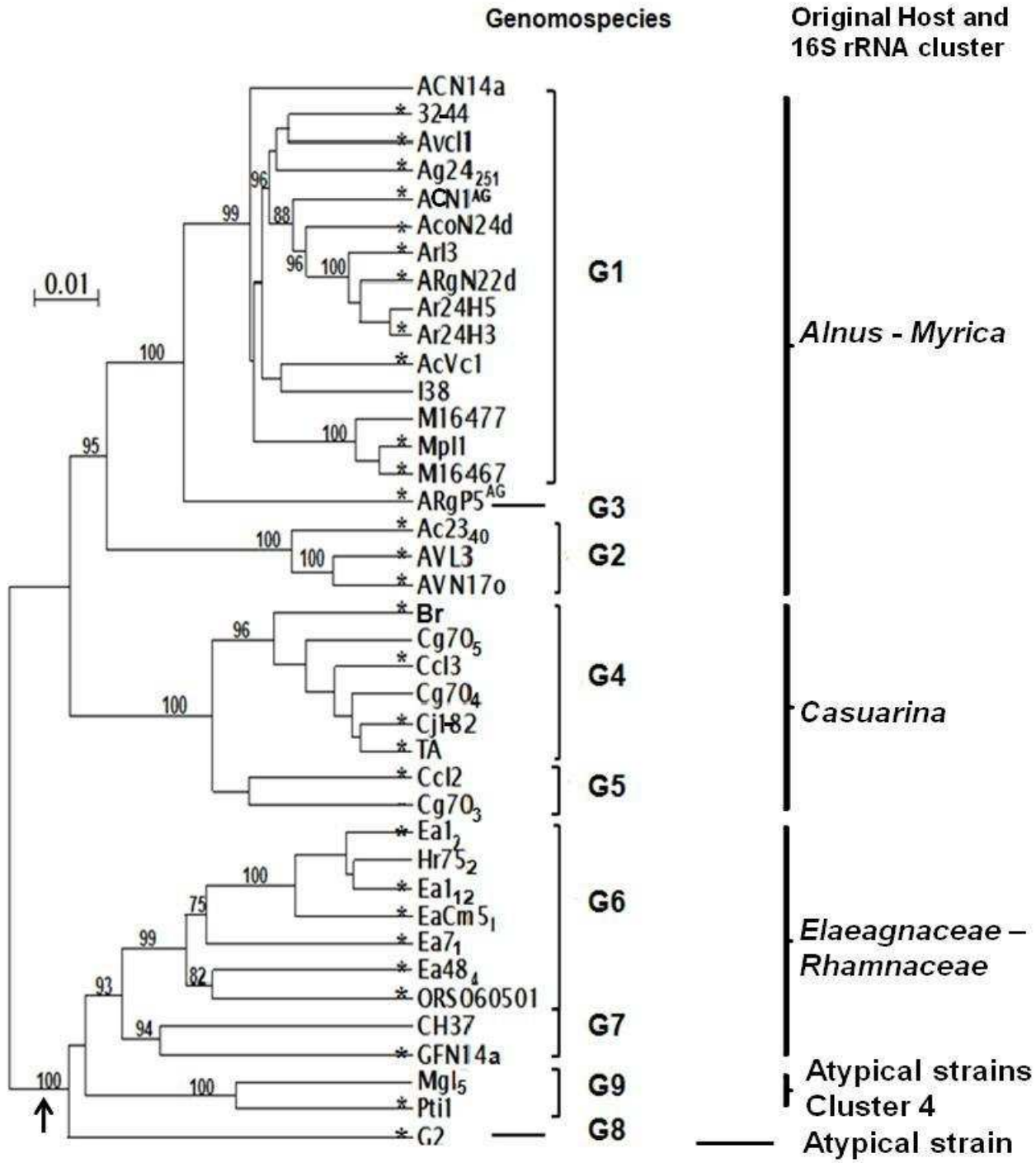

Fig. 1 UPGMA dendrogram based on the EGD calculated from AFLP data analysis of *Frankia* strains with four conditions. Numbers indicate significant bootstrap values obtained from 1000 data resamplings. * highlights strains for which DDH values was available from elsewhere. Arrow indicates significant gathering of atypical and standard Eleagnaceae-Rhamnaceae infective strains at the deep branching level.

genomospecies G1, G2, G4 and G6 with 10, 3, 4 and 5 isolates already tested by DDH, respectively (stars in Fig. 1). Conversely, isolates found to belong to different genomospecies by DDH analyses always gathered in different AFLP clusters. This clearly demonstrates that AFLP is a suitable method to unambiguously delineate *Frankia* genomospecies as previously determined by DDH.

In the present analysis, other isolates that were not analyzed previously by DDH, grouped together with DDH tested strains. Indeed, we confidently assumed that they readily belong to the same genomospecies that their analyzed neighbors. As a result, this allowed us to infer that sequenced strain ACN14a belongs to genomospecies G1.

AFLP threshold values for Frankia species definition. We questioned the possibility to define a threshold of genomic distances that would be a cutoff value for genomospecies delineation. For this purpose, we plotted AFLP based estimates of genome mispairing (CGM) against DDH re-association values. A strong and significant linear relationship (r^2= 0.87) was obtained. However, as displayed Fig. 2, the 60% DDH threshold - we previously found appropriate to delineate *Frankia* genomospecies based on ΔTm considerations (Akimov & Dobritsa, 1992b; An *et al.*, 1985b; Fernandez *et al.*, 1989b; Jamann *et al.*, 1992b) - did not correspond to a unique CGM cutoff value. More precisely, the lowest value of genomospecies pairwise CGM was 0.080 mismatch per site between G1 and G3, while the largest CGM value found within a single genomospecies was in G6, 0.098 mismatch per site.

Distinctive cutoffs can be found according to major phylogeny clades (table 3). Actually, it appeared that the value was between 0.071 and 0.080 mismatch per site in *Alnus-Myrica* and *Casuarina*-infecting genomospecies (16S cluster 1), while, it was over 0.098 and below 0.105 mismatch per site in Eleagnaceae-Rhamnaceae infecting genomospecies (e.g. 16S clusters 3 and 4). This cutoff discrepancy was revealed as well in phylogenetic dendrograms inferred from EGD values (Fig. 1). Cutoff differences for genomospecies definition were also found with the *nifH* molecular marker (Welsh *et al.*, 2009b).

Inconsistency of genomic distance cutoff over the *Frankia* genus could be related to dramatic differences in genome size between strains that belong to different host infecting

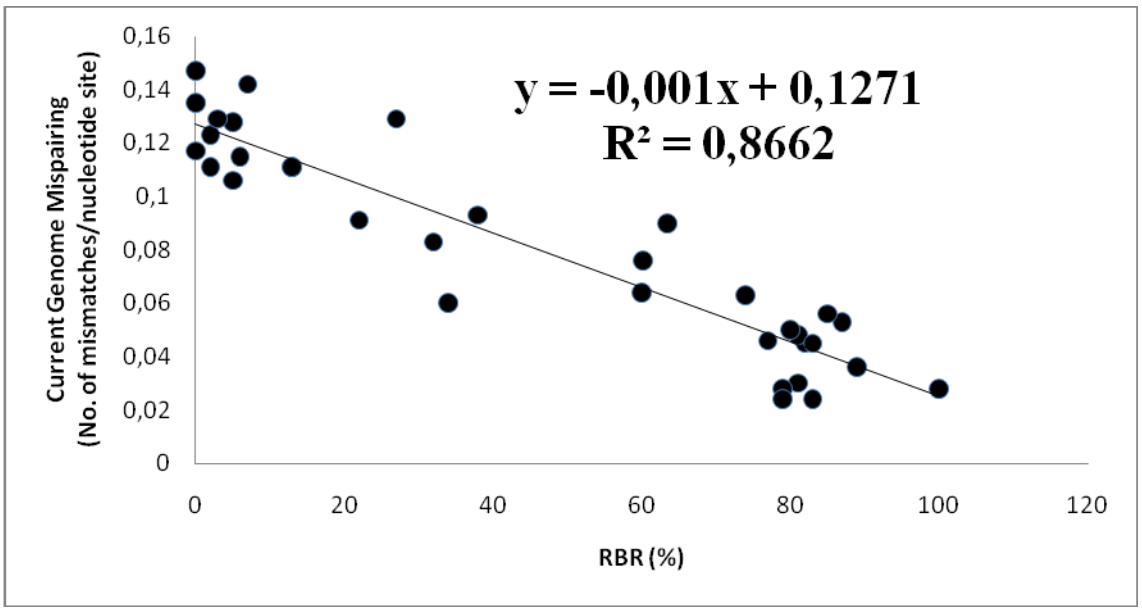

Fig. 2 Correlation between current genomic mispairing values (CGM) calculated from the concatenation of the four AFLP conditions and the DNA reassociation rates (RBR%).

Table 3. Current Genome Mispairing (CGM) limits for Frankia genomospecies delineation

Host-infectivity group	*Alnus-Myrica*			*Casuarina*		Elaeagnaceae		
Genomospecies*	G1	G2	G3	G4	G5	G6	G7	G8
Maximal intraspecies CGM	0,071	0,05	-	0,049	-	0,098	-	-
Minimal Interspecies CGM	0,080	0,102	0,080	0,083	0,083	0,110	0,105	0,107

*In this table are summarized only the genomospecies for which DDH values are available.

groups. Actually, ACN14a (G1, *Alnus-Myrica* infecting group), CcI3 (G4, *Casuarina* infecting group) and EaN1pec (G6, Eleagnaceae-Rhamnaceae infecting group) have genome sizes of 7.5, 5.4 and 9.03 Mb, respectively (Normand *et al.*, 2007b). It has been estimated that large genomes generate more numerous and more specific AFLP fragments than small ones (Fay *et al.*, 2005; Garcia-Pereira *et al.*, 2010; Meudt & Clarke, 2007a), while strains with reduced genomes that must keep essential and thus conserved genes would generate less strain specific AFLP fragments. As a result, it is likely that genomospecies with large genomes such as members of the Eleagnaceae-Rhamnaceae infecting group must deliver more strain specific fragments and thus must display much larger genomic distances between their members than groups with smaller genomes such as members of the *Casuarina* infecting group.

Providing the occurrence of different genomic distance cutoff values, we assumed it is possible to use AFLP data to assign novel strains to their respective genomospecies. As a result, *Casuarina* infecting strains CcI2 and $Cg70_3$ that display a CGM of 0.068 mismatch per site - thus below 0.071, the largest infra-specific CGM recorded in the 16S cluster 1 - were thought to belong to the same genomospecies G5. In the Eleagnaceae-Rhamnaceae infecting group, GFN14 and CH37 are distantly related (CGM = 0.094 mismatch per site), but still in the range of infra-specific distances of the group. It is thus likely they are members of the same genomospecies G9.

In addition, with a CGM value of 0.071 mismatch per site below all cutoff values, atypical strains MgI_5 and Pti1 - that were grouped in 16S cluster 4 by Huguet et al. (2001) - could confidently be assigned to the same genomospecies G7 by the present AFLP study.

Deep branchings of Frankia phylogeny. Remarkably, contrary to what was found with several other taxa (Bussel *et al.*, 2005; Koopman, 2005; Portier *et al.*, 2006a), bootstrap values strongly support the deep branchings of the *Frankia* phylogenetic tree obtained with AFLP data. In all instance, the phylogeny significantly grouped standardly infective *Frankia* strains according to their actual host infectivity groups *Alnus-Myrica*, *Casuarina*, and Elaeagnaceae-Rhamnaceae. The present results show that it was also the case for atypical strains G2 and MgI_5 unable to re-infect their original *Casuarina* and *Myrica*, respectively, but that were found to be able to infect Elaeagnaceae. Those readily Elaeagnaceae infective but

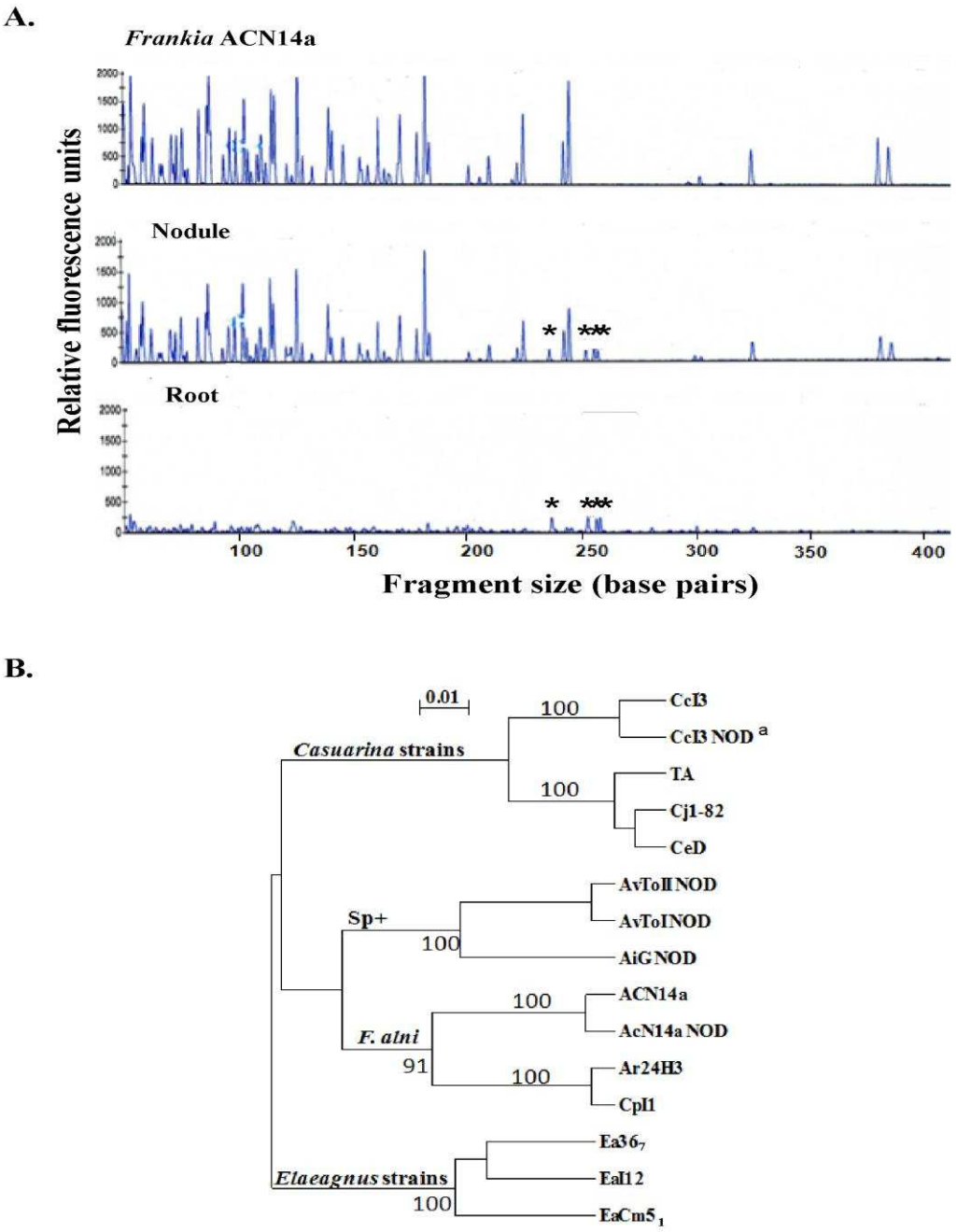

Fig. 3 A. AFLP- electrophoretic profiles yielded with PstIAA/HhaI+0 for AcN14a strain, composite DNA from nodules induced by the same strain and root *Frankia*-free DNA. Asterisks indicate peaks shared by nodule and root AFLP profiles. B. UPGMA dendrogram calculated from EGD values given by the AFLP analysis of cultured and uncultured *Frankia* strains. Numbers indicate significant bootstrap values for 1000 resamplings.

still atypical strains grouped coherently in the same major AFLP cluster gathering standard Elaeagnaceae-Rhamnaceae infecting strains (arrow, Fig. 1).

Superimposed to host infectivity groups, the AFLP tree grouped *Frankia* strains according to previously defined 16S clusters 1a, 1b, 3 and 4 (Fig. 1). Noticeably, cluster 4 (i.e. genomospecies G9) appeared here as a sister clade of cluster 3. This feature was not supported - nor in contradiction indeed - by 16S phylogeny results because the 16S marker is not resolutive enough to support significant bootstrap values at this level(Huguet *et al.*, 2001b). However, this feature fits with the assignment of MgI_5 to the Elaeagnaceae and Rhamnaceae-infective group, also obtained by using the symbiotic *nifH* gene (Welsh *et al.*, 2009b). It is thus worth noting the place of the atypical strain Pti1 (G9) together with MgI_5. Pti1 isolated from *Purshia tridentata* is known to be unable to re-infect any host (Nazaret *et al.*, 1991), likely because it is a defective strain lacking all symbiotic *nif* genes (Nazaret *et al.*, 1991). The actual position of Pti1 in G9 close to MgI_5, shows that Pti1 has a similar genomic background than MgI_5 and likely than most true Elaeagnaceae and Rhamnaceae-infective strains. A fact that could not be confirmed by using *nifH* for evident reasons.

The actual phylogenetic position of G9/cluster 4 members within or close to the true Elaeagnaceae and Rhamnaceae-infective group remains however yet questionable, because this position may result from artifact in tree reconstruction at the deep branching level caused by the well known long branch attraction phenomenon (Bergsten, 2005). Actually, as explained above, larger genomes produce more fragments that cause in turn longer branches than smaller ones. It can't thus be excluded that deep branchings may at least partly be induced by genome size differences rather than by true genomic similarities. This clearly indicates the need for additional and thorough genomic investigations to definitively solve the *Frankia* phylogeny at the deepest levels.

Genomospecies assignment of uncultivated Frankia present in nodules. As numerous *Frankia* members are only available in nodules because they are up to now recalcitrant to in vitro culture, they can't be assigned to genomospecies by DDH because contaminant plant DNA would hamper precise analyses. We thought that AFLP could be used instead because this methodology targets precise DNA sequences (i.e. restriction sites flanked

of selective nucleotide), allowing us to discriminate plant and bacterium-derived AFLP fragments and thus to overcome the occurrence of plant DNA contaminant in nodules.

Thus in a first step, we performed AFLP analyses on DNA extracted: (i) from nodules infected by known strains; (ii) from the corresponding pure culture strains; and (iii) from *Frankia*-free plant roots. In all cases but one, no noise from plant DNA was detected in the nodular DNA patterns. By using the selective AFLP condition PstI+AA/HhaI+0, only four fragments attributable to plant DNA appeared in the AFLP pattern obtained from nodular DNA, accounting for 1.3% more fragments than obtained with a *Frankia* pure culture (Fig. 3a). This weak noise that didn't interfere significantly in resulting phylogeny can be explained by very different G+C content between plant and *Frankia* genome (i.e about 44% and 70% for *Alnus glutinosa* and *Frankia* respectively). Actually, AFLPs performed with DNAs from nodules and their pure culture inciting strains yielded AFLP patterns with 0.020 and 0.032 mismatch per site between ACN14aNOD and ACN14a, and CcI3NOD and CcI3, respectively. They thus positioned very close together in phylogenomic dendrograms (Fig. 3b). Noticeably, such a small difference is in the range of unreliability reported above between AFLP duplicates done with the same strain, showing that the AFLP methodology is relevant to validly assess the phylogenomic position of *Frankia* present in nodules.

In a second step, we used the methodology to define the phylogenomic position of culture recalcitrant *Frankia.* In this experiment, we analyzed field nodules of *Alnus viridis* and *A. incana* hosting endophytic *Frankia*, which displayed a typical Sp+ phenotype known to repeatedly fail to culture in vitro (Moiroud, personal communication). AFLP patterns of the three Sp+ endophytes were found to be closely related, thus likely belonging to the same genomospecies (greatest CGM = 0.046 mismatch per site). These endophytes appeared highly related but likely different of genomospecies G1. Actually, CGM values were generally in a range indicating an out position (0.085 to 0.120 mismatch per site) except with ACN14a and the AiGNOD endophyte (0.066 mismatch per site). These results suggest that *Alnus* Sp+ strains have notably diverged from genomospecies G1 and may belong to another species or subspecies among *Alnus* strains.

Conclusion. As already shown in several other genera including *Burkholderia*, *Xanthomonas* or *Agrobacterium* (Coenye *et al.*, 1999; Portier *et al.*, 2006a; Rademaker *et al.*,

2000a), AFLP was presently revealed as a suitable alternative to DDH for the assignment of strains to known or new genomospecies within the whole *Frankia* genus. It allows us to assign new isolates to previously defined genomospecies. Moreover, the method also allows the genomospecies determination of culture recalcitrant ones. As a consequence, it is now thinkable to draw a complete phylogenomy of the genus. The present data also still confirm, but at the genomic level, that co-evolution with hosts strongly shaped the whole phylogeny of *Frankia* spp. as suspected by various authors using single locus analyses (Benson & Dawson, 2007b; Simonet *et al.*, 1999). This co-evolutive shaping was yet confirmed for culture recalcitrant Sp+ that infect *Alnus* plants, that were found in the present study typically included in the *Alnus*-*Myrica* infecting group. Finally, the present study revealed the true membership of the sequenced strain ACN14a to genomospecies G1, which is definitively the bona fide species *Frankia alni*. To our knowledge, this is the first time that the AFLP method has been successfully applied to type and identify bacterial genomospecies from complex samples. The use of AFLP directly on nodule DNA paves the way for describing genomospecies within the major groups of non-isolated *Frankia* strains infecting *Dryas*, Coriariaceae, Datiscaceae and *Ceanothus*.

Acknowledgements – Thanks are expressed to D. Chapulliot (UMR555-Ecologie Microbienne, Lyon, France) for his technical support. Hector H. Bautista Guerrero had a grant from CONACYT and M. Antonia Cruz was supported by a Mexico-France ECOS program. Part of this work was funded by the CONACyT grant 48044 to M. Valdes and by la Région Rhône-Alpes (Cluster 9 program).

REFERENCES

[1] V.N. Akimov, S.V. Dobritsa, Grouping of *Frankia* strains on the basis of DNA relatedness, Syst. Appl. Microbiol. 15 (1992) 372-379.

[2] C.S. An, W.S. Riggsby, B.C. Mullin, Relationships of *Frankia* isolates based on deoxyribonucleic acid homology studies, Int. J. Syst. Bacteriol. 35 (1985) 140-146.

[3] D.R. Benson, J.O. Dawson, Recent advances in the biogeography and genecology of symbiotic *Frankia* and its host plants, Physiol. Plant. 130 (2007) 318-330.

[4] D.R. Benson, W.B. Silvester, Biology of *Frankia* strains, actinomycete symbionts of actinorhizal plants, Microbiol. Rev. 57 (1993) 293-319.

[5] J. Bergsten, A review of long-branch attraction, Cladistics 21 (2005) 163-193.

[6] R.A. Bloom, B.C. Mullin, R.L. Tate, DNA restriction patterns and DNA-DNA solution hybridization studies of *Frankia* isolates from *Myrica pensylvanica* (bayberry), App. Environ. Microbiol. 55 (1989) 2155-2160.

[7] J.D. Bussel, M. Waycott, J.A. Chappill, Arbitrarily amplified DNA markers as characters for phylogenetic inference, Perspect. Plant Ecol. Evol. Syst. 7 (2005) 3-26.

[8] M.L. Clawson, A. Bourret, D.R. Benson, Assessing the phylogeny of *Frankia*-actinorhizal plant nitrogen-fixing root nodule symbioses with *Frankia* 16S rRNA and glutamine synthetase gene sequences, Mol. Phylogenet. Evol. 31 (2004) 131-138.

[9] A. Clerc, C. Manceau, X. Nesme, Comparison of Randomly Amplified Polymorphic DNA with Amplified Fragment Length Polymorphism To Assess Genetic Diversity and Genetic Relatedness within Genospecies III of *Pseudomonas syringae*., Appl. Environ. Microbiol. 64 (1998) 1180-1187.

[10] T. Coenye, L.M. Schouls, J.R. Govan, K. Kersters, P. Vandamme, Identification of *Burkholderia* species and genomovars from cystic fibrosis patients by AFLP fingerprinting, Int. J. Syst. Bacteriol. 49 Pt 4 (1999) 1657-1666.

[11] D. Costechareyre, A. Rhouma, C. Lavire, P. Portier, D. Chapulliot, F. Bertolla, A. Boubaker, Y. Dessaux, X. Nesme, Rapid and Efficient Identification of Agrobacterium Species by recA Allele Analysis : Agrobacterium recA Diversity, Microb. Ecol. (2010).

[12] J.O. Dawson, Ecology of actinorhizal plants, In: K. P. a. W. E. Newton (Ed.), Nitrogen-fixing Actinorhizal Symbioses, Springer, Dordrecht, The Netherlands, 2007, pp. 199-227

[13] G. Faten, I. Nouioui, M. Chair, A. Boudabous, M. Gtari, 16S-23S rRNA Intergenic spacer region variability in the genus *Frankia*, Microb. Ecol. In press (2010).

[14] M.F. Fay, R.S. Cowan, I.J. Leitch, The effects of nuclear DNA content (C-value) on the quality and utility of AFLP fingerprints, Ann. Bot. 95 (2005) 237-246.

[15] M.P. Fernandez, H. Meugnier, P.A.D. Grimont, R. Bardin, Deoxyribonucleic acid relatedness among members of the genus *Frankia*, Int. J. Syst. Bacteriol. 39 (1989) 424-429.

[16] M.J. Garcia-Pereira, A. Caballero, H. Quesada, Evaluating the relationship between evolutionary divergence and phylogenetic accuracy in AFLP data sets, Mol. Biol. Evol. 27 (2010) 988-1000.

[17] D. Gevers, F.M. Cohan, J.G. Lawrence, B.G. Spratt, T. Coenye, E.J. Feil, E. Stackebrandt, Y. Van de Peer, P. Vandamme, F.L. Thompson, J. Swings, Opinion: Re-evaluating prokaryotic species, Nat. Rev. Microbiol. 3 (2005) 733-739.

[18] M. Gtari, L. Brusetti, A. Hassen, D. Mora, D. Daffonchio, A. Boudabous, Genetic diversity among *Elaeagnus* compatible *Frankia* strains and sympatric-related nitrogen-fixing actinobacteria revealed by *nifH* sequence analysis, Soil Biol. Biochem. 39 (2007) 372-377.

[19] D. Hahn, Polyphasic taxonomy of the genus Frankia, Nitrogen-fixing Actinorhizal Symbiosis. In K. Pawlowski and W. E. Newton (Eds) (2007) 25-47.

[20] V. Huguet, J.M. Batzli, J.F. Zimpfer, P. Normand, J.O. Dawson, M.P. Fernandez, Diversity and specificity of *Frankia* strains in nodules of sympatric *Myrica gale, Alnus incana*, and *Shepherdia canadensis* determined by rrs gene polymorphism, Appl. Environ. Microbiol. 67 (2001) 2116-2122.

[21] V. Huguet, M. Mergeay, E. Cervantes, M.P. Fernandez, Diversity of *Frankia* strains associated to *Myrica gale* in Western Europe: impact of host plant (*Myrica* vs. *Alnus*) and of edaphic factors, Environ. Microbiol. 6 (2004) 1032-1041.

[22] S. Jamann, M.P. Fernandez, A. Moiroud, Genetic diversity of Elaeagnaceae-infective *Frankia* strains isolated from various soils, Acta Oecolog. 13 (1992) 395-405.

[23] P. Janssen, R. Coopman, G. Huys, J. Swings, M. Bleeker, P. Vos, M. Zabeau, K. Kersters, Evaluation of the DNA fingerprinting method AFLP as a new tool in bacterial taxonomy, Microbiology 142 (1996) 1881-1893.

[24] P. Janssen, K. Maquelin, R. Coopman, I. Tjernberg, P. Bouvet, K. Kersters, L. Dijkshoorn, Discrimination of *Acinetobacter* genomic species by AFLP fingerprinting, Int. J. Syst. Bacteriol. 47 (1997) 1179-1187.

[25] S.-C. Jeong, D.D. Myrold, Genomic fingerprinting of *Frankia* microsymbionts from *Ceanothus* copopulations using repetitive sequences and polymerase chain reactions, Can. J. Bot. 77 (1999) 1220-1230.

[26] W.J. Koopman, Phylogenetic signal in AFLP data sets, Syst. Biol. 54 (2005) 197-217.

[27] M. Lechevalier, Catalog of *Frankia* strains, The Actinomycetes 19 (1986) 131-162.

[28] M.P. Lechevalier, The taxonomy of the genus *Frankia*, Plant Soil 78 (1984) 1-6.

[29] B.A. Lindstedt, E. Heir, T. Vardund, G. Kapperud, Fluorescent amplified-fragment length polymorphism genotyping of Salmonella enterica subsp. enterica serovars and comparison with pulsed-field gel electrophoresis typing, J. Clin. Microbiol. 38 (2000) 1623-1627.

[30] E. Lumini, M. Bosco, PCR-restriction fragment length polymorphism identification and host range of single-spore isolates of the flexible *Frankia* sp. strain UFI 132715., Appl. Environ. Microbiol. 62 (1996) 3026-3029.

[31] E. Lumini, M. Bosco, Polymerase chain reaction - restriction fragment length polymorphisms for assessing and increasing biodiversity of *Frankia* culture collections., Can. J. Bot. 77 (1999) 1261-1269.

[32] H.M. Meudt, A.C. Clarke, Almost fogotten or latest practice? AFLP applications, analyses and advances, Trends Plant Sci. 12 (2007) 106-117.

[33] B.S. Mirza, A. Welsh, J.P. Rieder, M.W. Paschke, D. Hahn, Diversity of frankiae in soils from five continents, Syst. Appl. Microbiol. 32 (2009) 558-570.

[34] P. Mortimer, C. Arnold, FAFLP: last word in microbial genotyping ?, J. Med. Microbiol. 50 (2001) 393-395.

[35] C. Mougel, J. Thioulouse, G. Perriere, X. Nesme, A mathematical method for determining genome divergence and species delineation using AFLP, Int J Syst Evol Microbiol 52 (2002) 573-586.

[36] M.A. Murry, M.S. Fontaine, J.G. Torrey, Growth kinetics and nitrogenase induction in *Frankia* sp. HFPArI3 grown in batch culture, Plant Soil 78 (1984) 61-78.

[37] M.A. Murry, D. Zhang, M. Schneider, F.J. de Bruijn, Use of repetitive sequences and the polymerase chain reaction to fingerprint the genomes of *Frankia* isolates., Symbiosis 19 (1995) 223-240.

[38] S. Nazaret, B. Cournoyer, P. Normand, P. Simonet, Phylogenetic relationships among *Frankia* genomic species determined by use of amplified 16S rDNA sequences., J. Bacteriol. 173 (1991) 4072-4078.

[39] R. Nichols, Gene trees and species trees are not the same, Trends Ecol. Evol. 16 (2001) 358-364.

[40] P. Normand, P. Lapierre, L. Tisa, J.P. Gogarten, N. Alloisio, E. Bagnarol, C.A. Bassi, A.M. Berry, D.M. Bickhard, N. Choisne, A. Couloux, B. Cournoyer, S. Cruveiller, V. Daubin, N. Demange, M.P. Francino, E. Goltsman, Y. Huang, O. Kopp, L. Labarre, A. Lapidus, C. Lavire, J. Marechal, M. Martinez, J.E. Mastronunzio, B.C. Mullin, J. Niemann, P. Pujic, T. Rawnsley, Z. Rouy, C. Schenowitz, A. Sellstedt, F. Tavares, J.P. Tomkins, D. Vallenet, C. Valverde, L. Wall, Y. Wang, C. Medigue, D.R. Benson, Genome characteristics of facultatively symbiotic *Frankia* sp. strains reflect host range and host plant biogeography, Genome Res. 17 (2007) 7 - 15.

[41] P. Normand, S. Orso, B. Cournoyer, P. Jeannin, C. Chapelon, J. Dawson, L. Evtushenko, A.K. Misra, Molecular phylogeny of the genus *Frankia* and related genera and emendation of the family Frankiaceae, Int. J. Syst. Bacteriol. 46 (1996) 1-9.

[42] P. Portier, D. Costechareyre, D. Muller, D. Chapulliot, C. Oger, G. Perriére, L. Guéguen, V. Daubin, X. Nesme, Démasquage des gènes spécifiques d'une espèce génomique du complexe Agrobacterium tumefaciens par AFLP et multicapteur à ADN, Les Actes du BRG 7 (2008) 351-371.

[43] P. Portier, M. Fischer-Le Saux, C. Mougel, C. Lerondelle, D. Chapulliot, J. Thioulouse, X. Nesme, Identificationof genomic species in *Agrobacterium* Biovar 1 by AFLP Genomic Markers, Appl. Environ. Microbiol. 72 (2006) 7123-7131.

[44] J.L.W. Rademaker, B. Hoste, F.J. Louws, K. Kersters, J. Swings, L. Vauterin, P. Vauterin, F.J. De Bruijn, Comparison of AFLP and rep-PCR genomic fingerprinting with DNA-DNA homology studies: *Xanthomonas* as a model system, Int. J. Syst. Evol. Microbiol 50 (2000) 665-677.

[45] S. Rombauts, Y. Van de Peer, P. Rouzé, AFLPinSilico, simulating AFLP fingerprints, Bioinformatics 19 (2003) 776-777.

[46] Y. Shi, J. Ruan, DNA base composition and homology values in the classification of some *Frankia* strains, Acta Microbiol. Sin. 32 (1992) 133-136.

[47] P. Simonet, M. Bosco, C. Chapelon, A. Moiroud, P. Normand, Molecular characterization of *Frankia* microsymbionts from spore-positive and spore-negative nodules in a natural alder stand, Appl. Environ. Microbiol 60 (1994) 1335-1341.

[48] P. Simonet, E. Navarro, C. Rouvier, P. Reddell, J. Zimpfer, Y. Dommergues, R. Bardin, P. Combarro, J. Hamelin, A.M. Domenach, F. Gourbiere, Y. Prin, J.O. Dawson, P. Normand, Co-evolution between *Frankia* populations and host plants in the family Casuarinaceae and consequent patterns of global dispersal, Environ. Microbiol. 1 (1999) 525-533.

[49] E. Stackebrandt, W. Frederiksen, G.M. Garrity, P.A. Grimont, P. Kampfer, M.C. Maiden, X. Nesme, R. Rossello-Mora, J. Swings, H.G. Truper, L. Vauterin, A.C. Ward, W.B. Whitman, Report of the ad hoc committee for the re-evaluation of the species definition in bacteriology, Int. J. Syst. Evol. Microbiol. 52 (2002) 1043-1047.

[50] M. Valdés, *Frankia* Ecology, In: K. P. a. W. E. Newton (Ed.), Nitrogen-fixing Actinorhizal Symbiosis, Springer, Dordrecht, The Netherlands, 2007, pp. 49-71

[51] P. Vinuesa, C. Silva, M.J. Lorite, M.L. Izaguirre-Mayoral, E.J. Bedmar, E. Martinez-Romero, Molecular systematics of rhizobia based on maximum likelihood and Bayesian phylogenies inferred from rrs, atpD, recA and nifH sequences, and their use in the classification of Sesbania microsymbionts from Venezuelan wetlands, Syst. Appl. Microbiol. 28 (2005) 702-716.

[52] P. Vos, R. Hogers, M. Bleeker, M. Reijans, T. Van de Lee, M. Hornes, A. Frijters, J. Pot, J. Peleman, M. Kuiper, M. Zabeau, a new technique for DNA fingerprinting Nucleic Acids Res. 23 (1995) 4407-4414.

[53] A. Welsh, B.S. Mirza, J.P. Rieder, M.W. Paschke, D. Hahn, Diversity of frankiae in root nodules of *Morella pensylvanica* grown in soils from five continents, Syst. Appl. Microbiol. 32 (2009) 201-210.

[54] J.P. Zehr, B.D. Jenkins, S.M. Short, G.F. Steward, Nitrogenase gene diversity and microbial community structure: a cross-system comparison, Environ. Microbiol. 5 (2003) 539-554.

Fig. 1 UPGMA dendrogram based on the Evolutionary Genomic Divergence (EGD) calculated from the AFLP data analysis of *Frankia* strains with the four concatenated conditions. The numbers at the branches represent bootstraps values (1000 repetitions).

Chapter II: *Frankia* phylogeny depiction and species definition by Amplified Fragment Length Polymorphism (AFLP) and Multilocus Sequence Analysis (MLSA)

II.3. (Article 2) Multilocus-based phylogenetic reconstruction of Frankia genus

Authors

Bautista GHH[1,2], Herrera-Belaroussi A, Xu R[1], Normand P[1]. Valdés M[2] and Fernandez MP[1]

Laboratory addresses

[1] Université de Lyon, F-69622, Lyon, France and Université Lyon1, Villeurbanne, CNRS, UMR5557, Ecologie Microbienne, 43, Boulevard du 11 novembre 1918, F-69622 Villeurbanne Cedex, France

2Laboratorio de Microbiologia Agricola. Escuela Nacional de Ciencias Biologicas, IPN, MéxicoD.F.

Corresponding author :

Hector H. Bautista Guerrero

Ecologie Microbienne UMR5557, 43, Boulevard du 11 novembre 1918, F-69622 Villeurbanne Cedex, France

Tel: (33) 472 44 82 00

Fax: (33) 426 23 44 68

E-mail: hector.bautista@univ-lyon1.fr

Journal: Applied and Environmental Microbiology

Chapter II: *Frankia* phylogeny depiction and species definition by Amplified Fragment Length Polymorphism (AFLP) and Multilocus Sequence Analysis (MLSA)

Introduction

The actinobacteria *Frankia* establish a symbiotic nitrogen fixing interaction, actinorhizal symbiosis, with the roots of 25 genera of angiosperms collectively called Actinorhizal Plants (Benson & Silvester, 1993). The actinorhizal symbiosis has a world-wide distribution and accounts for an important percentage ,as great as 25% (Dixon and Wheeler 1986), of the total terrestrial nitrogen fixed, playing an essential role of the organic nitrogen cycling to the ecosystems, together with the Rhizobia-Legumes symbiosis..

The evolution of both actinorhizal partners and their diversification within an ecological frame had led to creation of biogeographical symbiotic patterns by processes still not well understood (Benson & Dawson, 2007a). The assessment of diversity within *Frankia* genus and its phylogeny depiction is fundamental for the understanding of ecological adaptation and species diversification processes.

Frankia remains a focus of systematic research, because its taxonomy and phylogeny are tasks not completely accomplished, due to the slow growing rate, filamentous morphology and non cultivability of several strains (Dobritsa, 1998a). Since the first in vitro isolation (Callaham et al., 1978), *Frankia* genus structure have been studied by means of polyphasic strategies. In this regard, molecular methods overcome in vitro culturing problems, and different molecular markers were analyzed for phylogenetic purposes (Hahn et al., 1999; Nalin et al., 1995; Nazaret et al., 1991; Normand et al., 1996; Simonet et al., 1991). Actually, as for many microbial groups, 16S rRNA sequence analysis has been adopted as the standard practice to establish a hierarchical organization of *Frankia* taxon (Clawson *et al.*, 2004b; Huguet *et al.*, 2001b; Normand *et al.*, 1996).

The comparative study of 16S rRNA sequence suggests that *Frankia* is the unique member of the Frankiaceae family and excludes the genera *Geodermatophilus* sp. and *Blastococcus* sp. . Conversely, *Acidothermus cellulolyticus*, a bacteria isolated from 55-60°C acidic water and mud samples collected in Yellowstone National Park, was found as the closest neighbour of this family (Normand et al 1996). The genus is a monophyletic group splitted into four well defined main clusters coherent with host infection groups: Cluster 1) Groups the *Myrica-Casuarina-Alnus* infective strains; Cluster 2) encompasses the non isolated symbionts of *Dryas*, *Coriaria*, *Datisca* and *Ceanothus* sp. (Rhamnaceae). The Cluster

3 includes the Elaeagnaceae and Rhamnaceae typical strains while the Cluster 4 gathers a divergent group of ineffective (non-nitrogen fixing) or non-infective strains (Normand et al., 1996).

Accurate identification of frankiae at the genus and major-cluster level is obtained not only by 16S rRNA sequencing but also with other markers as the protein coding genes i.e. nifH (Welsh *et al.*, 2009b), recA (Marechal et al., 2000) glnA (Clawson *et al.*, 2004b) and the ITS 16-23S rRNA (Gtari et al 2010). To date a vast sequence data set is available in electronic databases comprising culturable and endophyticstrains. Meanwhile is widely accepted that single marker-based studies may resolve gene phylogenies but can roughly reflect the evolutionary history of species because of the stochasticity of gene sorting during speciation in bacteria (Vinuesa et al 2005). Indeed single markers might fail in the accurate assignation of strains to species, likely due to high sequence conservation of single constitutive markers (Vinuesa et al 2005, Gervers et al 2005).

The lack of species definition is by far the major drawback of Frankia current classification and hinders the comprehension of ecological distribution and richness. About 20 genomospecies have been reliably delineated by DNA-DNA hybridization (DDH) in addition to some others not yet confirmed (Akimov & Dobritsa, 1992b; Bloom *et al.*, 1989a; Fernandez *et al.*, 1989b; Shi & Ruan, 1992). Up to now DNA reassociation rate is still considered the standard for bacterial species definition nevertheless it displays several short-comings (Coenye et al., 2005).Frankia species definition by DDH, represent a cumbersome task and is not applicable to non isolated specimens.

Alternative techniques looking for DDH replacement in Frankia genomospecies definition, includes the comparative analysis of hypervariable regions, as the intergenic sequences (IGS) between 16S and 23S rRNA and nifD-K genes, whether single (Nalin et al., 1995) or combined (Lumini and Bosco 1999) these two regions are very discriminant but roughly permit to define Frankia genomospecies, as other single locus they are prone to recombinatorial events and due to their small size signal saturation is easily achieved. Genomic fingerprinting methods as REP-PCR and RAPDs go direct to subspecies and strain level (ref), being well suited for diversity studies and strain typing, however they can hardly be used to robustly evaluate pairwise genomic distances because they provide only limited number of fragments with low reproducibility {Meudt, 2007 #3801}. Among fingerprinting

genomic approaches the AFLP technique highly improves the genomic relatedness estimation, thus Bautista et al. (2010) designed an AFLP protocol for species definition by comparing the entire genomic polymorphism (current genome mispairing) between strains and then estimating the evolutionary genomic distance (Mougel et al 2002) with respect to a common ancestor. Moreover, their protocol permits to overcome the noncultibability of strains as shown by the almost inexistent influence of the host (Betulaceae and Casuarinaceae) DNA in *Frankia* correct typing. The AFLP data bases are still poor and lab to lab data portability has not been yet estimated, in addition, more investigations need to be made in order to determine the low influence of other host plant families DNA in *Frankia* typing.

Recently, another technique, the multilocus sequence analysis (MLSA), is becoming the standard in microbial molecular systematic and the most promising method for replacing the DDH. This approach consists in the simultaneous comparative analysis of partial sequences of several protein-coding genes; the diversity and relationship of different isolates across related taxa are then assessed by using an appropriate phylogenetic or cladistic approach (Gervers et al 2005). The ideal approach would rely on the use of genes evolving more rapidly than 16S rRNA and that are ubiquitous within a taxon (Gerves and Coenye 2007). The analysis of multiple genes provides a buffer against the distorting effects of recombination at a single locus (Zeigler et al 2003).Actually the MLSA has been used to resolve the species and candidate species in multiple bacterial genera (Hanage 2006), and the results suggest that in several cases species correspond to well-resolved sequence clusters (Fraser et al 2010) providing a criterion for delimiting the extent of microbial species. Moreover, MLSA can give an accurate prediction of whole-genome relatedness with a level of resolution comparable to DDH (Konstantinidis et al 2006; Zeigler 2003). Besides the species delimitation, MLSA seeks the accurate assignment of genera (Bishop *et al.*, 2009a; Martens *et al.*, 2008) and their placement into broader taxonomic groups (Sawabe et al., 2007) aiming to provide a global and reliable overview of interorganismal relationships and species diversification.

The MLSA scheme has not so far been explored in *Frankia*, a pioneer study carried out the analysis of the genus by the concatenation of the 16S rRNA with the *glnA* gene (Benson and Clawson, 2004). We approached *Frankia* genus phylogenetic study and species definition with a five locus-based MLSA scheme. We aim to clarify the genus structure,

monophyly and its relationships with other taxons. At lower level, the major subdivisions, their appearance during the evolution and species delineation were studied. The portability of data will permit the generation of a data base prone to be easily expanded by future studies.

II.3.1. Material and Methods

Bacterial strains and culture condition. Cultured *Frankia* strains, unisolated endophytes as well as their host plant and geographic origin are listed in Table 1. The strains were grown at 28°C in liquid F medium (*Elaeagnaceae*-infective strains) (Simonet *et al.*, 1985), F medium containing Tween 80 (*Alnus*-infective strains) or BAP medium (other strains) (Murry *et al.*, 1984c) with weekly manual shaken.

Selection of gene loci and design of primers. The five loci studied are spread throughout AcN14a genome (Fig.1) and represent different COG functional categories: energy production and ATP biosynthetic genes (*atpD*), chromosomal replication genes

Table 1. Liste of *Frankia* strains and actinorhizal nodules included in this study.

Trivial designation	Source host	Geographical origin	Reference
Alnus-Myrica-Casuarina infectivity group			
$Ac2_{18}$	*Alnus cordata*	Miribel, France	Unpublished
$Ac23_{23}$	*Alnus cordata*	Saou (France)	Unpublished
$Ac23_{40}$	*Alnus crispa*	Orléans, France	(18)
$Ac24I_5$	*Alnus cordata*	Orléans, France	Unpublished
ACN14a	*Alnus crispa*	Tadoussaq, Canada	(44)
AcoN24d	*Alnus cordata*	Orléans, France	(52)
AcVc1	*Alnus cordata*	Corsica, France	(18)
Ag21D1	*Alnus glutinosa*	Corsica, France	Unpublished
$Ag24_{251}$	*Alnus glutinosa*	Orléans, France	(18)
$Ai96_6$	*Alnus incana*	101 (Lacrans (01) Carrière Famy	Unpublished
AirI1	*Alnus incana ssp. rugosa*	Vermont, USA	(35)
AJ01	*Alnus japonica*	Makabe, Ibaraki Pref., Japan	Unpublished
Ar24H5	*Alnus rubra*	Orléans, France	(55)
ARgN22d	*Alnus rugosa*	Québec, Canada	(43)
$ARgP5^{AG}$	*Alnus rugosa*	Québec, Canada	(44)
ArI3	*Alnus rubra*	Oregon, USA	(43)
$Av59_7$	*Alnus viridis*	Alpe du Grand Serre, France	Unpublished
$AvBi200_{NOD}$*	*Alnus viridis*	Bionassay, France	
$AvBi201_{NOD}$	*Alnus viridis*	Bionassay, France	
AvcI1	*Alnus viridis ssp. crispa*	Ontario, Canada	(4)
AVL3	*Alnus viridis*	Lautaret, France	(18)
AVN17o	*Alnus viridis*	La Toussuire, France	(18)
CpI1	*Comptonia peregrina*		
I38	*Alnus incana*	La pallud, France	Unpublished
M16467	*Myrica pensilvanica*	New Jersey, USA	(11)
M16477	*Myrica pensilvanica*	New Jersey, USA	Unpublished
$Mg60_2^{AG}$	*Alnus glutinosa*	Landes, France	(18)
MpI_1	*Myrica pensilvanica*		
BR	*Casuarina equisetifolia*	Brazil	(50)
CcI_2	*Casuarina cunninghamiana*	Florida, USA	(62)
CcI_3	*Casuarina cunninghamiana*	Florida, USA	(62)
CeD	*Casuarina equisetifolia*		
$Cg70_3$	*Casuarina glauca*	Inde	Unpublished
$Cg70_4$	*Casuarina glauca*	Inde	Unpublished
CjI-82	*Casuarina junghuniana*	Thailand	(16)
TA	*Allocasuarina torulosa*	Australie	(62)

Table 1. Continuation

Elaeagnaceae-Rhamnaceae infectivity group			
BMG5.3	*Elaeagnus angustifolia*	Tunisia	(24)
Cg701	*Casuarina glauca*	Inde	
CH37	*Hippophaë rhmanoides*	Nogent sur Marne, France	(49)
ChI_7	*Colletia hystrix*	Chile	(10)
$Colletia_{NOD}$	*Colletia spinossisima*	Argentine	
Ea112	*Elaeagnus angustifolia*	Ecully, France	(18)
$Ea35_2$	*Elaeagnus angustifolia*	Sutri, Italy	(29)
$Ea36_7$	*Elaeagnus angustifolia*	Ecully, France	(18)
$Ea48_1$	*Elaeagnus angustifolia*	St Etienne de Tinée, France	Unpublished
$Ea48_4$	*Elaeagnus angustifolia*	St Etienne de Tinée, France	(29)
$Ea7_1$	*Elaeagnus angustifolia*	Toulon, France	(29)
Ea84	*Elaeagnus angustifolia*	Pont en Royan, France	(29)
EAN1pec	*Elaeagnus angustifolia*	Illinois (U.S.A)	(34)
EUN1f	*Elaeagnus umbellata*	Illinois, USA	(34)
G_2	*Casuarina equisetifolia*	Guadeloupe	(17)
$Hr75_2$	*Hippophaë rhmanoides*	Station 7 N75, France	Unpublished
ORS060501	*Colletia spinossisima*	Argentina	(19)
R43	*Casuarina equisetifolia*		
Nod-/Fix- strains			
CaI_1	*Ceanothus americanus*	Adisson, USA	(35)
CeSI5	*Casuarina equisetifolia*	USA	(9)
Cn3	*Coriaria nepalensis*		(41)
EuI1a	*Elaeagnus umbellata*	Petersham, USA	(2)
EuI1c	*Elaeagnus umbellata*	USA	(3)
MgI_5	*Myrica gale*	New York, USA	(35)
Pti1	*Purshia tridentata*	Wyoming, USA	(2)
Dryas-Coriariaceae-Datiscaceae infectivity group			
Coriaria M_{NOD}	*Coriaria myrtifolia*	France	
Coriaria R_{NOD}	*Coriaria ruscifolia*	Mexique	
$Datisca_{NOD}$	*Datisca glomerata*	Argentine	
$Dryas_{NOD}$	*Dryas drumondii*	Canada	

*NOD sufixe indicates the nodular samples.

(*dnaA*), cell division genes (*ftsZ*), transcription and RNA metabolism genes (*rpoB*) and carbohydrate metabolism genes (*pgk*). In addition such loci are present as unique copy in the three reference *Frankia* genomic sequences. The gene *atpD* encodes the beta subunit of the membrane ATP synthase, *dnaA* encodes for the initiator of the chromosomal replication, *rpoB* encodes the beta subunit of RNA polymerase, *ftsZ* encodes for an essential protein implicated in the cell division cycle and the *pgk* product (phosphoglycerate kinase) participates in the carbohydrate metabolism. The design of amplification primers for the genes *atpD*, *dnaA*, *ftsZ* and *pgk* was based in the genomic sequences of the three *Frankia* strains hitherto reported (Normand *et al.*, 2007a) while primers used for *rpoB* amplification were reported elsewhere (Kim *et al.*, 1999). Some of the primers characteristics are summarized in Table 2.

DNA extraction. Total DNA from pure cell cultures was extracted with the DNeasy Plant Mini kit (QIAGEN) according to the manufacturer's instructions. Composite nodular DNA was extracted from one single lobe according to Hughet et al. (Hughet *et al.*, 2001). After washing, the outer layers were removed, then the lobe was crushed in 300 µL of TCP extraction buffer (100 mM Tris-HCl [pH 7], 0.5 M NaCl, 50 mM EDTA [pH 8], 2% [wt/vol] cethylmethylammonium bromide [Sigma, St. Louis, Mo.], 1% [wt/vol] polyvinylpolypyrrolidone [Sigma]). The homogenate was incubated at 65°C for 1h and centrifuged twice at 6000 *X g* for 20 min. The supernatant was extracted with an equal volume of Chloroform-isoamyl alcohol (24:1, vol/vol) and centrifuged at 13000 *X g* for 20 min. DNA from aqueous phase was precipitated in ethanol for at least 2h at -20°C. The sample was then centrifuged at 13,000 *X g* for 30 min, and the resulting DNA pellet was washed with 70% (vol/vol) ethanol, air dried, and dissolved in 10 µL of tris-EDTA (TE) buffer (pH 7.5).

PCR amplification . The five loci were amplified by a simple PCR reaction in a final volume of 50 µL. Each PCR reaction contained 5 µL of template ADN (50-200ng), 5 µL of 10 X PCR buffer, 2.5 µL of each primer (10 mM), 5 µL of a dNTP mix (2 mM), 4µL of $MgCl_2$ (25 mM), 5 µL DMSO, 2.5 U of *Taq* DNA polymerase and 23 µL of sterile MiliQ water. The reaction conditions were initial denaturation at 94 °C for 5 min, followed by 30 cycles of the denaturation at 94 °C for 1 min, annealing for 1 min at the primer-pair-specific annealing temperature (Table 2), and extension at 72 °C for 1 min. A final extension step was

Table 2. Primers used for amplification and sequencing

Gene	Primer name and Sequence (5'→ 3')	Annealing Temperature
atpD	**atpDfdT7F** *TAATACGACTCACTATAGGGACCGGSATCAAGGTCATCGAC **atpDrvT3R** ATTAACCCTCACTAAAGGGACCGAGGATGGCGATGATGTC	62 °C
dnaA	**dnAfdT7F** TAATACGACTCACTATAGGGGAGGARTTCACCAACGACTTCAT **dnArvT3R** ATTAACCCTCACTAAAGGGACRGAAGTGCTGGCCGATCTT	62 °C
ftsZ	**ftsZfdT7F** TAATACGACTCACTATAGGGCCGTCAACCGGATGATCGAA **ftsZrvT3R** ATTAACCCTCACTAAAGGGAGCSGCGTTGATCTCGAACAG	62 °C
pgk	**pgkFwdT3** ATTAACCCTCACTAAAGGGATGAGGACGATCGACGACCTGC **pgkRevT7** TAATACGACTCACTATAGGGCGCSAGGAAGGTGAAGCACAT	62 °C
rpoB	**rpoB-KIM1** CGACCACTTCGGCAACCG **rpoB-TBB2** TACGGCGTCTCGATGAASCC	58 °C

*underlined sequences correspond to the standard promoter sequences of RNA polymerase T3 and T7.

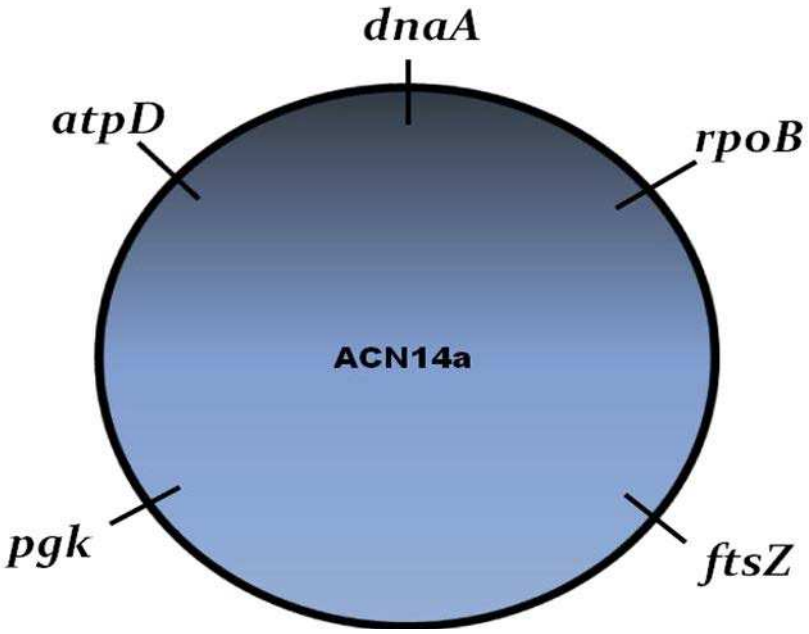

Fig 1. Schematic representation of the physical location of the five loci across the ACN14a *Frankia* genome.

performed at 72 °C for 10 min. Amplified products were resolved on a 1.0% agarosa gel. The amplification products were purified on QIAquick spin columns (QIAGEN).

16SrRNA amplification was performed with the universal primers pA and pH (Edwards *et al.*, 1989). Initial PCR amplification was achieved in a 100 µL mixture composed by 2 µL of template ADN (10 ng), 10 µL of 10X PCR buffer, 10 µL of each primer (10 mM), 10 µL of a dNTP mix (2 mM), 4 µL of $MgCl_2$ (25 mM), 5 µL DMSO, 5 U of *Taq* DNA polymerase and 53 µL of sterile MiliQ water. Amplification conditions were initial denaturation at 94 °C for 5 min, followed by 30 cycles of the denaturation at 94 °C 45 s, annealing for 55 min at 55 °C, and extension at 72 °C for 1 min. A final extension step was performed at 72 °C for 5 min.

Sequencing. Both strand sequencing was performed by GenoScreen (Lille, France) employing the DyeDeoxy Terminator Cycle Sequencing Kit and an ABI Prism 3730 automated DNA sequencer (Applied Biosystems).

Sequence analysis: Sequences was edited and alignments were performed with the BioEdit and ClustalW, respectively (Hall, 1999; Thompson *et al.*, 1994). *Acidothermus cellulolyticus*, *Blastococcus saxobsidens*, *Geodermaphilus obscurus*, *Modestobacter multiseptatus*, *Salinispora arenicola* and *Stackebrandtia nassauensis* as external groups selection was based on a recent study (Wu *et al.*, 2009).

Statistics for each locus, such as the number and proportion of polymorphic sites, mean G+C content and ratio of mean synonymous substitutions per synonymous site/mean nonsynonymous substitutions per nonsynonymous site (d*S*/d*N*) were calculated using the START2 program (Jolley *et al.*, 2001). Aligment concatenation was made using DnaSP5 (Librado & Rozas, 2009). Maximum Likelihood analyses were carried out in the software TREEFINDER (Jobb *et al.*, 2004). Following the Akaike Information Criterion (AIC) (Posada & Buckley, 2004) the GTR + Γ model was used as the best-fitting model of nucleotide substitution for the concatenate data. The node reliability was assessed using approximate bootstrap test "LR-ELW edge support" with 1000 replications (Jobidon & Thibault, 1980). The Splits decomposition analysis and Pairwise Homoplasic Index for recombination detection (PHI test) (Bruen *et al.*, 2006) were performed Splits Tree 4 software (Huson & Bryant, 2006).

II.3.2. Results

PCR amplification and particular features of each locus

Partial sequences of five housekeeping genes were analyzed for 65 *Frankia* strains. They represented 57.35%, 38.87%, 48.87% and 43.66% of *pgk, dnaA, atpD, ftsZ* entire gene respectively. For *rpoB* 11.67% of the entire gene (3426 pb in AcN14a) was analyzed, corresponding to a very discriminative region of this gene in a *Mycobacterium* sp. study (Devulder *et al.*, 2005). The main characteristics of the different loci here analyzed are summarized in Table 3.

Mean G+C content ranged from 66.4 % (*dnaA*) to 73.6 % (*pgk*). These values are in the range of total G+C% content reported for the genus (66-72%). Mean genetic divergences ranged from 0.035 (*rpoB*) to 0.144 (*pgk*) and in all cases were higher than 16S rRNA (0.22) value (Table 3).

The purpose of the Tajima's D test is to distinguish between a DNA sequence evolving randomly (neutrally) versus one evolving under a non-random process. The values obtained for 4 loci are relatively close to 0 indicating they are submitted to a neutral selection and with similar evolutionary rates. The *rpoB* negative D value could be due either to a purifying selection event or the presence of many singletons in the sequence analyzed (Yang, 2006). The d*N*/d*S* ratio is used to estimate the degree of selection operating on each locus. For all markers d*N*/d*S* values are <1, suggesting they are under neutral selection.

Single protein coding gene analysis

Individual phylogenetic trees for each protein–coding locus were constructed (Fig. 2). The level of variation observed (Table 3) within the sequences led to use, nucleotide sequences instead of amino-acid sequences for phylogenetic analysis. Considerable variations in tree topology were observed for all the separate genes. The *atpD*, *dnaA*, *ftsZ* and *pgk* phylogenies sustained the delineation of the genus *Frankia*, while *rpoB* sequence analysis harbored an aberrant topology in which members of external taxa (outgroups) were interspersed with *Frankia* members. Besides, the consolidation of the four 16S-major clades was only supported by the *ftsZ*, *pgk* and *dnaA*; nonetheless their branching pattern was

variable. At lower taxonomic level, *atpD*, *dnaA*, *ftsZ* and *pgk* fairly discriminated all genomospecies in coherent groups statistically well supported.

Table 3. Characteristics of loci of the 65 independent *Frankia* strains.

Locus	Fragment size (bp)	Mean G+C content (%)	Allele number	Number of polymorphic sites (%)	Overall mean divergence	Tajima's D test	dN/dS	Phi test*
atpD	**714**	**68.8**	**45**	**196 (27.5)**	**0.073**	**0.0443**	**0.096**	**0.029**
dnaA	**623**	**66.5**	**41**	**196 (31.4)**	**0.082**	**1.237**	**0.057**	**0.011**
ftsZ	**718**	**69.7**	**42**	**287 (39.9)**	**0.063**	**0.911**	**0.049**	**0.66**
pgk	**695**	**73.6**	**45**	**333 (47.9)**	**0.144**	**1.385**	**0.2057**	**0.42**
rpoB	**399**	**69.6**	**39**	**122 (30.6)**	**0.035**	**-0.790**	**0.070**	**0.023**

*P value for phi test of recombination <0.05

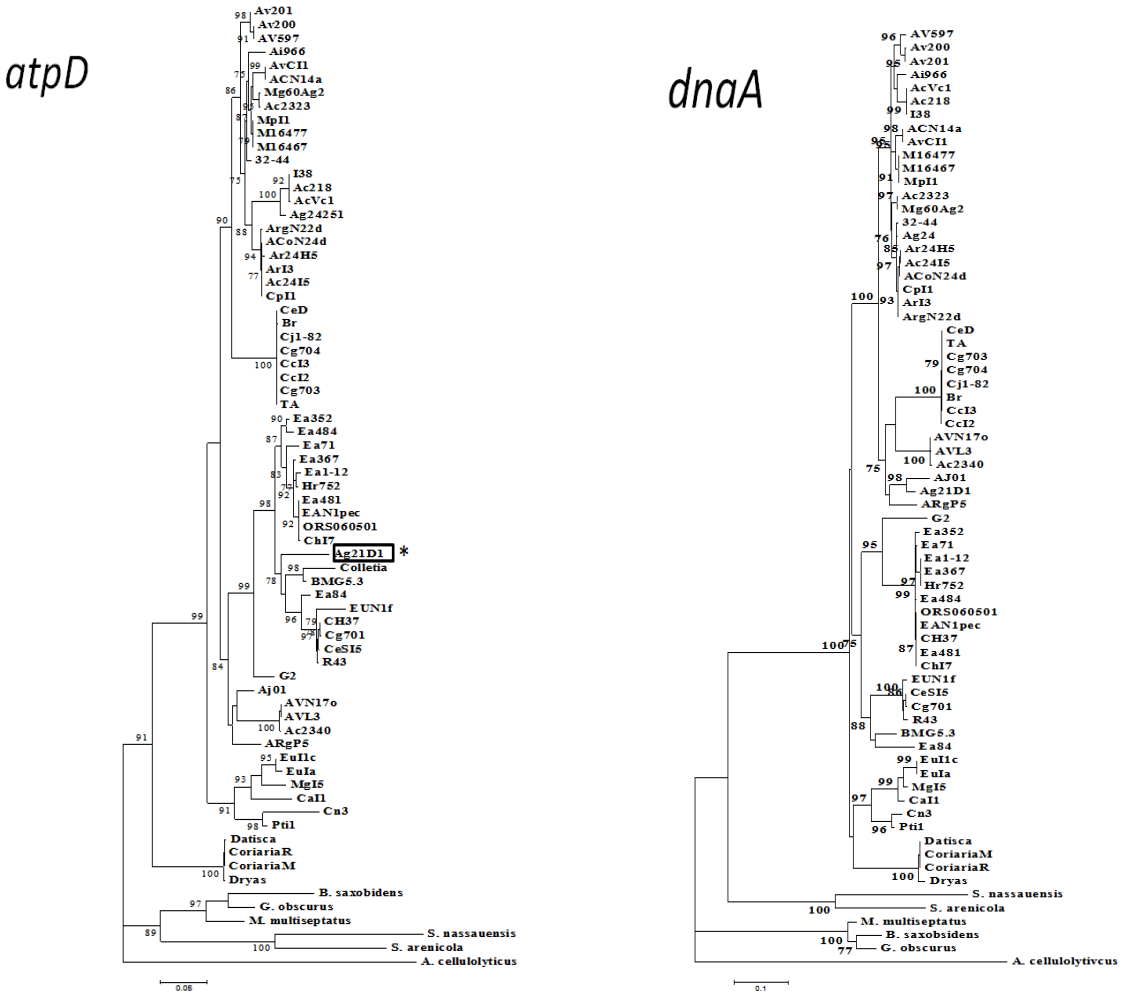

Figure 2. Phylogenetic reconstruction based on individual analyses of the *atpD*, *dnaA*, *ftsZ*, *pgk* and *rpoB* genes. Analyses were conducted using the ML method and the GTR+G model of substitution. LR-ELW values ≥ 75 (using 100 replicates) are indicated at branching points. Bars, indicated stimated substitutions. * the evidence of lateral transfer of the Ag21D1.

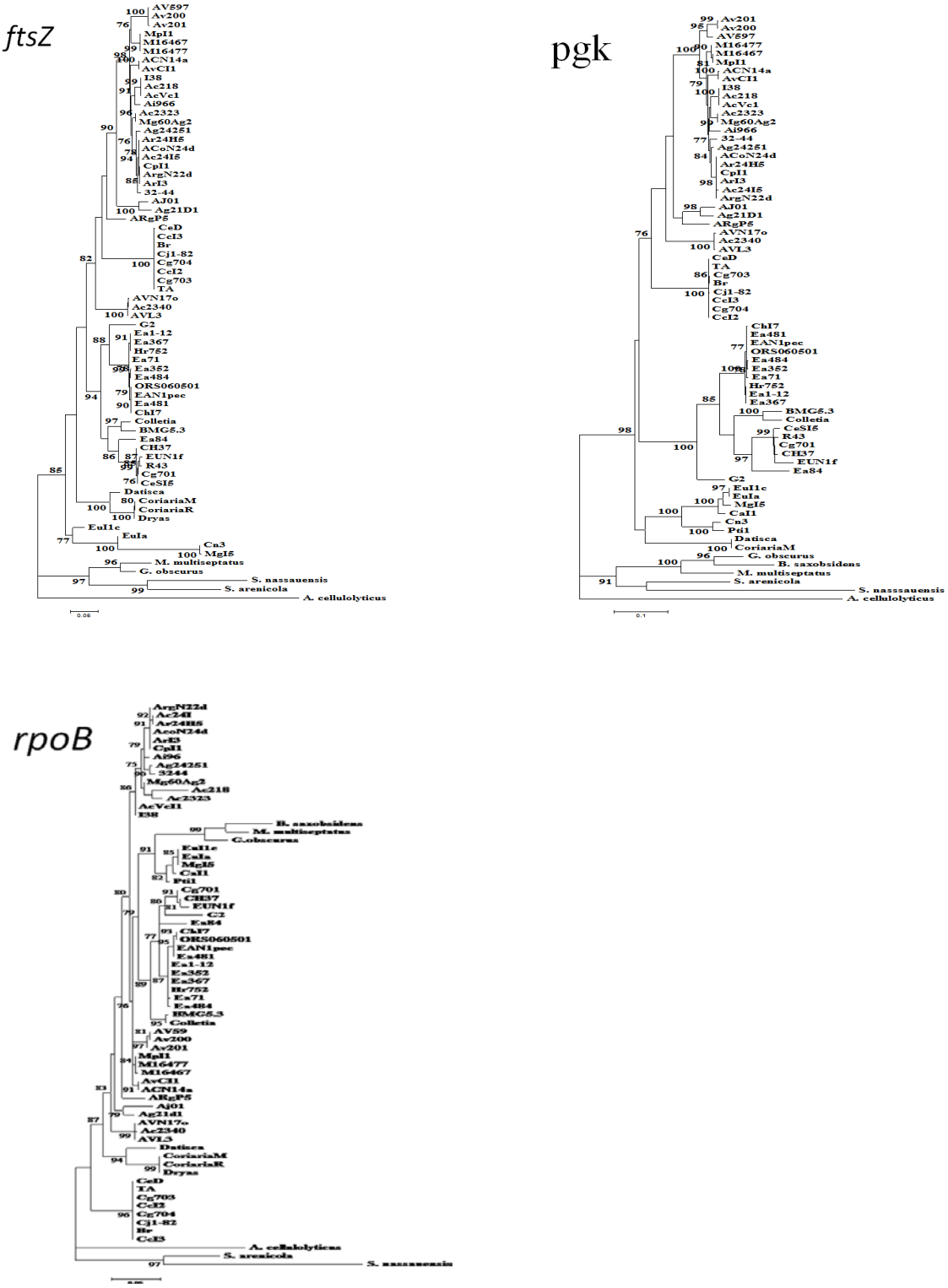

Figure 2. Continuation

MLSA-based Frankia phylogeny and species delineation

The concatenated alignment of five loci resulted in a supergene of 3165 nt with a mean G+C content of 67.8 mol% and a mean pairwise distance of 0.104.

The phylogenetic analysis of concatenated sequences harbored a robust topology with high statistical support of the nodes at the different branching level. The dendrogram display a similar topology to that obtained with 16S rRNA sequence dividing the genus into four major clusters (Fig. 3): The Cluster I (P value= 92) contains *Alnus-Myrica* and *Casuarina* infective strains (including the proposed *F. alni* G1 type species) , the Cluster II (P-value= 100) gathered *Elaeagaceae and Rhamnaceae* infective strains (P-value=96), the Cluster III (P-value= 100) comprises the nod-/fix- strains and finally the Cluster IV (P-value=100) regrouped *Dryas-Coriariaceae-Datiscaceae* non isolated endophytes. Moreover the high support of the external nodes allows to suggest that clusters I and II share a common ancestor, having the cluster III as the closest neighbor. Contrary to 16S based tree topology, the cluster IV was undoubtedly positioned at the basal position of the dendrogram.

The genomospecies already defined by DDH and AFLP analysis were here identified as coherent cluster of strains exhibiting overall mean distances ranging from 0,027 to 0,03 (Fig. 3). The EUN1f was the only member of the genomospecies GS6 reported by Fernandez et al. (1989), here it clusters with atypical *Casuarina* strains group (0,015 of mean distance) son belonging to the same genomospecies, to avoid a conflict we will name this genospecies G9 (on the basis of AFLP genomospecies definition).

In addition to described genomospecies, the analysis based on clustering and overall mean genetic distances of concatenated sequences suggests the delineation of one novel genomospecies: the uncultured cluster IV strains (0,015 of mean distance).

Moreover the analysis corroborated the inclusion of the uncultured Sp+ (sporulating *in planta*) strains into the *F. alni* genomospecies (0,027 of mean distance) as we early suggested by AFLP studies.

Besides, the recognition of GS4 and GS5 (Corresponding to genomospecies G6 in this study) as separate groups has been conflicting since they yielded < 50% of DNA reassociation rate (Fernandez *et al.*, 1989a) but conversely, evolutionary genomic divergence values calculated with AFLP data included them in the same genomic species. By means of MLSA

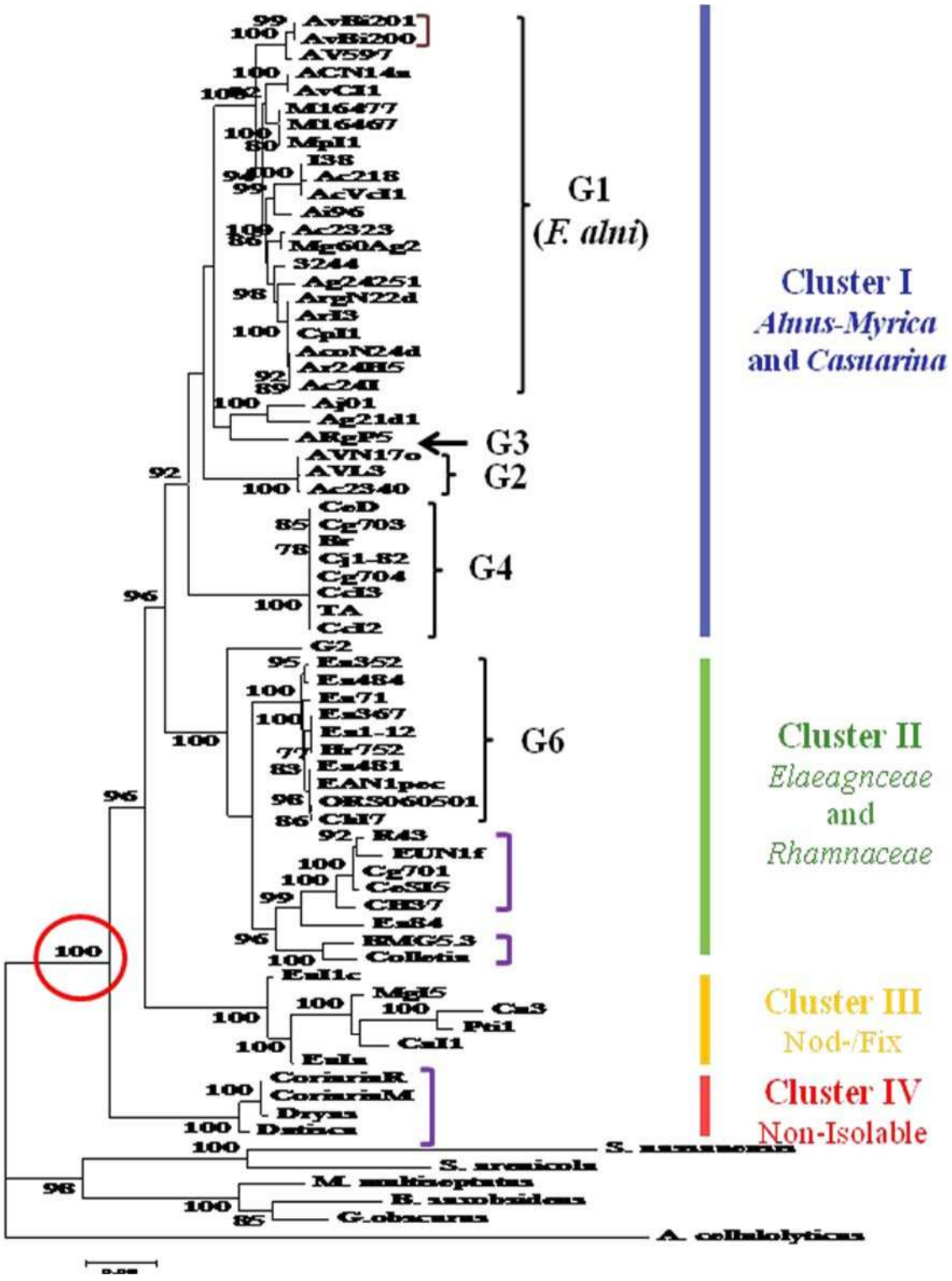

Fig. 3 Phylogenetic reconstruction based on the concatenated *atpD*, *dnaA*, *ftsZ*, *pgk* and *rpoB* gene sequences. Analyses were conducted using the ML method. LR-ELW values of ≥ (using 100 replicates) are indicated at the branching points. Bar, 0, 05% estimated substitutions.genomospecies (G) are named on the basis of the AFLP classification.

they are clustered in a coherent and likely genetically homogeneous group, moreover they display a mean genetic distance of 0,015 which place them in the same genomospecies.

At the dendrogram tips, several subgrouping with strong statistical support is shown, notably in the widely sampled *F. alni* and G6 species, these groups might represent subspecies.

16S and five-loci phylogenies comparison

The comparative analysis of the topologies yielded by the sequence analysis of the 16S rRNA gene (Fig. 4a) and the five loci-concatenate (Fig. 4b) for the same set of strains, harbored differences at the node statistical support. Moreover, with 16S *Dryas* endophyte clustered with the *Casuarinaceae* typical strains in contrast to the basal placement defined by the concatenate topology. These results confirmed the less discriminatory power of 16S compared with a multigenic approach.

With the inclusion of 16S sequence to the concatenation yielded a supergene of 4570 nt. The phylogenetic analysis of this sequence harbored a topology almost identical to that obtained with the concatenated five-loci (Fig. 4c), only minimal differences in the branching pattern within *F. alni* are evidenced. Thus the four major clades and the genospecies are clearly delineated and no relevant differences in the statistical support values are exhibited.

Recombination

Evidence for recombination in at the five loci was investigated by using the phi test implemented in the SplitsTrees software. By performing PHI recombination test (< 0,05 as statistical significant value) *atpD*, *dnaA* and *rpoB* loci exhibited significant evidence of recombination with p= 0.029, 0.011 and 0.023 respectively. Recombination evidence is observed in *atpD* phylogeny for the Ag21D1 strain. Such isolate always group with the *Alnus*-infective strains was grouped with the *Elaeagnaceae*-infective strains (Fig. 3).

Besides, split decomposition analysis displayed a star-like network with parallelogram depiction indicative of homologous recombination (Fig. 5). In this regard, genomic group comprising the EUN1f exhibit several recombination events. Conversely, *Casuarina* and Sp+ *Alnus* groups exhibit minimal recombination, thus adopting a clonal structure.

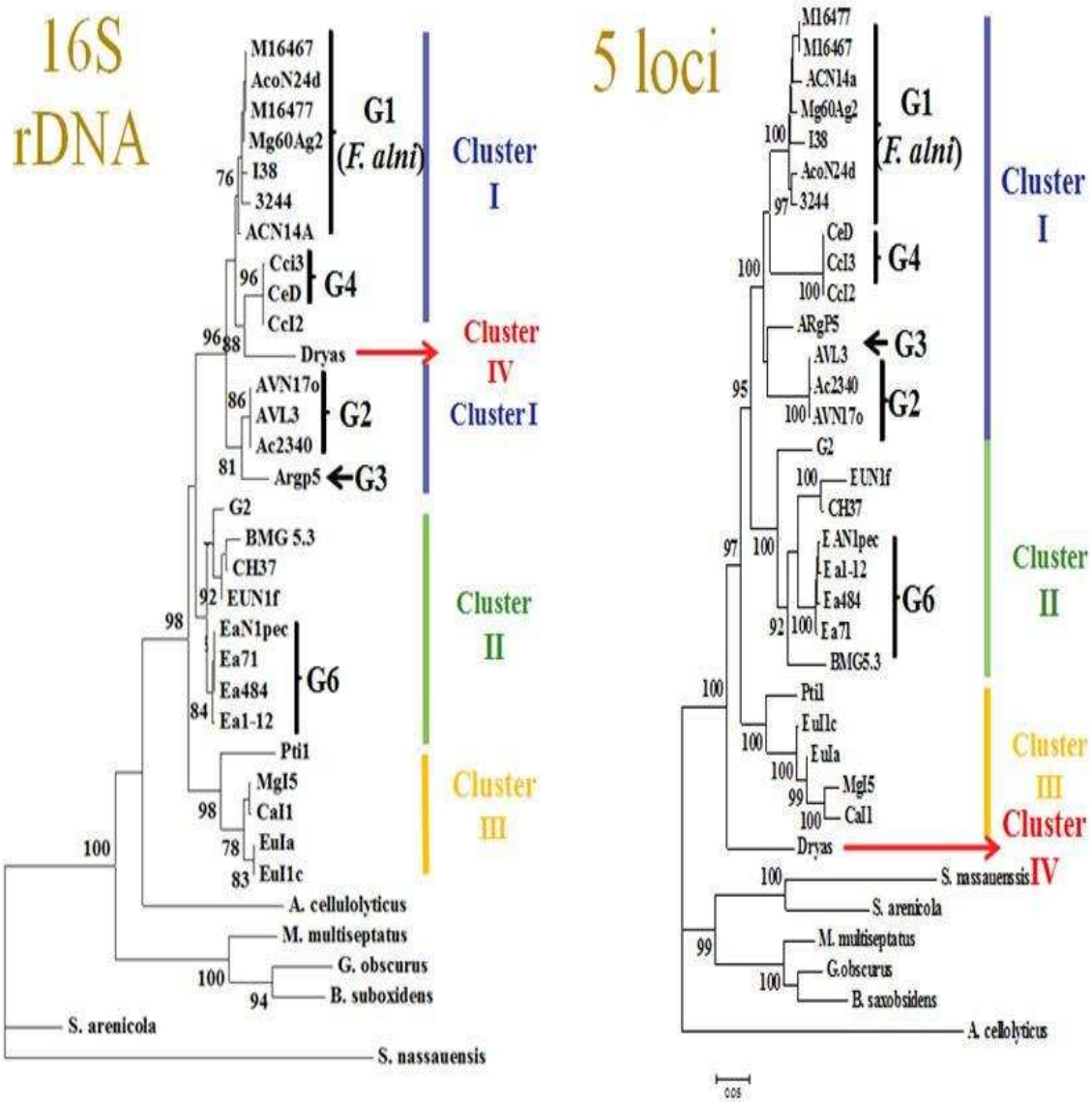

Fig. 4 Maximum-likelihood trees of *Frankia* strains calculated with the GTR+G substitution model a) Dendrogram, based on the complete sequence of 16S rRNA b) Dendrogram based on the concatenation of five loci: *atpD*, *dnaA*, *ftsZ*, *pgk* and *rpo*B) Dedrogram constructed with the concatenation of six-loci *atpD*, *dnaA*, *ftsZ*, *pgk*, *rpoB* and 16S rRNA. Numbers at the nodes correspond to the statistical support LR-ELW (100 replicates). Scale represents sustitutions per nucleotide.

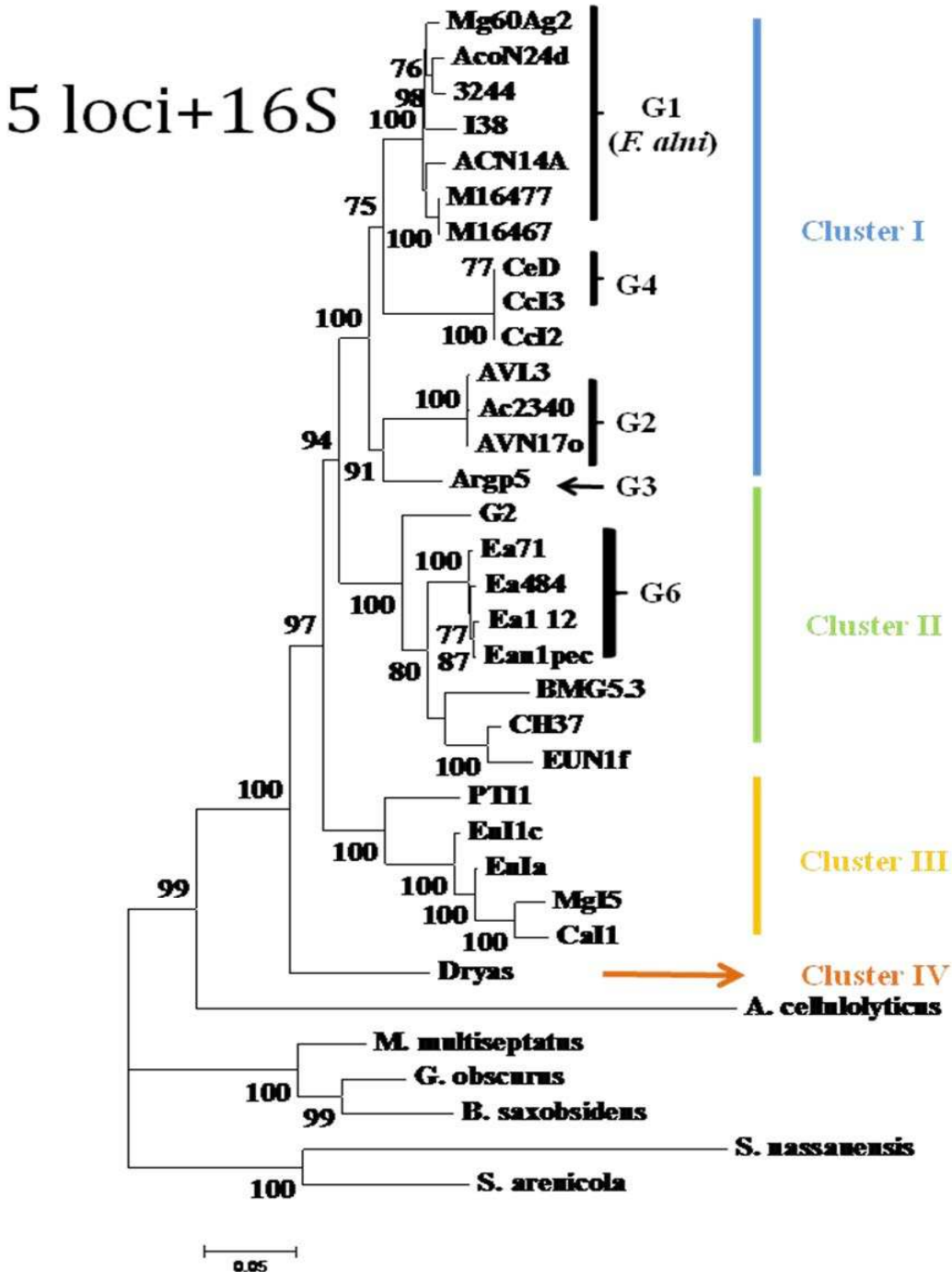

Fig. 4 Continuation

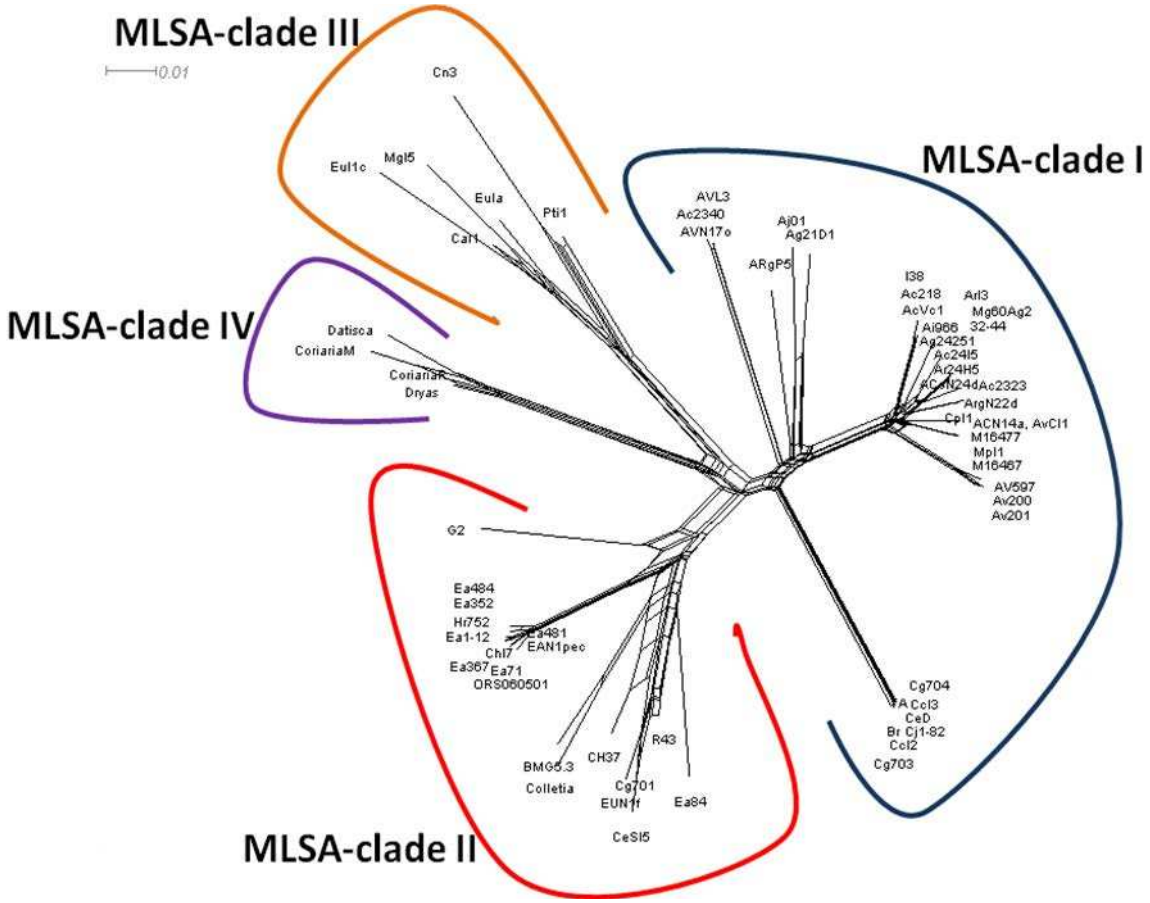

Fig 5. Split decomposition analysis show a bushy network. NeighbourNet graph of the concatenated sequences of five loci with significant (p= 0.0) evidence of recombination with PHI test constructed in SplitsTree v4.0.

II.3.3. Discussion

Frankia diversity has been analyzed by means of several approaches in order to generate a precise taxonomical classification and accurate phylogenetic depiction of the genus (Clawson *et al.*, 2004a; Lechevalier, 1984a; Normand *et al.*, 1996). The culturing difficulties of *Frankia* isolates and the non cultivability of some groups of frankiae hindered diversity analysis by classical methods and often forced researchers, to include a limited number of strains for comparative studies, consequently, the retrieved data are scarce.

MLSA is currently considered the most promising method for bacterial phylogeny depiction and species delineation (Gevers *et al.*, 2005a). Thus, a MLSA approach based on the sequence of five housekeeping genes *Frankia* phylogenetic construction was applied for the first time to a large set of *Frankia* reference strains and uncultured microsymbionts. The results obtained were compared to the standard phylogeny based on 16S rRNA.

We found in this study that single-gene phylogenies cannot reliably reflect the evolution of *Frankia*, since single-locus topologies displayed differences in the major clusters placement (data not swhon) and in with *rpoB* marker the genus consolidation was not likely supported (Fig. 3). Differences among single-locus topologies have been attributed to the occurrence of recombinatorial and lateral transfer events (Gogarten & Townsend, 2005; Ochman *et al.*, 2000). The housekeeping genes here used, did not seem to be immune to those events, since three of them were suspected to be submitted to recombination events by the Phi test. In addition, a clear lateral transfer event was detected in the *atpD* gene for the Ag21D1 strain.

One of the advantages attributed to MLSA approach is the capacity to eliminate aberrant grouping thanks to the buffering of conflicting signals achieved when sequences are concatenated (Gevers *et al.*, 2005a). Such buffering was confirmed in this work since the five-loci based phylogeny gave a robust representation of the *Frankia* genus, allowing the grouping of strains into coherent well supported clusters (Fig. 3). With this analysis the *Frankia* genus was divided in four major coherent clusters corresponding to those described with 16S sequence analysis (Clawson *et al.*, 2004a; Normand *et al.*, 1996) but in addition allowed the formation of coherent well supported internal groupings.

In this regard, the MLSA is currently addressed to determine the existence of species within bacterial groups and accurately delineate their borders (Gevers *et al.*, 2005a), thus the subgroups found likely correspond to genomospecies. For this study, we counted with DNA-DNA reassociation values (Fernandez *et al.*, 1989a) and evolutionary genomic divergence values (EGD) obtained by AFLP analysis (Bautista GHH. et al 2010, See Chapter II: Part 1) that support the MLSA subgroups likely they must acquire the species status. In this way, the major clusters I and II (the best sampled in this study) were confirmed as complex of species.

The Cluster I contains symbionts of the *Alnus*, *Casuarinaceae* and *Myricaceae*. This group comprises *F. alni* and at least three other genomic species, as before detected by AFLP (Bautista GHH. et al 2010). The orphan strains Ag21D1 and Aj01 might represent additional species each. We concluded that *Casuarinaceae* infective strains (G4) irrefutably belong to Cluster I but forming a divergent group. Even if the three groups of host plants belong to the Hamamelidae (Swensen & Mullin, 1997), the symbiosis between *Casuarinaceae* and G4 strains display some particularities i.e. high specificity and tightness of the association. Moreover a difference of about 2.1 Mpb exists between CcI3 (G4) and *F. alni* suggesting a great variability between this cluster. In spite of the high resolutive power of our analysis the parental relationships between the other genomospecies could not be defined.

The cluster II comprises *Elaeagnaceae* and *Rhamnaceae* infective strains. Four genomospecies can be identified, two of them defined by orphan strains (G_2 and Ea84). The genomospecies containing the EUN1f (GS6 according to Fernandez et al. 1989) reference strain seem to us particularly important, because it also includes three *Elaeagnaceae* infective atypical *Casuarina* strains (Cg701, CesI5 and R43) sharing as feature the lost of the ability to renodulate their original host.

Additionally we could define the G6. While the DDH values suggested the division into two different genomospecies (GS4 and GS5 in Fernandez et al, 1989), fingerprinting techniques (Lumini & Bosco, 1999; Nazaret *et al.*, 1991) and AFLP's EGD data concluded that GS4 and GS5 were not differentiable. Taking in count these molecular data and our MLSA results we suggest this group represent a single species, G6. The DDH disagreement can be attributed to bias in the DNA identity estimation by this technique, caused by the dissymmetric comparison of genomes of difference sizes.

Cluster III comprises isolates exhibiting as particular feature the lack of capacity to nodulate and to fix nitrogen (Nod-/Fix- strains). All strains -Cn3, MgI5, EuI1c, and Pti1- were isolated from hosts belonging to divergent families (Table 1), *Coriariaceae*, *Myricaceae*, *Elaeagnaceae* and *Rosaceae*, respectively. The genetic divergence of original host plants might be parallel to the corresponding strains, explaining thus the long branching within this ,although significantly well supported, group (Fig. 3).

The cluster IV comprises the closely related uncultured symbionts of *Dryas*, *Coriariaceae* and *Datiscaceae*. To date, these families of actinorhizal plants have not harbored cultivable strains. Original morphologic traits such as hyphae diameter and diazovesicles size and shape suggest a divergent evolution of this group (Becking, 1970b). The cluster IV placement within the phylogeny of the genus had been controversial (Fig. 4); however, our analysis indicates this group represents the most ancient branching in the *Frankia* evolutionary history. It is worthy noting the remarkable genetic similarity among its members; they indeed are regarded as a single species.

Coevolution has been suggested as one of the evolutionary forces guiding *Frankia* speciation. Coevolutionary signals have been detected in isolated studies about specific actinorhizal models (Huguet *et al.*, 2005b; Simonet *et al.*, 1999). A detailed inspection of the MLSA topology gives insights into some groups likely involved in such evolutionary scheme, specifically the *Hamamelidae* actinorhizal families (*Betulaceae*, *Myricaceae* and *Coriariaceae*). First we found the *Casuarinaceae* infective group, with *Gymnostoma* sp. as the only exception (Navarro *et al.*, 1997), which is composed for several typical strains isolated from different geographic sites (G4) and exhibiting high genetic identity (they cannot be differentiated by MLSA) and high specificity towards its host. This is likely the only genogroup infecting *Casuarina* and *Allocasuarina* inhabiting out of its site of origin (Simonet *et al.*, 1999). Moreover no *Casuarina* typical strains have been isolated from other actinorhizal plants.

The second case is represented by the *A. viridis ssp. viridis* strains that were splitted in two groups, G2 and AvBi Sp+ subclade of *F. alni*. The two groups are very divergent however they shared the infective-specificity toward a common host. This hypothesis must be confirmed by analyzing more strains, unfortunately no more G2 have been isolated. In this

regard the Sp+ strains can be a source of information since the Sp+ strains prevalence is high at its original stand and thus new collect at the same site can be carried out nodules.

Three *Myrica pensylvanica* infective strains clustered in a homogeneous subgroup within *F. alni* species, this pattern could suggest a coevolutionary event between the two partners. *Myrica* has been considered as a promiscuous host, because some species can be nodulated by *Frankia* from different clusters during cross inoculation trials, however Huguet et al (2005) demonstrated that in natural conditions some species of *Myrica* exhibit specificity towards their nodulating strains.

In this study we managed a MLSA approach in order to study *Frankia* genetic diversity and delineate a robust phylogeny of the genus. The use of partial sequences of five housekeeping genes reliably resolved the phylogenetics incongruences harbored by single-gene based analysis. Moreover the inclusion of a 6th marker into de concatenation did not to increase the resolution power of the method. On the other hand, sampling extent positively favored the statistical support of the clusters. The phylogeny depicted enable us to emit several hypotheses about the processes guiding the diversification of *Frankia* species; the data obtained here suggest coevolution has a considerable impact in these issues because of the evident influence of the host on *Frankia* evolution. Additional studies must be addressed to verify this hypothesis.

MLSA lead to coherent strain grouping, likely representing some of the species, previously described by DDH and AFLP data. A high correlation between the MLSA genomic distances and the evolutionary genomic divergence calculated from AFLP data was found. All this renders the MLSA and advisable method (together with the AFLP) to replace the DDH in *Frankia* the *bona fide* species delineation.

Bibliography

1. **Akimov, V., and S. Dobritsa.** 1992. Grouping of *Frankia* strains on the basis of DNA relatedness. Syst Appl Microbiol **15:**372-379.
2. **Baker, D.** 1987. Relationships among pure-cultured strains of *Frankia* based on host specificity. Physiol Plant **70:**245-248.
3. **Baker, D., W. Newcomb, and J. Torrey.** 1980. Characterization of an ineffective actinorhizal microsymbiont, *Frankia* sp. EuI1 (Actinomycetales). Canadian Journal of Microbiology **26:**1072-1089.
4. **Baker, D., and J. G. Torrey.** 1980. Characterization of an effective actinorhizal microsymbiont, *Frankia* sp. AvcI1 (Actinomycetales). Canadian Journal of Microbiology **26:**1066-1071.
5. **Becking, J. H.** 1970. Frankiaceae *fam. nov.* (Actinomycetales) with one new combination and six new species of the genus *Frankia* Brunchorst 1886, 174. International Journal of Systematic Bacteriology **20:**201-220.
6. **Benson, D. R., and W. B. Silvester.** 1993. Biology of *Frankia* strains, actinomycete symbionts of actinorhizal plants. Microbiol. Rev. **57:**293-319.
7. **Bishop, C. J., D. M. Aanensen, G. E. Jordan, M. Kilian, W. P. Hanage, and B. G. Spratt.** 2009. Assigning strains to bacterial species via the internet. BMC Biol **7:**3.
8. **Bruen, T. C., H. Philippe, and D. Bryant.** 2006. A simple and robust statistical test for detecting the presence of recombination. Genetics **172:**2665-81.
9. **Burggraaf, A., and W. Shipton.** 1983. Studies on the growth of *Frankia* isolates in relation to infectivity and nitrogen fixation (acetylene reduction). Can J Bot **61:**2774-2782.
10. **Caru, M.** 1993. Characterization of native Frankia strains isolated from chilean shrubs (Rhamnaceae). Plant and Soil **157:**137-145.
11. **Clawson, M., and D. Benson.** 1999. Natural diversity of Frankia strains in actinorhizal root nodules from promiscuous hosts in the family Myricaceae. Appl Environ Microbiol **65:**4521-4527.
12. **Clawson, M. L., A. Bourret, and D. R. Benson.** 2004. Assessing the phylogeny of Frankia-actinorhizal plant nitrogen-fixing root nodule symbioses with Frankia 16S rRNA and glutamine synthetase gene sequences. Mol Phylogenet Evol **31:**131-8.
13. **Coenye, T., D. Gevers, Y. Van de Peer, P. Vandamme, and J. Swings.** 2005. Towards a prokaryotic genomic taxonomy. FEMS Microbiol Rev **29:**147-67.
14. **Cournoyer, B., and C. Lavire.** 1999. Analysis of *Frankia* evolutionary radiation using glnII sequences. FEMS Microbiology Letters **177:**29-34.
15. **Devulder, G., M. Perouse de Montclos, and J. P. Flandrois.** 2005. A multigene approach to phylogenetic analysis using the genus Mycobacterium as a model. Int J Syst Evol Microbiol **55:**293-302.
16. **Diem, H., D. Gauthier, and Y. Dommergues.** 1983. An effective strain of *Frankia* from *Casuarina* sp. Can J Bot **61:**2815-2821.
17. **Diem, H., D. Gauthier, and Y. Dommergues.** 1982. Isolation of *Frankia* from nodules of *Casuarina equisetifolia*. Can J Microbiol **28:**526-530.
18. **Fernandez, M., H. Meugnier, P. Grimont, and R. Bardin.** 1989. Deoxyribonucleic acid relatedness among members of the genus *Frankia*. Int J Syst Bacteriol **39:**424-429.
19. **Gauthier, D., L. Frioni, H. Diem, and Y. Dommergues.** 1984. The *Colletia spinosissima-Frankia* symbiosis. Oecolog Plant **5:**231-239.

20. **Gevers, D., F. M. Cohan, J. G. Lawrece, and o. authors.** 2005. Opinion: Re-evaluating prokaryotic species. Nat Rev Microbiol **3:**733-739.
21. **Gevers, D., F. M. Cohan, J. G. Lawrence, B. G. Spratt, T. Coenye, E. J. Feil, E. Stackebrandt, Y. Van de Peer, P. Vandamme, F. L. Thompson, and J. Swings.** 2005. Opinion: Re-evaluating prokaryotic species. Nat Rev Microbiol **3:**733-9.
22. **Gil, R., F. J. Silva, J. Pereto, and A. Moya.** 2004. Determination of the core of a minimal bacterial gene set. Microbiol Mol Biol Rev **68:**518-37.
23. **Gogarten, J. P., and J. P. Townsend.** 2005. Horizontal gene transfer, genome innovation and evolution. Nat Rev Microbiol **3:**679-87.
24. **Gtari, M., L. Brusetti, G. Skander, D. Mora, A. Boudabous, and D. Daffonchio.** 2004. Isolation of Elaeagnus-compatible Frankia from soils collected in Tunisia. FEMS Microbiol Lett **234:**349-55.
25. **Hall, T. A.** 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symposium Series **41:**95-98.
26. **Hughet, V., J. MacCray Batzli, J. F. Zimpfer, P. Normand, J. O. Dawson, and M. P. Fernandez.** 2001. Diversity and Specificity of Frankia Strains in Nodules of Sympatric Myrica gale, Alnus incana, and Shepherdia canadensis determined by rrs gene polymorphism. Applied and Environmental Microbiology **67:**2116-2122.
27. **Huguet, V., E. O. Land, J. G. Casanova, J. F. Zimpfer, and M. P. Fernandez.** 2005. Genetic diversity of *Frankia* microsymbionts from the relict species *Myrica faya* (Ait.) and *Myrica rivas-martinezii* (S.) in Canary Islands and Hawaii. Microb Ecol **49:**617-25.
28. **Huson, D. H., and D. Bryant.** 2006. Application of phylogenetic networks in evolutionary studies. Mol Biol Evol **23:**254-67.
29. **Jamann, S., M. Fernandez, and A. Moiroud.** 1992. Genetic diversity of Elaeagnaceae-infective *Frankia* strains isolated from various soils. Acta Oecolog **13:**395-405.
30. **Jobb, G., A. von Haeseler, and K. Strimmer.** 2004. TREEFINDER: a powerful graphical analysis environment for molecular phylogenetics. BMC Evol Biol **4:**18.
31. **Jobidon, R., and J. R. Thibault.** 1980. Optimal surface sterilization of Alnus crispa seeds with merthiolate. FOR. SCI. **26:**483-484.
32. **Jolley, K. A., E. J. Feil, M. S. Chan, and M. C. Maiden.** 2001. Sequence type analysis and recombinational tests (START). Bioinformatics **17:**1230-1.
33. **Kim, B. J., J. H. Lee, M. A. Lyu, S. J. Kim, G. H. Bai, G. T. Chae, E. C. Kim, C. Y. Cha, and Y. H. Kook.** 1999. Identificación of mycobacterial species by comparative analysis of the RNA polymerase gene (rpoB). Journal of clinical microbiology **37:**1714-1720.
34. **Lalonde, M., H. Calvert, and S. Pine.** 1981. Isolation and use of *Frankia* strains in actinorhizae formation, p. 296-299. *In* W. N. AH Gibson (ed.), Current Perspectives in Nitrogen Fixation. Australian Academy of Science, Canberra.
35. **Lechevalier, M.** 1986. Catalog of *Frankia* strains. The Actinomycetes **19:**131-162.
36. **Lechevalier, M.** 1984. The taxonomy of the genus *Frankia*. Plant Soil **78:**1-6.
37. **Librado, P., and J. Rozas.** 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics **25:**1451-2.
38. **Marechal, J., B. Clement, R. Nalin, C. Gandon, S. Orso, J. H. Cvejic, M. Bruneteau, A. Berry, and P. Normand.** 2000. A *rec*A gene phylogenetic analysis confirms the close proximity of *Frankia* to *Acidothermus*. Int. J. Syst. Evol. Microbiol. **50:**781-785.
39. **Martens, M., P. Dawyndt, R. Coopman, M. Gillis, P. De Vos, and A. Willems.** 2008. Advantages of multilocus sequence analysis for taxonomic studies: a case study using 10

housekeeping genes in the genus Ensifer (including former Sinorhizobium). Int J Syst Evol Microbiol **58:**200-14.

40. **Mirza, B. S., A. Welsh, J. P. Rieder, M. W. Paschke, and D. Hahn.** 2009. Diversity of frankiae in soils from five continents. Syst Appl Microbiol **32:**558-70.
41. **Mirza, M., D. Hahn, and A. Akkermans.** 1992. Characterization of *Frankia* strains from *Coriaria nepalensis*. Syst Appl Microbiol **15:**289-295.
42. **Murry, M. M., M. S. Fontaine, and J. G. Torrey.** 1984. Growth kinetics and nitrogenase induction in Frankia sp. HFP ArI3 grown in batch culture. Plant and Soil **78:**61-78.
43. **Normand, P., and M. Lalonde.** 1982. Evaluation of *Frankia* strains isolated from provenances of two *Alnus* species. Can. J. Microbiol. **28:**1133-1142.
44. **Normand, P., and M. Lalonde.** 1986. The genetics of actinorhizal Frankia: a review. Plant Soil **90:**429-453.
45. **Normand, P., P. Lapierre, L. Tisa, J. P. Gogarten, N. Alloisio, E. Bagnarol, C. A. Bassi, A. M. Berry, D. M. Bickhard, N. Choisne, A. Couloux, B. Cournoyer, S. Cruveiller, V. Daubin, N. Demange, M. P. Francino, E. Goltsman, Y. Huang, O. Kopp, L. Labarre, A. Lapidus, C. Lavire, J. Marechal, M. Martinez, J. E. Mastronunzio, B. C. Mullin, J. Niemann, P. Pujic, T. Rawnsley, Z. Rouy, C. Schenowitz, A. Sellstedt, F. Tavares, J. P. Tomkins, D. Vallenet, C. Valverde, L. Wall, Y. Wang, C. Medigue, and D. R. Benson.** 2007. Genome characteristics of facultatively symbiotic *Frankia* sp. strains reflect host range and host plant biogeography. Genome Research **17:**7 - 15.
46. **Normand, P., S. Orso, B. Cournoyer, P. Jeannin, C. Chapelon, J. Dawson, L. Evtushenko, and A. K. Misra.** 1996. Molecular phylogeny of the genus *Frankia* and related genera and emendation of the family Frankiaceae. Int. J. Syst. Bacteriol. **46:**1-9.
47. **Ochman, H., J. G. Lawrence, and E. A. Groisman.** 2000. Lateral gene transfer and the nature of bacterial innovation. Nature **405:**299-304.
48. **Posada, D., and T. R. Buckley.** 2004. Model selection and model averaging in phylogenetics: advantages of akaike information criterion and bayesian approaches over likelihood ratio tests. Syst Biol **53:**793-808.
49. **Prin, Y., L. Maggia, B. Picard, H. Diem, and P. Goullet.** 1991. Electrophoretic comparison of enzymes from 22 single-spore cultures obtained from *Frankia* strain ORS140102. FEMS Microbiol Lett **77:**223-228.
50. **Puppo, A., L. Dimitrijevic, H. Diem, and Y. Dommergues.** 1985. Homogeneity of superoxide dismutase patterns in *Frankia* strains from Casuarinaceae. FEMS Microbiol Lett **30:**43-46.
51. **Sawabe, T., K. Kita-Tsukamoto, and F. L. Thompson.** 2007. Inferring the evolutionary history of vibrios by means of multilocus sequence analysis. J Bacteriol **189:**7932-6.
52. **Simonet, P., A. Capellano, E. Navarro, R. Bardin, and A. Moiroud.** 1984. An improved method for lysis of *Frankia* with achromopeptidase allows detection of new plasmids. Can J Microbiol **30:**1292-1295.
53. **Simonet, P., E. Navarro, C. Rouvier, P. Reddell, J. Zimpfer, Y. Dommergues, R. Bardin, P. Combarro, J. Hamelin, A. M. Domenach, F. Gourbiere, Y. Prin, J. O. Dawson, and P. Normand.** 1999. Co-evolution between *Frankia* populations and host plants in the family Casuarinaceae and consequent patterns of global dispersal. Environ. Microbiol. **1:**525-533.
54. **Simonet, P., P. Normand, A. Moiroud, and M. Lalonde.** 1985. Restriction enzyme digestion patterns of Frankia plasmids. Plant and Soil **87:**49-60.

55. **Simonet, P., L. N. Thi, A. Moiroud, and R. Bardin.** 1989. Diversity of *Frankia* strains isolated from a single alder stand. Plant Soil **118:**13-22.
56. **Stackebrandt, E., W. Frederiksen, G. M. Garrity, P. A. Grimont, P. Kampfer, M. C. Maiden, X. Nesme, R. Rossello-Mora, J. Swings, H. G. Truper, L. Vauterin, A. C. Ward, and W. B. Whitman.** 2002. Report of the ad hoc committee for the re-evaluation of the species definition in bacteriology. Int J Syst Evol Microbiol **52:**1043-7.
57. **Thompson, J. D., D. G. Higgins, and T. J. Gibson.** 1994. Clustal W - Improving the sensitivity of progressive multiple alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Research **22:**4673-4680.
58. **Valdés, M.** 2007. Frankia Ecology. Nitrogen-fixing Actinorhizal Symbiosis. In K. Pawlowski and W. E. Newton (Eds)**:**49-71.
59. **Welsh, A., B. S. Mirza, J. P. Rieder, M. W. Paschke, and D. Hahn.** 2009. Diversity of frankiae in root nodules of Morella pensylvanica grown in soils from five continents. Syst Appl Microbiol **32:**201-10.
60. **Welsh, A. K., J. O. Dawson, G. J. Gottfried, and D. Hahn.** 2009. Diversity of Frankia populations in root nodules of geographically isolated Arizona alder trees in central Arizona (United States). Appl Environ Microbiol **75:**6913-8.
61. **Zeigler, D. R.** 2003. Gene sequences useful for predicting relatedness of whole genomes in bacteria. International Journal of Systematic and Evolutionary Microbiology **53:**1893-1900.
62. **Zhang, Z., M. F. Lopez, and J. G. Torrey.** 1984. A comparison of cultural characteristics and infectivity of *Frankia* isolates from root nodules of *Casuarina* species. Plant and Soil **78:**79-90.

CHAPTER III: (Article 3) THE *Alnus*-EFFECTIVE Sp+ *Frankiae* FORM TWO DIVERGENT PHYLOGENETIC GROUPS STRONGLY CORRELATED TO HOST PLANT SPECIES

The Alnus-effective Sp+ Frankiae form two divergent phylogenetic groups strongly correlated to host plant species

Authors

Bautista GHH[1], Herrera-Belaroussi A[1], Xu R[1], Pepin R[1]. and Fernandez MP[1]

Laboratory addresses

[1] Université de Lyon, F-69622, Lyon, France and Université Lyon1, Villeurbanne, CNRS, UMR5557, Ecologie Microbienne, 43, Boulevard du 11 novembre 1918, F-69622 Villeurbanne Cedex, France

Corresponding author :
Hector H. Bautista Guerrero
Ecologie Microbienne UMR5557, 43, Boulevard du 11 novembre 1918, F-69622 Villeurbanne Cedex, France
Tel: (33) 472 44 82 00
Fax: (33) 426 23 44 68
E-mail: hector.bautista@univ-lyon1.fr

Article en préparation

Chapter III: The *Alnus*-effective Sp+ *Frankia*e form two divergent phylogenetic groups strongly correlated to host plant species

Introduction

The filamentous bacterium *Frankia* has been reported for its ability to establish nitrogen-fixing symbiosis with actinorhizal plants, resulting in nodules on plant roots. Since the first reproducible isolation of *Frankia* in 1978 (Callahan et al., 1978), the diversity of its morphological states in pure culture have been largely described, allowing to differentiate three main structures: hyphae (filamentous structure), vesicles (nitrogen-fixing site) and, as most of actinobacteria, sporangia (containing multitude of spores). Spores are likely involved in bacterial dispersion and maybe in the conservation of infective propagules in old culture media and soil (Schwintzer, 1990b). Except some isolates, such as the ArgP5 strain, for which the quantity of sporangia generated has been considered as a particular phenotypic feature, sporulation is in general a consistent property for all *Frankia* strains, occurring in pure culture under a range of conditions. However, once in symbiosis with actinorhizal plants, sporulation does not seem universal. Although filamentous structure of *Frankia* and its terminal swollen vesicles can consistently be observed inside infected nodule cells, spores may or may not be formed within nodule and two *Frankia*-infected nodule phenotypes (Sp+ and Sp-) have been described depending on the presence or the absence of sporangia (van Dijk & Merkus, 1976).

Actually, the Sp+ nodule phenotype has been described in nine of the 25 Actinorhizal plant genera : *Alnus* (*A. glutinosa, A. rubra, A. viridis, A. incana* et *A. cordata*), *Comptonia* (*C. peregrina*), *Myrica* (*M. cerifera* et *M. gale*), *Ceanothus* (*C. velutinus*), *Elaeagnus* (*E. umbellata*), *Hippophae* (*H. rhamnoïdes*), *Casuarina* (*C. cunninghamiana, C. equisetifolia*, *C. glauca*, *C. obesa*), *Purshia* (*P. tridentata*) et *Dryas* (*D. drummondii*) (Schwintzer, 1990a; Torrey, 1987). However the most striking abundance of Sp+ nodules has been documented within the *Alnus* genus.

Considerable evidence indicates that the ability of *in-planta* sporulation would be controlled by the *Frankia* strain infecting the host (Schwintzer, 1990a; Simonet *et al.*, 1994). Indeed, shortly after the first observation of *Frankia* sporangia in nodules (van Dijk & Merkus, 1976), cross-inoculation experiments were performed from different host species (*Alnus, Comptonia* and *Myrica*) as trap plants inoculated with either crushed Sp+ or Sp- nodules (VanDijk, 1978 ; VandenBosh and Torrey, 1985). Interestingly, nodules of the two different phenotypes were obtained and the plant inoculated with Sp+ nodules originated

nodules of the same phenotype (VanDijk, 1978 ; VandenBosh and Torrey, 1985). These results suggested that *in-planta* sporulation is a stable trait of the microsymbiont, hence only Sp+ strains generate Sp+ nodules. (VandenBosch & Torrey, 1985. However, host plant could be involved in the number of sporangia formed within Sp+ nodules and the seasonal sporulation {Schwintzer, 1990 #1266).

Since its description, no successful isolation of Sp+ strains has been achieved, while a large collection of Sp- has been obtained (Torrey, 1987). More precisely, all *Frankia* strains isolated from Sp+ nodules were non-infective or failed to form sporangia when reinoculated in different host species and thus were considered as Sp- strains (Quispel et Tak, 1978 ; Burgraaf et al., 1981 ; Normand et Lalonde 1982).

Because of the absence of Sp+ isolated strain, our understanding of *in-planta* sporulation phenomenon has been deeply limited. A phylogenetic divergence between Sp+ and Sp- strains was observed for the first time from nodules of *A. viridis* collected at the same alder stand, based on partial 16S rRNA sequence analysis (Simonet *et al.*, 1994). Posterior phylogenetic analyses have sporadically included Sp+ strains and in all cases they branched apart. However the number of specimens included was too low, leaving the phylogenetic position of the Sp+ group unclear. Complete 16S rRNA sequence analyses grouped Sp+ strains within the *F. alni* species, that is why the formation of a novel group to contain these strains was not supported (Normand *et al.*, 1996). Moreover the population dynamic of the Sp+ *Frankia* as well as host specificity range of these strains remain an unanswered question.

The aim of this study was therefore to carry out a phylogenetic study on the basis of a multigenic MLSA approach in order to confirm the phylogenetic positioning and genetic diversity of *Frankia* Sp+ strains. This work focused on *Alnus* strains including a large set of reference strains (thus Sp-) together with Sp+ nodular endophytic symbionts from three *Alnus* species, altogether representative of a large collection of geographic sites and ecological conditions.

III.1. Material and Methods

Bacterial strains and culture condition. The isolated *Frankia* strains (thus all Sp- phenotype) used in this study are listed in the Table 1. The strains were grown at 28°C in F

medium containing Tween 80 or in BAP medium (Murry *et al.*, 1984c) with weekly manual shaking.

Nodules. Actinorhizal nodules were collected from three species of *Alnus sp.* (*A. glutinosa* = Ag, *A. viridis* = Av and *A. incana* = Ai), on 16 sites in France and one site in Sweden (collected by A. Sellstedt). Among the French sites, 13 were located in Rhône-Alpes region, three other were located in Bourgogne and Forez regions. Nodules from Sweden were collected in greenhouse plants inoculated with a local Sp+ nodule inoculum (ref). Some of these 16 French sites (Allemont, Bérarde, Forez, Gévoudaz, Thury, Ornon and Toussuire) were previously studied (Table 1) For each site, except Sweden, three different samples, representative of the local heterogeneity were collected.

The distribution of host plant sampled at each site is illustrated figure 1, except for the Sweden site where only nodules from *A. glutinosa* were collected. Host plants were sometime growing in sympatry, such as *A. viridis* and *A. incana* at Ornon or *A. glutinosa* and *A. incana* at Le Bourget en Huile (Marais bord du Gelon).

The majority of the nodules were collected from June 2009 to October 2009, except nodules from Ornon, which were collected in August 2008. For each site (except Sweden), nodules were collected from at least 5 distinct trees. They were all washed in autoclaved distilled water and kept few days at 4°C for analysis, before being stored at -20°C.

Microscopic nodule examinations. Nodule phenotype (Sp+ or Sp-) was determined **o**n an average of three nodules per site. From each nodule, two adjacent lobes were used for microscopic observation and DNA extraction, respectively. Nodule phenotype was currently identified on ten to fifteen transversal hand sections stained with Lactophenol Blue (Réactifs RAL, Martillac, France) for 2 min at room temperature (Kurdali *et al.*, 1990). A Sp+ phenotype was identified by the presence of a great number of sporangia. When no or few sporangia were observed a double-staining protocol applied on semithin sections (adapted from (Pépin & Boumendil, 1982) was used to confirm the phenotype, . All steps were performed at room temperature. Longitudinal sections of nodules were fixed for 3 h in 2% glutaraldehyde–0.5% paraformaldehyde in a 0.2M Mac Ilvaine (citrate-phosphate) buffer at pH 7. They were subsequently washed overnight in 0.2M Mac Ilvaine buffer (pH 7). After dehydration in a graded series of ethanol solutions, the samples were embedded in LR-White resin. Polymerization of the resin was carried out at 56°C for 24 hours in capsules to avoid

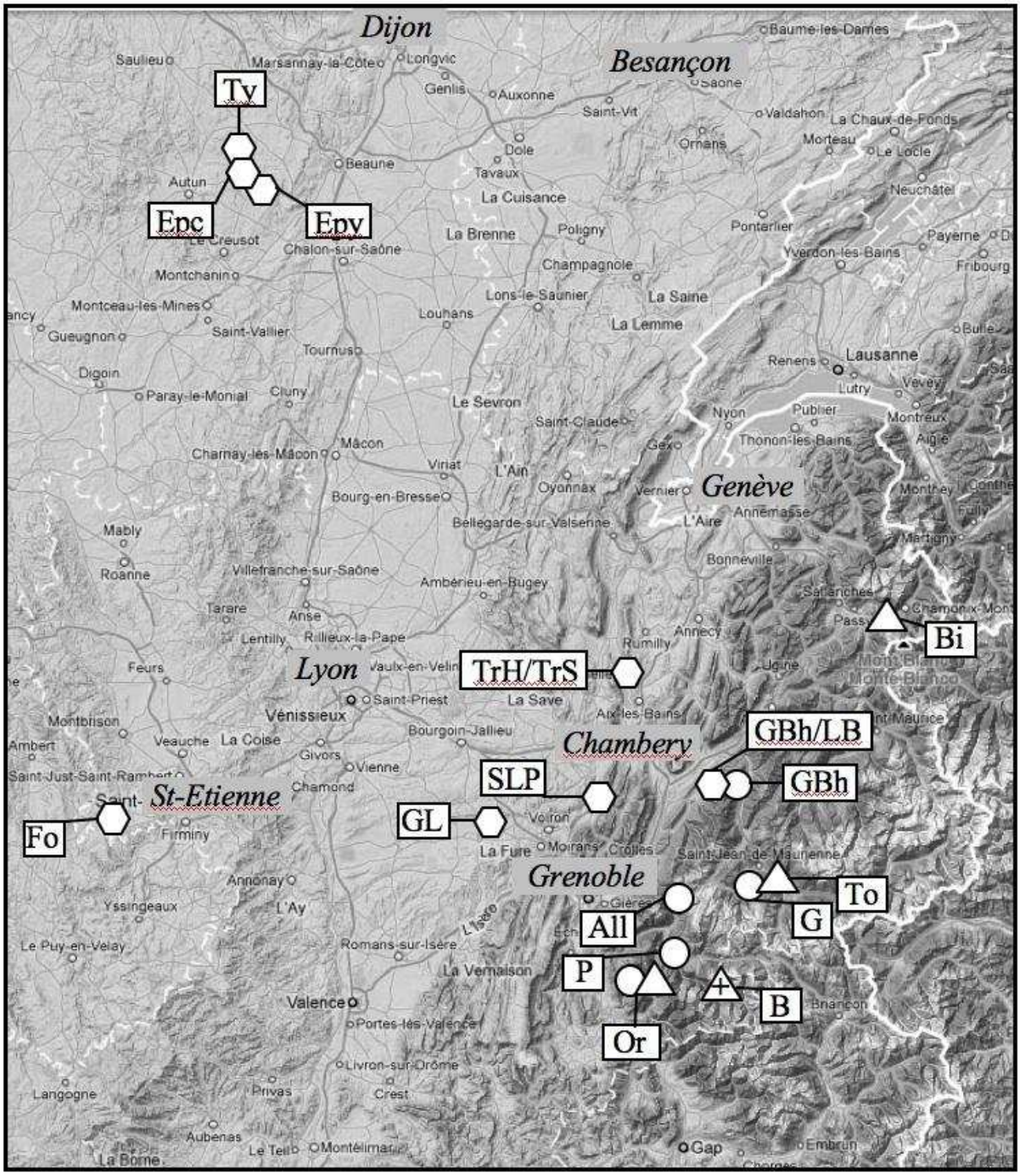

Figure 1. Geographical location of the different sampled sites. Correspondence between the nomenclature of sites are found in the table 1.

Table 1. Liste of Frankia reference strains and actinorhizal nodules

Strain	Host plant	Geographic origin*	Phenotype
Pure strain			
3244	*Alnus incana ssp. rugosa*	Vermont, USA	Sp-
Ac218	*Alnus cordata*	Miribel, France	Sp-
Ac2323	*Alnus cordata*	Saou (France)	Sp-
Ac2340	*Alnus cordata*	Orléans, France	Sp-
Ac24I5	*Alnus cordata*	Orléans, France	Sp-
ACN14a	*Alnus crispa*	Tadoussaq, Canada	Sp-
AcoN24d	*Alnus cordata*	Orléans, France	Sp-
AcVc1	*Alnus cordata*	Corsica, France	Sp-
Ag21d1	*Alnus glutinosa*	Corsica, France	Sp-
$Ag24_{251}$	*Alnus glutinosa*	Orléans, France	Sp-
Ai966	*Alnusincana*	101 (Lacrans (01), France	Sp-
Aj01	*Alnus japonica*	Makabe, Ibaraki Pref., Japan	Sp-
Ar24H5	*Alnus rubra*	Orléans, France	Sp-
ArgN22d	*Alnus rugosa*	Québec, Canada	Sp-
ARgP5	*Alnus rugosa*	Québec, Canada	Sp-
ArI3	*Alnus rubra*	Oregon, USA	Sp-
AV597	*Alnus viridis*	Alpe du Grand Serre, France	Sp-
AvCI1	*Alnus viridis ssp. crispa*	Ontario, Canada	Sp-
AVL3	*Alnus viridis*	Lautaret, France	Sp-
AVN17o	*Alnus viridis*	La Toussuire, France	Sp-
I38	*Alnus incana*	La pallud, France	Sp-
M16467	*Myrica pensilvanica*		Sp-
M16477	*Myrica pensilvanica*		Sp-
Nodules			
AgEpc	*Alnus glutinosa*	Epinac	Sp-
AgEpy	*Alnus glutinosa*	Epertully	Sp-
AgFo3	*Alnus glutinosa*	Forez	Sp-
AgGBh	*Alnus glutinosa*	Le bourget en Huile - Marais Bord du Gelon	Sp+
AgGL1	*Alnus glutinosa*	Réserve naturelle du Grand Lemps	Sp-
AgLB	*Alnus glutinosa*	Le bourget en Huile - Le Blanchet	Sp-

AgTrH1	*Alnus glutinosa*	Le bourget du lac - Le Tremblay (zone humide)	Sp+
AgTrS1	*Alnus glutinosa*	Le bourget du lac - Le Tremblay (zone sèche)	Sp+
AgTyI.1	*Alnus glutinosa*	Thury site I	Sp+
AgTyII.3	*Alnus glutinosa*	Thury site II	Sp+
AgTyIV	*Alnus glutinosa*	Thury site IV	Sp-
AgTyV	*Alnus glutinosa*	Thury site V	Sp-
AiAII	*Alnus incana*	Allemont	Sp-
AiG1	*Alnus incana*	Gevoudaz	Sp-
AiGBh	*Alnus incana*	Le bourget en Huile - Marais Bord du Gelon	Sp+
AiOrII	*Alnus incana*	Ornon	Sp-
AiP	*Alnus incana*	La Paute	Sp+
AvBI.5	*Alnus viridis*	La Berrarde site I	Sp+
AvBIII	*Alnus viridis*	La Berrarde site III	Sp+
AvBIV	*Alnus viridis*	La Berrarde site IV	Sp+
AvBi200	*Alnus viridis*	Bionnassay	Sp+
AvOr2	*Alnus viridis*	Ornon	Sp+
AvToI.1	*Alnus viridis*	Toussuire	Sp-

*The nodule location are also depicted in the figure 1.

oxygene. Then, 0.8 µm semithin sections were cut using a Leica Ultracut E ultramicrotome and dried on glass slides by heating 10 minutes at 40°C and 5 minutes at 100°C. They were finally stained with Coleman's Shiff reagent (Hotchkiss, 1948) for 10 minutes at 40-45°C followed by 1% Malachite for 40 secondes at 40-45°C. Nodule semithin sections were observed under light microscopy using a Nikon OptiPhot2 and microphotographs were taken using a Nikon Type D70.

DNA extraction, PCR amplification and sequencing.

DNA extraction was performed from *Frankia* pure strains by using DNAeasy Plant Mini (Qiagen) according to manufacturer's instructions.

Total nodule DNA was extracted from one single lobe as previously described (Hughet *et al.*, 2001) with slight modifications. After washing, the outer layers were removed, then the lobe was crushed in 300 ml of TCP extraction buffer (100 mM Tris-HCl [pH 7], 0.5 M NaCl, 50 mM EDTA [pH 8], 2% [wt/vol] cethylmethylammonium bromide [Sigma, St. Louis, Mo.], 1% [wt/vol] polyvinylpolypyrrolidone [Sigma]). The homogenate was incubated at 65°C for 1h and centrifuged twice at 6000 *X g* for min. The supernatant was extracted with an equal volume of Chloroform-isoamyl alcohol (24:1, vol/vol) and centrifuged at 13000 *X g* for 20 min. DNA from aqueous phase was precipitated in ethanol for at least 2h at -20°C. The sample was then centrifuged at 13,000 *X g* for 30 min, and the resulting DNA pellet was washed with 70% (vol/vol) ethanol, air dried, and dissolved in 10 µL of tris-EDTA (TE) buffer (pH 7.5).

The five loci were amplified by a simple PCR reaction in a final volume of 100 µl. Each PCR reaction contained 5 µL template ADN, 10 µl of 10X PCR buffer, 5 µl of each primer (10 mM) (Table 2), 10 µL of a dNTP mix (2.5 mM), 4 µL of $MgCl_2$ (25 mM), 5 µL DMSO, 2.5 U Taq DNA polymerase and 60.5 µL of sterile MiliQ water. The reaction conditions and sequencing were already reported (Bautista GHH et al, 2010).

Sequence analysis

The consensus sequence for each gene fragment was determined by alignment of the forward and reverse sequences using BioEdit version 7.0.5.3 (http://www.mbio.ncsu.edu/BioEdit/bioedit.html) (Hall, 1999) and the alignments were

performed with the Clustal W as implemented in BioEdit (Thompson *et al.*, 1994). Statistics for each locus, such as the number and proportion of polymorphic sites, mean G+C content and ratio of mean synonymous substitutions per synonymous site/mean nonsynonymous substitutions per nonsynonymous site (d*S*/d*N*) were calculated using the START2 program (Jolley *et al.*, 2001). Aligment concatenation was made using DnaSP5 (Librado & Rozas, 2009). Maximum Likelihood analyses were carried out in TREEFINDER (Jobb *et al.*, 2004) by using the GTR + Γ model as the best-fitting model of nucleotide substitution was chosen. The node reliability was assessed using LR-ELW edge support (Jobidon & Thibault, 1980).
For each locus, the sequences obtained for all the samples were compared, alleles sequences that differed from each other by one or more polymorphisms were assigned a unique allele number with help of the NRBD tool (http://linux.mlst.net/nrdb/nrdb.htm). Each unique allele profile as defined by the allele numbers of the five loci was assigned a sequence type (ST) number.
The method of split decomposition was used to assess the degree of tree-like structure present in the alleles found for the concatenation.

III.2. Results

Microscopic phenotype determination

In most cases hand sections were sufficient for phenotype determination. When sporangia were scarced or numerous starch granules present in nodule tissues, the double staining protocol, which differentiated in green stained sporangia and in red stained hyphae and vesicles (Fig. 2), allowed to achieve a precise diagnostic. Nodule phenotypes for each site are listed at Table 1.

Phylogenetic analysis of *Alnus Frankia* strains

The sequence of an internal portion of five housekeeping genes was obtained for isolated *Frankia* strains and nodular samples exhibiting Sp+ and Sp- phenotypes. The number of sequence type (ST) for each loci ranged from 29 to 31, suggesting that all loci are evolving at similar rates. The locus *pgk* yielded the maximal divergence with 38.7 of polymorphic sites

Table 2. Characteristics of loci included in the MLSA analysis.

Locus	Fragment size (bp)	Mean G+C content (%)	Allele number	Number of polymorphic sites (%)	Overall mean divergence	Tajima's D test	dN/dS	Phi test*
atpD	714	68.8	32	196 (27.5)	0.050	0.56	0.096	0.34
dnaA	623	66.5	32	158 25.3)	0.046	0.22	0.057	0.54
ftsZ	718	69.7	32	177 (24.7)	0.036	0.59	0.049	0.59
pgk	695	73.6	31	269 (38.7)	0.113	0.53	0.2057	0.86
rpoB	399	69.6	29	79 (19.8)	0.038	-0.690	0.070	0.015

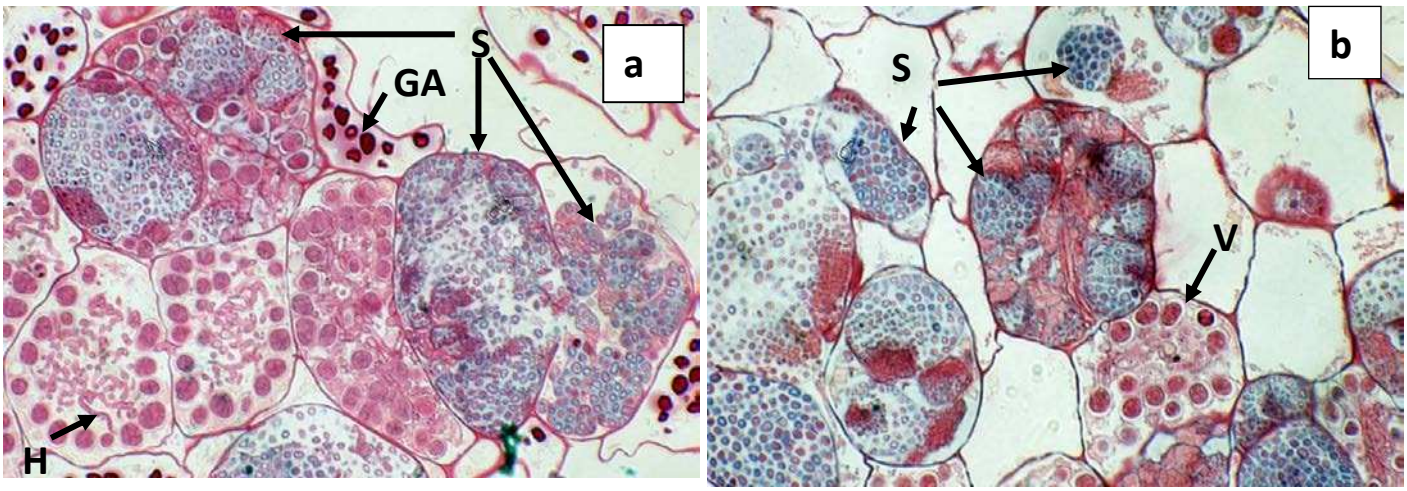

Fig 2: Photograph of semithin sections (0,8 µm) from the AV17Nb nodule stained with the double-staining protocol. (S) Sporangia stained in blue and in red the (H) Hyphae (V) Vesicles and (GA) Starch granules.

and displayed an overall mean divergence of 0.113. According with the Tajima's D test and the d*N*/d*S* ratio (Table 2), the five loci are subjected to stabilizing selection within the population of samples here studied. Individual phylogenies for each locus were constructed and compared (data not shown) and since they were largely congruent the five fragments were concatenated for a final sequence of 3189 nt per sample presenting an overall mean divergence of 0.038. The phylogeny was constructed with the Maximum likelihood method (Fig. 3) using *Frankia* EaN1pec (*Elaeagnaceae*-infective strain) as outgroup. The tree depicted the classical phylogeny of the major cluster I (See Chapter II), which includes the effective *Frankia* strains of *Alnus* sp. and *Myrica* sp, *Casuarina* sp. and is subdivided in several genomospecies. The Sp+ strains grouped in two different and coherent clusters, supported by high LR-ELW (100). Sp+ groups are consistently correlated with the source host plant as follows:

(i) A. viridis and A. incana Sp+ strains

A. viridis and *A. incana* Sp+ strains formed two closely related groups (LR-ELW = 100) clearly correlated with host *Alnus* species and they were included within *F. alni*, since they exhibited 96. 7% of sequence similarity.

The sub-cluster of *A. viridis* Sp+ strains contains nodular symbionts originating from three geographically separate sites. These strains represent a clonal group as they exhibit 99.9% of overall sequence similarity and yielded only two different ST. This group might correspond to that already described by Normand et al. (1996). In their work, *A. viridis* Sp+ strains displayed 98.8% of 16S rRNA similarity with respect to *F. alni* and thus were not considered as a different group. In our study this group displayed 96.5% of divergence with *F. alni*. According to our previous extensive MLSA study we concluded this group should be included in *F. alni* species.

The second sub-cluster grouped *A. incana* Sp+ strains. This group comprises the samples collected at four different geographic sites. This group also seems clonal since all its members exhibited a single ST and shared 99.6% of sequence similarity. *A. incana* and *A. viridis* Sp+ formed a sub-branching of the major group I and their close relatedness suggests they share a common ancestor (LR-ELW= 96). Nodules from Ornon where *A. incana* and *A. viridis* inhabit sympatrically showed a strong correlation between plant and endophyte genotypes since they fell all within their corresponding *A. incana* or *A. viridis* Sp+ group of strains

depending on their original host. It is worth noting that the *A. viridis* Sp- nodules from this site (AvOr2) grouped out from GS 2.

(iii) A. glutinosa Sp+ strains

A. glutinosa symbionts collected in four different French sites and a fifth one collected in Sweden formed a coherent (LR-ELW= 99) group, forming a divergent cluster within the *Alnus-Myrica* infective strains. This group, described here for the first time exhibits a single ST (divergence was found only in the *pgk* sequence) with 93% of sequence similarity with respect to *F. alni*. Compared with likely clonal structure of *A. glutinosa* Sp + strains, Sp- nodules fell in different clusters, either close to Sp+ strains as was the case for AgLB, AgEpy and AgEpc, AgTyII.3, or within genomospecies 1 and 2 (AgGBh and AgGL1).

According to the thresholds established for the species definition in *Frankia* by MLSA (Chapter II of this study) this group of strains might represent a new genomospecies.

Recombination

Evidence for recombination in the five loci was investigated by using the program SplitsTree. The *rpoB* is the only locus exhibiting significant evidence of recombination, p= 0.087. The split decomposition analysis of the concatenate, displayed a star-like network with parallelogram depiction consistent with a recombinational population structure. The linear Sp+ branches suggest restricted allelic recombination (Fig. 4), except between the *A. viridis* and *A. incana* Sp+ strains (See zoom in the Fig. 4)

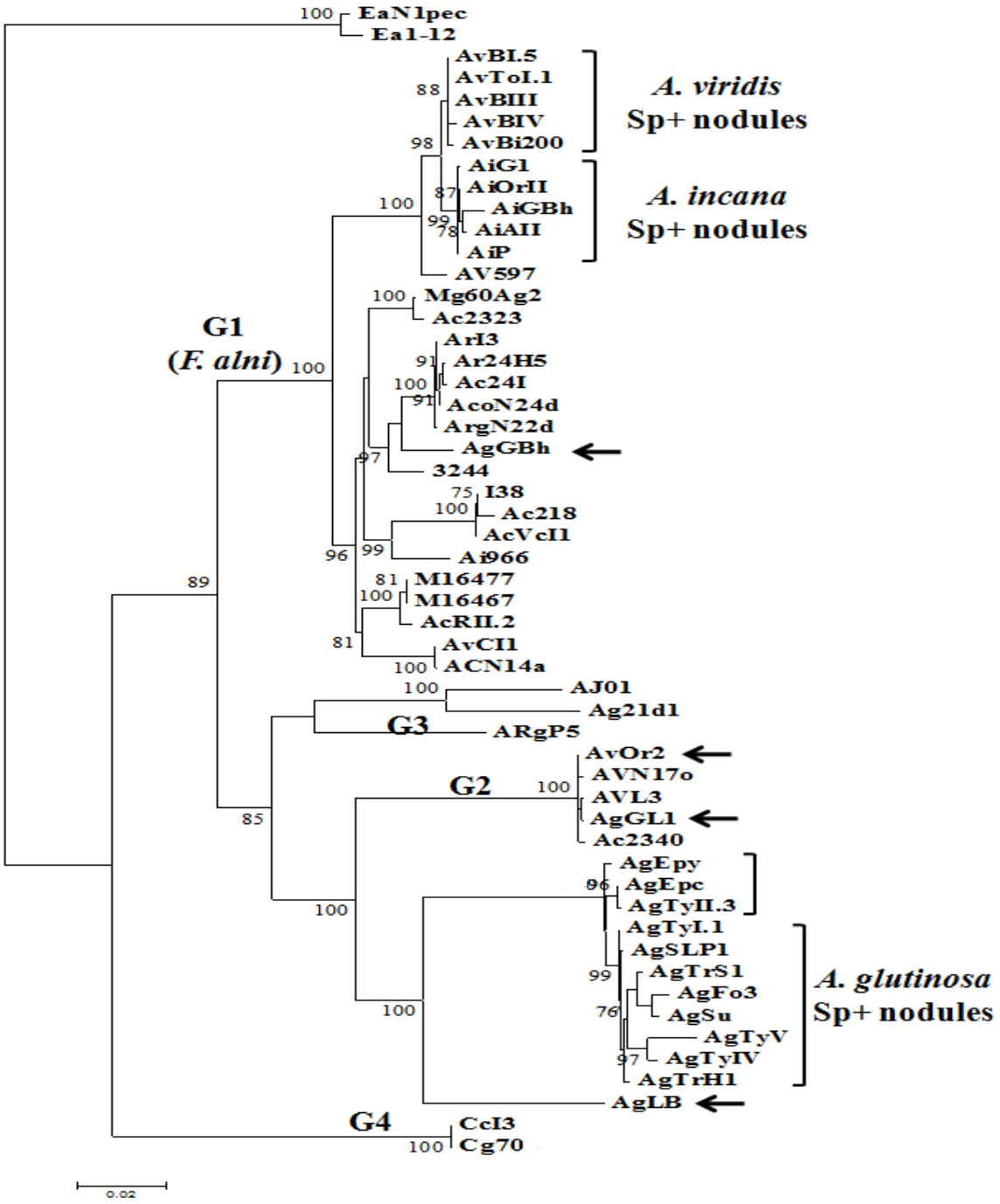

Fig. 3 Phylogenetic reconstruction based on the concatenated *atpD*, *dnaA*, *ftsZ*, *pgk* and *rpoB* gene sequences. Analyses were conducted using the ML method. LR-ELW values of ≥ (using 100 replicates) are indicated at the branching points. Bar, 0, 02% estimated substitutions.genomospecies (G) are named on the basis of the AFLP classification. Arrows indicates the Sp- strains.

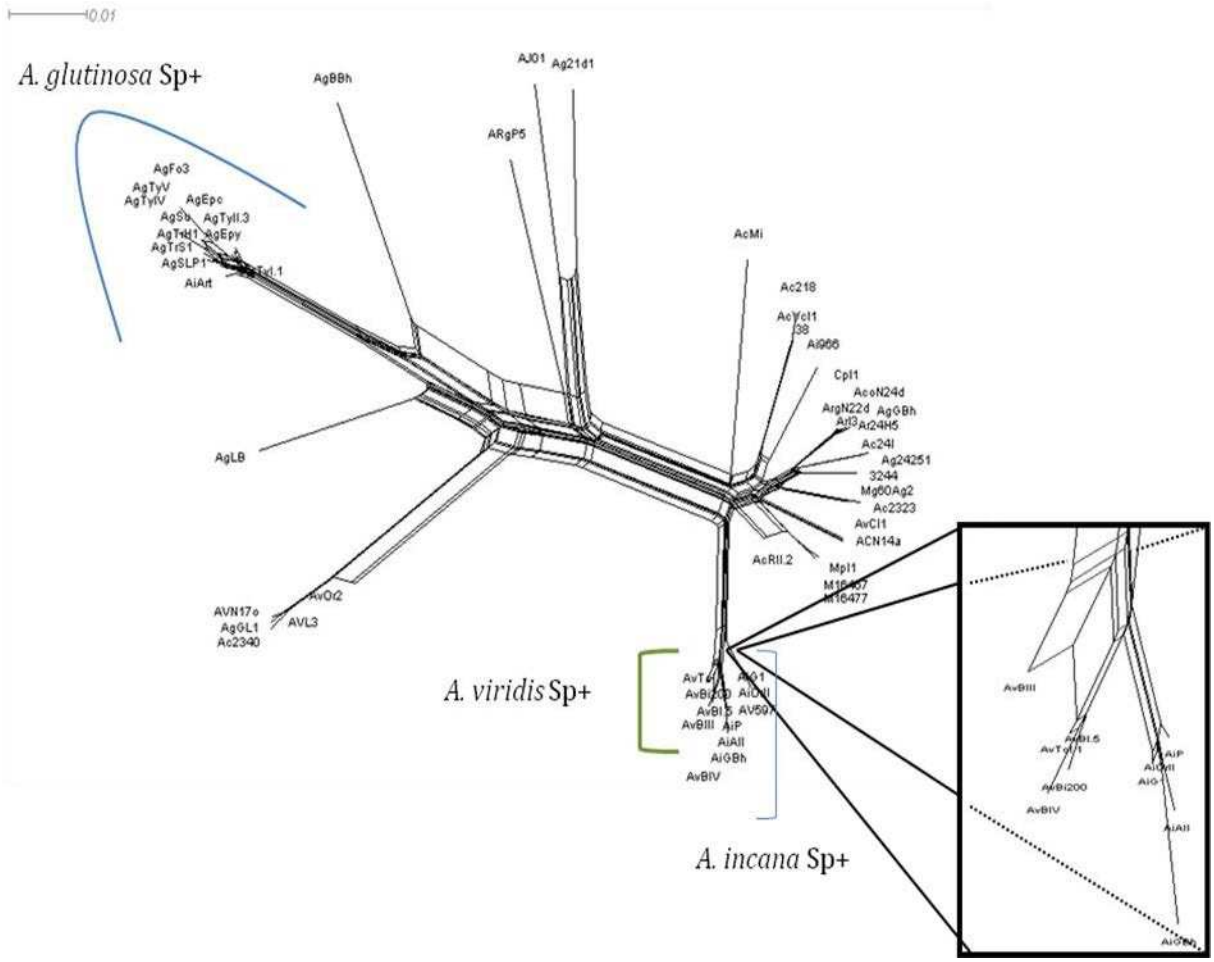

Fig 4. Split decomposition analysis show a bushy network. NeighbourNet graph of the concatenated sequences of five loci with significant (p= 1.27) evidence of recombination with PHI test constructed in SplitsTree v4.0. The zoom indicates the *A. viridis* Sp+ and *A. incana* Sp+ branching.

III.3. Discussion

All *Frankia* strains to date isolated *in vitro* produce spores (Benson & Silvester, 1993). However the spore production *in planta* seems to be restricted to particular groups of *Frankia* strains, and is not widespread within the genus. Hence two major groups of *Frankia* strains can be delineated on the basis of sporulation *in planta*, Sp+ and Sp- (van Dijk & Merkus, 1976).

Frankia, as several actinobacteria posses the ability to sporulate, since in the genome of the three sequenced strains, key genetic sporulation determinants such as the transcription factor WhiB and the genes *ssgA and ssgB* are present and in several copies (Normand *et al.*, 2007a; Traag & van Wezel, 2008), suggesting that differences between the two phenotypes is likely due to regulatory sporulation processes. The phylogenetic positioning of the Sp+ strains within *Frankia* genus was until now controversial and scarcely studied (Normand *et al.*, 1996; Simonet *et al.*, 1994)

In this study we carried out an extensive phylogenetic study including a number of Sp+ or Sp- endophytic *Alnus* symbionts. A MLSA approach, using partial sequence of the five housekeeping genes, already tested to depict the phylogeny of the genus was performed (Chapter II of this work).

In the concatenated-based dendrogram the Sp+ strains formed two coherent and very divergent clusters (94 % of similarity), clearly separated from the Sp- strains. The phylogeny displayed, also allowed state that the Sp+ *Frankia* exhibit marked host specificity. Cross-inoculation trials results support the narrow infection spectrum of Sp+ strains, for example *A. incana* Sp+ strains did not infect *A. glutinosa* (Schwintzer, 1990a; van Dijk *et al.*, 1988).

Without exception all the Sp+ specimens grouped in agreement with the host plant from which they were collected. While *A. incana* and *A. viridis* Sp+ grouped in *F. alni* species the *A. glutinosa* Sp+ and three Sp- strains formed a new genomic group within *Alnus* strains which can doubtless have the ranking of new species. Moreover a high genetic homogeneity was found, and hence all the samples displayed the same sequence type no matter the geographic provenance of the samples.

The *A. viridis* and *A. incana* strains are closely related and together form a major subrgroup that is then divided into two groups. The two groups exhibit also a clonal behaviour as they

yielded a low number of sequence type, respectively. It is interesting to note that in any sampled site, the *A. viridis* Sp+ and *A. incana* Sp+ were found cohabiting in sympatry, suggesting that host plant specialization was acquired before the dispersion of the two types of strains. Besides, there is no congruence between the phylogenetic relatedness of two hosts and high similarity of Sp+ strains infecting them, indeed the species *A. glutinosa* and *A. incana* belongs to the same genomic group, while *A. viridis* represents and ancient lineage of the genus (Navarro *et al.*, 2003). Hence, coevolution of Sp+ and their host seems to be a recent event.

In addition to the high specificity for the host, the clonal behaviour was the other major characteristic of the three Sp+ genomic groups. Low genetic diversity within the group of Sp+ *Frankia* strains inhabiting the same alder stand was reported (Simonet *et al.*, 1994). In this study distant geographic sites were sampled and the same sequence type was found for each Sp+ group, this fact was also evidenced by the splittree test, suggesting that such groups represents rather stable clones with the presence of recombination events restricted to the interior of each population (Fig. 3). The presence of identical sequence types implies that strains from the same genomic group share at minimum four of the five alleles.

Besides, it was expected to achieve maximal discrimination between Sp+ and Sp- strains with the *ftsZ* gene, because it is a housekeeping function directly implicated in the sporulation process (Grantcharova *et al.*, 2005) and therefore it might be subjected to different selective pressure in the two type of strains. However it was *pgk* the gene with the maximum discriminative power while *ftsZ* proved to be the second marker most conserved in the analysis. *ftsZ* is also the main cell division protein in bacteria (McCormick *et al.*, 1994), being that the reason of its highly conserved sequence.

It has been reported that genetic monomorphic microorganisms with apparent clonal populations structures are either recently evolved or have recently experienced a bottleneck (Achtman, 2008). By analyzing the phylogenetic tree depicted by the MLSA analysis, one can state the *A. viridis* and *A. incana* Sp+ lineages represent recent branching events of *F. alni*. Moreover the *A. glutinosa* Sp+ ancestor belongs to a divergent genetic cluster, though the Sp+ group speciation is quite recent.

In an ecological frame, the Sp+ *Frankia* strains seem to represent the final stage in the succession of *Frankia* population in Alder stands (van Dijk, 1984). The geographical distribution of the Sp+ seems to vary with the age of the substrate. It has been shown that in

alder stands the Sp+ nodules are common in the old regions and in abandoned fields while the Sp- dominate the young regions (Simonet *et al.*, 1994; van Dijk, 1984) suggesting that a given actinorhizal plant stand is first colonized is made by *Frankia* Sp- strains, this population is being later displaced by the Sp+ type strains.

The *Frankia* Sp+ strains are refractory to the *in vitro* isolation and culture, moreover in soils devoid of host plants, *Frankia* strains induce the formation of Sp- nodules and no Sp+ (van Dijk, 1984), suggesting that Sp+ show low adaptation to saprophytic life-style and consequently a high metabolic dependence of the host. Conversely, the Sp+ strains are characterized by their low nitrogen fixation efficiency, providing to the host a lesser benefice than the Sp- (Monz & Schwintzer, 1989). In summarize, the Sp+ strains likely exhibits a selfish behaviour, by efficiently establishing symbiosis and performing a poor nitrogen fixation. Similar behaviour has been described in the Rhizobia-Legumes symbiosis (Provorov & Vorobyov, 2008; Rankin *et al.*, 2007; Schumpp & Deakin, 2010), in which some strains literally "cheat" their partners taking benefits but not reciprocating them. Thus the hypothesis is that Sp+ strains exhibits elevated host dependence, by this reason they likely developed strategies to become more efficient in nodulation but they do not represent a potential benefice to their host. All this allow to suggest that the Sp+ could be a good model to study the host-sanction model as they likely are cheater microorganisms.

In summary this study shows the existence of the Sp+ *Frankia* strains, as a stable phenotype and as independent phylogenetic group. These strains exhibit clonal behaviour and high host specificity. The sporulation *in planta* is likely the result of the expression and/or regulation of the genetic information contained in the Sp+ *Frankia* genome rather than a processes dictated by the host plant. It remains unknown the reason of the Sp+ population establishment at advanced successional stages of actinorhizal plant stands, it may first suggest this relationships implies a more equilibrated interaction between the two symbionts. The second hypothesis depicts an inverse overview, in which Sp+ act as selfish organisms ignoring the symbiotic rules and looking only for their fitness.

The extent of the influence of physiological stage of the host as well as the ecological characteristics of the stand over the Sp+ population is another subject of study. Due to their particular ecological properties the Sp+ strains are an interesting model of the actinorhizal symbiosis, that encouraged the application of genomic and population diversity studies in

order to unveil the genetic and evolutionary mechanisms implicated in this symbiotic relationship.

Bibliography

1. **Achtman, M.** 2008. Evolution, population structure, and phylogeography of genetically monomorphic bacterial pathogens. Annu Rev Microbiol **62:**53-70.
2. **Benson, D. R., and W. B. Silvester.** 1993. Biology of *Frankia* strains, actinomycete symbionts of actinorhizal plants. Microbiol. Rev. **57:**293-319.
3. **Callaham, D., P. Del Tredici, and J. Torrey.** 1978. Isolation and cultivation *in vitro* of the actinomycete causing root nodulation in *Comptonia*. Science **199:**899-902.
4. **Fernandez, M., H. Meugnier, P. Grimont, and R. Bardin.** 1989. Deoxyribonucleic acid relatedness among members of the genus *Frankia*. Int J Syst Bacteriol **39:**424-429.
5. **Grantcharova, N., U. Lustig, and K. Flardh.** 2005. Dynamics of FtsZ assembly during sporulation in Streptomyces coelicolor A3(2). J Bacteriol **187:**3227-37.
6. **Hall, T. A.** 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symposium Series **41:**95-98.
7. **Hotchkiss, R. D.** 1948. A microchemical reaction resulting in the staining of polysaccharide structures in fixed tissue preparations. Arch Biochem **16:**131-41.
8. **Hughet, V., J. MacCray Batzli, J. F. Zimpfer, P. Normand, J. O. Dawson, and M. P. Fernandez.** 2001. Diversity and Specificity of Frankia Strains in Nodules of Sympatric Myrica gale, Alnus incana, and Shepherdia canadensis determined by rrs gene polymorphism. Applied and Environmental Microbiology **67:**2116-2122.
9. **Huson, D. H., and D. Bryant.** 2006. Application of phylogenetic networks in evolutionary studies. Mol Biol Evol **23:**254-67.
10. **Jobb, G., A. von Haeseler, and K. Strimmer.** 2004. TREEFINDER: a powerful graphical analysis environment for molecular phylogenetics. BMC Evol Biol **4:**18.
11. **Jobidon, R., and J. R. Thibault.** 1980. Optimal surface sterilization of Alnus crispa seeds with merthiolate. FOR. SCI. **26:**483-484.
12. **Jolley, K. A., E. J. Feil, M. S. Chan, and M. C. Maiden.** 2001. Sequence type analysis and recombinational tests (START). Bioinformatics **17:**1230-1.
13. **Kurdali, F., G. Rinaudo, A. Moiroud, and A. Domenach.** 1990. Competition for nodulation and $^{15}N_2$-fixation between a Sp^+ and a Sp^- Frankia strain from *Alnus incana*. Soil Biol Biochem **22:**57-64.
14. **Librado, P., and J. Rozas.** 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics **25:**1451-2.
15. **McCormick, J. R., E. P. Su, A. Driks, and R. Losick.** 1994. Growth and viability of Streptomyces coelicolor mutant for the cell division gene ftsZ. Mol Microbiol **14:**243-54.
16. **Monz, C., and C. Schwintzer.** 1989. The physiology of spore-negative and spore-positive nodules of *Myrica gale*. Plant Soil **118:**75-87.
17. **Murry, M. M., M. S. Fontaine, and J. G. Torrey.** 1984. Growth kinetics and nitrogenase induction in Frankia sp. HFP ArI3 grown in batch culture. Plant and Soil **78:**61-78.
18. **Navarro, E., J. Bousquet, A. Moiroud, A. Munive, D. Piou, and P. Normand.** 2003. Molecular Phylogeny of Alnus (Betulaceae), inferred from Nuclear Ribosomal DNA ITS sequences. Plant & Soil **00:**000-000.
19. **Normand, P., P. Lapierre, L. Tisa, J. P. Gogarten, N. Alloisio, E. Bagnarol, C. A. Bassi, A. M. Berry, D. M. Bickhard, N. Choisne, A. Couloux, B. Cournoyer, S.**

Cruveiller, V. Daubin, N. Demange, M. P. Francino, E. Goltsman, Y. Huang, O. Kopp, L. Labarre, A. Lapidus, C. Lavire, J. Marechal, M. Martinez, J. E. Mastronunzio, B. C. Mullin, J. Niemann, P. Pujic, T. Rawnsley, Z. Rouy, C. Schenowitz, A. Sellstedt, F. Tavares, J. P. Tomkins, D. Vallenet, C. Valverde, L. Wall, Y. Wang, C. Medigue, and D. R. Benson. 2007. Genome characteristics of facultatively symbiotic *Frankia* sp. strains reflect host range and host plant biogeography. Genome Research **17:**7 - 15.

20. **Normand, P., S. Orso, B. Cournoyer, P. Jeannin, C. Chapelon, J. Dawson, L. Evtushenko, and A. K. Misra.** 1996. Molecular phylogeny of the genus *Frankia* and related genera and emendation of the family Frankiaceae. Int. J. Syst. Bacteriol. **46:**1-9.
21. **Pépin, R., and J. Boumendil.** 1982. Préservation de l'ultrastructure du sclérote de Sclerotina tuberosa (champignon discomycètes); un modèle pour le préparation des échantillons imperméables et hétérogènes. Cytologia **47:**359-377.
22. **Posada, D., and T. R. Buckley.** 2004. Model selection and model averaging in phylogenetics: advantages of akaike information criterion and bayesian approaches over likelihood ratio tests. Syst Biol **53:**793-808.
23. **Schwintzer, C.** 1990. Spore-positive and spore-negative nodules, p. 177-193. *In* J. T. CR Schwintzer (ed.), The Biology of *Frankia* and Actinorhizal Plants. Academic Press, New York, NY.
24. **Schwintzer, C. R.** 1990. Spore-positive and spore-negative nodules, p. 177-193. *In* C. R. a. T. Schwintzer, J.D. (ed.), The biology of Frankia and actinorhizal plants. Academic Press, Inc., San Diego.
25. **Simonet, P., M. Bosco, C. Chapelon, A. Moiroud, and P. Normand.** 1994. Molecular characterization of *Frankia* microsymbionts from spore-positive and spore-negative nodules in a natural alder stand. Appl. Environ. Microbiol **60:**1335-1341.
26. **Smolander, A., and V. Sundman.** 1987. *Frankia* in acid soils of forests devoid of actinorhizal plants. Physiol Plant **70:**297-303.
27. **Thompson, J. D., D. G. Higgins, and T. J. Gibson.** 1994. Clustal W - Improving the sensitivity of progressive multiple alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Research **22:**4673-4680.
28. **Torrey, J.** 1987. Endophyte sporulation in root nodules of actinorhizal plants. Physiol Plant **70:**279-288.
29. **Traag, B. A., and G. P. van Wezel.** 2008. The SsgA-like proteins in actinomycetes: small proteins up to a big task. Antonie Van Leeuwenhoek **94:**85-97.
30. **van Dijk, C.** 1984. Ecological aspects of spore formation in *Frankia-Alnus* symbiosis. Ph.D. thesis. Leiden, The Netherlands, State University.
31. **van Dijk, C.** 1978. Spore formation and endophyte diversity in root nodules of *Alnus glutinosa* (L.) Vill. New Phytol **81:**601-615.
32. **van Dijk, C., and E. Merkus.** 1976. A microscopical study of the development of a spore-like stage in the life cycle of the root nodule endophyte of *Alnus glutinosa* (L.) Gaertn. New Phytol **77:**73-91.
33. **van Dijk, C., A. Sluimer, and A. Weber.** 1988. Host range differentiation of spore-positive and spore-negative strain types of *Frankia* in stands of *Alnus glutinosa* and *Alnus incana* in Finland. Physiol Plant **72:**349-358.

34. **VandenBosch, K., and J. Torrey.** 1985. Development of endophytic *Frankia* sporangia in field- and laboratory-grown nodules of *Comptonia peregrina* and *Myrica gale*. Amer J Bot **72:**99-100.

CHAPTER IV: PHYLOGENETIC ANALYSIS OF ECOLOGICALLY RELEVANT MARKERS

Introduction

The multilocus analysis based on the housekeeping gene sequences (MLSA), has shown to be an accurate tool in the phylogenetic history description of many bacterial genera (Fraser *et al.*, 2010). Many housekeeping genes can evolve faster than the 16S rRNA and are subjected to selectively neutral mutation rates. Variation in their sequences reliably reflect the phylogeny of the group in study and the evolution of its core genome (Gevers *et al.*, 2005a).

However, they present limitations in resolving phylogenetic relationships of highly clonal bacteria (Hanage *et al.*, 2006). For instance in epidemiology studies, bacteria associated with distinct human or animal disease and historically placed in different species were found to be a unique clone with distinctive biology and ecology expression (Hanage *et al.*, 2006).

The phenotypic properties of an isolate play an essential role on its ecological behaviour, however these characteristics are not included in the MLSA or MLST schemes, mainly because such properties are subjected to lateral transfer events and their mutation rate is higher, masking the phylogenetic background of the studied group. Nonetheless they describe fairly the adaptation and widespread of a particular population to an environment or to a host in the case of symbiotic and host-pathogen interactions.

The speciation, dispersion and adaptation processes of *Frankia* are not well understood. One can infer that allopatric events allow the widespread of the strain within the ecological frames. Two facts support this idea, 1) the ubiquitous properties of some group of *Frankia* strains (i. e. *Alnus* and *Elaeagnus*-compatible strains), and 2) the saprophytic lifestyle of some strains and their presence in ecosystems with no occurrence of actinorhizal plants (Benson & Dawson, 2007a).

One strain can follow allopatric speciation when it is forced to adapt in two divergent and geographically separate sites. The presence or absence of its host should play an important role in such adaptation; however its genomic plasticity and the expression of accessory functions will be essential to its survival. In such situation the functional genes become an important source of information that can report the evolutionary processes that are present in this place at spatial temporal scale. On the other hand they can fail to provide phylogeographical signals between the two different populations (Vitorino *et al.*, 2008).

Accessory genes have been mostly useful in epidemiology for clonality detection and strain diversity; the MLSVT (multi-virulence locus sequence typing) utilize some virulence

factors-coding genes instead of housekeeping genes. Such approach is useful in genotyping, discrimination between pathogenic clones and also to study the host-pathogen interaction (Zhang *et al.*, 2004).

In a parallel perspective for the symbiotic microorganisms, the genes involved in the symbiosis establishment can also give insights of the effects of host ecology and evolution on microbial partners. The analysis of the symbiosis related functions from an evolutionary ecology perspective might ultimately provide powerful new tools in the microsymbiont evolution (Laguerre *et al.*, 2001).

In the nitrogen fixing symbiosis the genes directly involved in nitrogen fixation process have been considered as promising targets in phylogenetic studies (Mierzwa *et al.*, 2010). Nitrogen fixation is carried out by the diazotrophic bacteria but it is mostly regulated by the host plant (Schumpp & Deakin, 2010), thus the microbial functions implicated are likely under selective pressure conducted by the host plant. In addition, clones from a same genomic group of diazotrophs exhibit differences in the nitrogen fixation efficiency under different environmental conditions (Benson & Silvester, 1993), implying an additional effect of environment in the shaping of diversification.

In this regard the *nifH* gene (coding for the nitrogenase reductase) is one of most widely used markers in diazotrophs phylogeny construction. The topology of the *Frankia* phylogeny based on *nifH* sequences is quite similar to that of 16S rRNA in terms of host compatibility groups (Jeong *et al.*, 1999; Welsh *et al.*, 2009a). To date comparative *nifH* sequences are utilized to establish the diversity of *Frankia* strains within the root nodules (Welsh *et al.*, 2009a) and even in the soil (Mirza *et al.*, 2009b) (the second niche of these actinomycetes). Although its phylogenetic relevance is clear, its discriminative power is low, thus it is not applicable to discriminate closer strains. In some diazotrophs the *nifH* is subject to lateral transfer events (Bolhuis *et al.*).

Genes coding for the glutamine synthase enzyme have been used as ecological markers. It is implied in the assimilation and incorporation of nitrogen fixation product (NH_3). *Frankia* owns two different glutamine synthases coded by the genes *glnA* and *glnII* (Hosted *et al.*, 1993). The former is not directly involved in nitrogen fixation and likely is not submitted to host selection (Clawson *et al.*, 2004a). Conversely, *glnII* is directly involved in symbiosis and yield a discriminatory power similar to 16S rRNA and is useful to place not characterized isolates within the major *Frankia* clades, however it fails in discriminating between closer isolates (Cournoyer & Lavire, 1999).

Nodulation genes (*nod*) are restricted to internalized symbionts and confined to symbiotic selective pressure becoming promising ecological evolutionary markers. The *nodA* phylogeny in Bradyrhizobia is consistent with the overall genetic variability of strains and their ability to perform efficient symbiosis with many different plants (Moulin *et al.*, 2004). Moreover, the nodulation-related genes of *Burkholderia*–specific clades, form a species complex, suggesting that these genes diverged over a long period within *Burkholderia* without substantial horizontal gene transfer between species complexes (Bontemps *et al.*, 2009).

Unfortunately, in *Frankia* the nodulation genes are still not identified and we lack of data about the establishment of the actinorhizal symbiosis, hence no potential targets can be defined to use it as molecular markers.

In an attempt for detecting *Frankia* functions potentially involved in the symbiosis initiation, Bagnarol et al. (2007) analyzed the reorganization of the *Frankia* proteome triggered by host plant chemical signals. The most striking response was the upregulation of the FeSOD, which was present as a major spot in all the obtained proteomic profiles. Conversely elements of the iron metabolism and transport, siderophore coding genes were down or upregulated in a strain dependent manner (Bagnarol *et al.*, 2007). Moreover in Rhizobiaceae, siderophore production and/or transport mutants produced inefficient nodules (Benson *et al.*, 2005), suggesting that siderophores determinants might not be involved in the infection step (Barton *et al.*, 1996) but they might be related to the effective establishment of symbiosis .

These results make the *sodF* and siderophore implicated genes, interesting ecological markers to be utilized as phylogenetic markers since an evolutionary ecology perspective might ultimately provide new insights in the *Frankia* diversification and speciation. Therefore in this last chapter of this study the occurrence of siderophore biosynthesis genes (PartI) and *sodF* (PatII) are studied in a large panel of *Frankia* strains. The results are divided in two subchapters.

The Siderophore study was started by Bagnarol, E. (2007) as part of her PhD Thesis dissertation. My personal contribution to this work comprised the ^{59}Fe uptake mediated by siderophore and xenosiderophores in *Frankia*. Moreover I participated in the research of siderophore determinants in some strains as well as in the phylogenetic analysis.

IV.1. (Article 4) Siderophore biosynthesis and xenosiderophore uptake in Frankia spp.

Running title

Frankia siderophores

Authors

Hector H. Bautista Guerrero[1], Emilie Bagnarol[1], Maria Antonia Cruz Hernandez[2], Petar Pujic[1], Pascale Fournier[1], Jean-Marie Meyer[3], Maria P. Fernandez[1]

Laboratory addresses

[1] Université de Lyon, F-69622, Lyon, France and Université Lyon1, Villeurbanne, CNRS, UMR5557, Ecologie Microbienne, 43, Boulevard du 11 novembre 1918, F-69622 Villeurbanne Cedex, France

[2] Centro de Biotecnologia Genomica, IPN, col. Narciso Mendoza CP 88710, Reynosa, Tam, Mexico.

[3] Laboratoire de Microbiologie et de Génétique, CNRS/Université Louis-Pasteur FRE 2326, 67000 Strasbourg, France

Corresponding author

Hector H. Bautista Guerrero

Ecologie Microbienne UMR5557, 43, Boulevard du 11 novembre 1918, F-69622 Villeurbanne Cedex, France

Tel: (33) 472 44 82 00

Fax: (33) 426 23 44 68

E-mail: hector.bautista@univ-lyon1.fr

Journal: International Microbiology Subbmision No. 10-068

Summary

Frankia siderophore biosynthesis and transport is likely involved at actinorhizal symbiosis initiation step. This work aimed at analyzing siderophore biosynthesis determinants in *Frankia* strains belonging to different infectivity and phylogenetic groups. Genomes mining on three sequenced strains and detection by PCR amplification on 39 genetically divergent strains were performed. Two kinds of siderophore, belonging to two structural groups, hydroxamates (HYD) and dihydroxybenzoates (DHB) may be synthesized through Non Ribosomal Peptide Synthases (NRPS)-independent synthesis or NRPS synthesis respectively. The DHB gene clusters, absent in strains nodulating *Casuarina,* display a heterogeneous organization with low synteny. Structure predictions were made by *in silico* analysis of active sites of Non Ribosomal Peptide Synthases (NRPS). The HYD siderophore gene cluster is present in all strains tested but two. High synteny was found among *Frankia* strains and with closer Actinobacteria as well as with the rhizobactin1021 synthesis cluster from *Sinorhizobium meliloti*. Phylogenetic trials of the HYD genes clustered this α-proteobacteria with Actinobacteria suggesting lateral transfer events. Functional and therefore structural similarity between rhizobactin1021 and the siderophore synthesized by *Frankia* strain ACN14a was demonstrated by siderophore-mediated $^{59}Fe^{3+}$ uptake assays. We concluded that siderophore production is widespread in *Frankia* and that nitrogen fixation and iron uptake mecanisms are linked as suggested by the inhibition of siderophore production in nitrogen fixing conditions and the positionning of *dtxR* regulator in an operon involved in nitrogen fixation.

Introduction

Iron is essential for growth of microorganisms since it is required for many central physiological processes such as electron transport, energy production, photosynthesis and nitrogen fixation (Guerinot & Yi, 1994). To overcome the limiting availability of iron at biological pH values, and under aerobic conditions, organisms can synthesize powerful acquisition systems, known as siderophores, and express specific receptor proteins to transport them into the cell (Neilands, 1981a; Neilands, 1981b; Neilands, 1982). Siderophores are low molecular weight Fe(III)-specific binding molecules, secreted under iron-limiting

conditions, by microorganisms and by plants (phytosiderophores). Bacterial siderophores exhibit a wide range of chemical structures (Winkelmann, 1991), and are generally divided into 3 main groups based on the nature of their iron-binding ligand: hydroxamate, catecholates and a mixed functional group. Many of them are polypeptides involving non ribosomal peptide synthases (NRPS) for their synthesis, while several others do not possess a peptide nature, such as those containing hydroxamate-chelating groups (Challis, 2005).

Nitrogen fixing microorganisms should have high requirements for iron which is needed, as component of the nitrogenase complex and for providing reductants for nitrogen fixation through the electron transport chains. Moreover, leghemoglobin, an iron-containing protein necessary for protecting the oxygen-labile nitrogenase, can represent 25-30% of the total cell proteins (Appleby, 1984) and iron-containing cytochromes are produced three times as much in symbiosis as under free living conditions (Sangwan & O'Brian M, 1992). Siderophore production may not be essential for the first steps of the infection process (Barton *et al.*, 1996; Stevens *et al.*, 1999) but contributes to improving the efficiency of nitrogen fixation. In *Rhizobiaceae*, mutants unable to synthesize or transport siderophores, still nodulate but the nodules produced do not contain leghemoglobin and do not fix nitrogen (Benson *et al.*, 2005; Gill & Neilands, 1989). As a consequence, siderophore production has been detected in most nitrogen fixing symbionts : cyanobacteria (Boyer *et al.*, 1987; Simpson & Neilands, 1976), *Azospirillum* (Bachhawat & Ghosh, 1987), *Rhizobium* (Barton *et al.*, 1994; Guerinot, 1991) and *Frankia* (Arahou *et al.*, 1998; Aronson & Boyer, 1992; Aronson & Boyer, 1994; Boyer & Aronson, 1994; Boyer *et al.*, 1999).

Frankia is an actinobacterium which fixes nitrogen both in aerobic cultures and in symbiosis with 25 genera of angiosperm known collectively as actinorhizal plants. Actinobacteria and particularly *Streptomyces* are known to produce a wide variety of secondary metabolites, among which different kinds of siderophores. Few studies reported the growth of *Frankia* under iron limited conditions. Two new hydroxamate type siderophores have been identified: the frankobactin and the frankobactinA produced by the *Frankia* sp. strains 52065 and CeS15 isolated from *Ceanothus* and *Casuarina* respectively. Frankobactin is a water soluble dihydroxamate siderophore containing a phenyl oxazolin ring or an open oxazoline ring (frankobactin A), 4 amino acids and 2 hydroxamate groups (Boyer & Aronson, 1994; Boyer *et al.*, 1999). The involvement of siderophore during nitrogen fixation and nodule development in actinorhizal symbiosis is unknown. Induction of *Frankia* strains with

host plant phenolic extracts, revealed differential expression of proteins involved in siderophore biosynthesis and transport (Bagnarol *et al.*, 2007). Among the differentially

expressed proteins, an ABC type Fe(III) siderophore transport system, a putative isochorismatase family protein involved in enterochelin (an iron chelator in *E.coli*) biosynthesis and a putative enterochelin esterase which catalyzes ferric-enterochelin hydrolysis (Langman *et al.*, 1972) were detected as down or up-regulated according to the strain. This suggests that *Frankia* siderophores are regulated by plant interaction and that siderophore requirement and production would be strain specific.

Frankia genus presents high genetic and physiologic diversity. Strains can be divided into 3 main genetic clusters based on 16S rRNA sequence analysis, each cluster corresponding to a different infectivity group (distinct host range) (Benson *et al.*, 2004; Normand *et al.*, 1996): cluster 1 divided in subclade 1a and 1b for "*Alnus* and *Myrica*–infective strains", and "*Casuarina*-infective strains" respectively, cluster 2 for unisolated *Rosaceae* and *Coriariaceae*-infective strains and cluster 3 for "*Elaeagnaceae*-infective strains". Three genomes representative of these three main clusters were sequenced (Normand *et al.*, 2007a). Few strains were tested for siderophore production and they mainly originated from the two (out of 25) genera *Casuarina* (cluster 1b) and *Hyppophae* (cluster 3). The slow growth rate of *Frankia* and the lack of mutagenesis and transformation techniques for this genus have restricted progress in this field of research.

This paper aims at evaluating the distribution and diversity of siderophore genetic determinants among a large panel of *Frankia* collection strains belonging to different infectivity groups and their possible expression in relation with nitrogen fixation. We report here the organization of siderophore biosynthesis genes in the three sequenced strains. From these data, siderophore structure predictions were made. A PCR-based protocol was used to detect the occurrence of those genes and their phylogeny in a large panel of strains. Bioassays for siderophore production and use were conducted using the enzymatic CAS protocol, which detects any iron-binding uncharacterized molecules (Schwyn & Neilands, 1987), and incorporation studies were performed in nitrogen fixing and non-fixing conditions using [^{59}Fe]-labelled siderophores.

IV.1.1. MATERIAL and METHODS

Strains studied and growth conditions

The three sequenced strains, ACN14a, CcI3 and EaN1pec (Normand *et al.*, 2007c), isolated from *Alnus, Casuarina* and *Elaeagnus* and belonging to cluster1a, cluster1b and

Table 1 *Frankia* strains tested and PCR amplification results for hydroxamate and dihydroxy benzoate siderophores

Strain	**Plant Source**	Reference	**DHB pb[#]**	**HYD pb[#]**	**CAS test**
Cluster 1a strains[†]					
ACN14a	*Alnus viridis ssp. crispa*	(Normand & Lalonde, 1986)	750	650	+
CpI1	*Comptonia peregrina*	(Callaham *et al.*, 1978)	-[1]	**650**[2]	+
ARgN22d	*Alnus rugosa*	(Normand & Lalonde, 1982)	-	650	
ArI3	*Alnus rubra*	(Normand & Lalonde, 1982)	-	650	-
Ar24H3	*Alnus rubra*	(Simonet *et al.*, 1989)	-	650	
Ar24H5	*Alnus rubra*	(Simonet *et al.*, 1989)	-	**650**	
$Ar240_2$	*Alnus rubra*	(Simonet *et al.*, 1989)	-	650	-
AirI1	*Alnus incana ssp. rugosa*	(Lechevalier, 1986)	-	650	
$Ag24_{251}$	*Alnus glutinosa*	(Fernandez *et al.*, 1989a)	-	650	
AvcI1	*Alnus viridis ssp. crispa*	(Baker & Torrey, 1980)	**700**	650	
AVN17o	*Alnus viridis*	(Fernandez *et al.*, 1989a)	-	650	
$Ac23_{40}$	*Alnus crispa*	(Fernandez *et al.*, 1989a)	-	650	
$ARgP5^{\Delta G}$	*Alnus rugosa*	(Normand & Lalonde, 1986)	-	650	-
Ag21D1	*Alnus glutinosa*	Unpublished*	620	-	
I38	*Alnus incana*	Unpublished	-	650	
M16467	*Myrica pensilvanica*	(Clawson & Benson, 1999)	**650**	**650**	+
Cluster 1b strains[†]					
CcI3	*Casuarina cunninghamiana*	(Zhang *et al.*, 1984)	-	650	-
TA	*Allocasuarina torulosa*	(Zhang *et al.*, 1984)	-	650	

BR	*Casuarina equisetifolia*	(Puppo *et al.*, 1985)	-	**650**	-
Cj1-82	*Casuarina junghuniana*	(Diem *et al.*, 1983)	-	650	-
$Cg70_3$	*Casuarina glauca*	Unpublished	-	650	
$Cg70_4$	*Casuarina glauca*	Unpublished	-	650	-
$Cg70_5$	*Casuarina glauca*	Unpublished	-	**650**	
$Cg70_9$	*Casuarina glauca*	Unpublished	-	650	-
CcI2	*Casuarina cunninghamiana*	(Zhang *et al.*, 1984)	-	650	
Cluster 3 strains[†]					
CeSI5 (atypical)[‡]	*Casuarina equisetifolia*	(Burggraaf & Shipton, 1983)	-	-	
G_2 (atypical)[‡]	*Casuarina equisetifolia*	(Diem *et al.*, 1982)	-	570	+
$Ea1_{12}$	*Elaeagnus angustifolia*	(Fernandez *et al.*, 1989a)	620	**570**	+
$Ea1_2$	*Elaeagnus angustifolia*	(Fernandez *et al.*, 1989a)	600	**580**	
$EaCm5_1$	*Elaeagnus angustifolia*	(Fernandez *et al.*, 1989a)	600	570	
$Ea48_2$	*Elaeagnus angustifolia*	(Jamann *et al.*, 1992a)	620	580	
$Ea48_8$	*Elaeagnus angustifolia*	(Jamann *et al.*, 1992a)	620	580	
$Ea48_4$	*Elaeagnus angustifolia*	(Jamann *et al.*, 1992a)	-	580	
$Ea8_4$	*Elaeagnus angustifolia*	(Jamann *et al.*, 1992a)	**530**	570	
$Ea7_4$	*Elaeagnus angustifolia*	(Jamann *et al.*, 1992a)	**600**	580	
$Ea35_4$	*Elaeagnus angustifolia*	(Jamann *et al.*, 1992a)	400	580	
$Ea35_2$	*Elaeagnus angustifolia*	(Jamann *et al.*, 1992a)	450	580	
$Ea7_1$	*Elaeagnus angustifolia*	(Jamann *et al.*, 1992a)	600	580	
$Ea7_3$	*Elaeagnus angustifolia*	(Jamann *et al.*, 1992a)	600	580	

[†] As in Normand et al 1996.

‡ Atypical strains are not able to re-nodulate their original host but nodulate Elaeagnaceae. * * All unpublished strains came from A. Moiroud collection (Lyon laboratory).

[#] HYD: hydroxamate primers, DHB: dihydroxy benzoate primers.

[I] -, No amplification.

[a] Numbers in bold underlined are the fragments which were sequenced and employed in the phylogenetic analysis.

cluster 3 respectively, were used to determine candidate genes and specific probes for PCR amplifications. A panel of 39 *Frankia* collection strains, among which 16 belong to cluster 1a, 9 to cluster 1b and 14 to cluster 3 were studied. Because of their non cultivability, cluster 2 strains were not included (Table1). Cultures were performed in BAP medium (Murry *et al.*, 1984a) modified to contain sodium propionate as carbon source and ammonium chloride as nitrogen source with (+Fe) or without (-Fe) iron.

Genome analysis

Genome sequences from ACN14a, EaN1pec and CcI3 strains were compared and Blast searches were performed using the Genoscope MAGE platform: (http://www.genoscope.cns.fr/agc/mage/wwwpkgdb/MageHome/index.php?webpage=mage). Siderophore structure predictions were made by analysis of active sites of NRPS-domains using "Domain Search Program for NRPS and PKS" software (http://linux1.nii.res.in/~zeeshan/searchnrps.html).

PCR amplification

DNA extractions were performed according to Simonet et al., (1986) or using the QIAquick Gel Extraction Kit (QIAGEN, Hilden, Germany). Two kinds of siderophore gene clusters were detected in *Frankia* strains: dihydroxybenzoate (DHB), and hydroxamate (HYD). PCR amplifications targeting a siderophore biosynthesis gene homolog from each cluster (DHB and HYD) were undertaken on the 39 strains. Primers DHB1 (5'-TGCGCGACGAGCTSGACAC-3') with DHB2 (5'-AGGTGCTCGATCTCGACCCT-3') were used to amplify part of a putative NRPS (FRAAL 4160, Franean1_5942) and of a 2,3-dihydroxybenzoate AMP ligase (FRAAL 4159, Franean1_5941). Primers HYD1 (5'-CGYACYGTCTGGCGGCCGGA-3') with HYD2 (5'-GTAGCCCTGGCTGTCGCGGTA-3') were used to amplify a putative siderophore biosynthesis gene (FRAAL 6423, Francci3_4056, Franean1_0660), homologous to *iucA* from *Escherichia coli* and *rhb*C from *Sinorhizobium meliloti* respectively (Fig. 1). Amplicon sizes of 751bp and 592bp for ACN14

and EaN1pec, respectively, were expected with DHB primers, and 657, 653 and 555bp for ACN14a, CcI3 and EaN1pec, respectively, were expected with HYD primers. DNA amplification reactions were carried out in a final volume of 25µl using 1 x PCR buffer, 1.5mM $MgCl_2$, 0.5mM of each primer, 10µM of each dNTP, 0.5 U Taq DNA polymerase. PCR cycles were as follow: (1) 95°C for 5min, (2) 95°C for 1min, 64.5°C for 45 sec, 72°C for 2 min (35 cycles), and (3) 72°C for 5min.

Sequencing and phylogenetic analysis

Direct sequencing of PCR-amplified DNA was performed (Genome express, Grenoble, France) on 11 *Frankia* strains chosen to provide a representative range of amplicon length and to belong to different infectivity groups (table 1). Seven PCR fragments (from 2 *Casuarina* strains, 3 *Alnus* strains and 2 *Elaeagnaceae* strains) obtained with HYD primers, and four fragments (from 2 *Alnus* strains and 2 *Elaeagnus* strains) obtained with DHB primers were sequenced. Searches for similarity to known DNA and protein sequences in databases were performed using Blast research software available on the NCBI website (http://www.ncbi.nlm.nih.gov). Phylogenetic analyses were performed on the sequenced PCR-amplicons and on the complete sequences identified in the 3 sequenced genomes of *Frankia* and in other non-Frankiaceae. Derived protein sequences were aligned using the multiple alignment clustalX (Higgins & Sharp, 1988). Phylogenetic trees were inferred using the neighbour joining (NJ) algorithm (Saitou & Nei, 1987).

CAS assay

Thirteen *Frankia* strains were tested for siderophore production: ACN14a, Cj1-82, ORS020608, G2, $Cg70_4$, $Cg70_9$, CcI3, Ea112, M16467, CH37, CpI1, Ar24O2, $ARgP_5$. Precultures were performed in BAP medium without iron. Iron was removed from the vessels using HCl and from the medium using 8-hydroxyquinoleine 5% w/vol in chloroform, followed by 3 chloroform extraction steps to remove 8-hydroxyquinoleine (Nicholas, 1952). After 6 days at 28°C under stirring conditions, precultures were used to inoculate BAP media containing 0, 0.5 or 20µM of $FeCl_3$. Two flasks of 200ml were used for each condition.

Ten days cultures were centrifuged and their supernatants were examined for siderophores production using the chrome azurol sulfonate (CAS) assay (Schwyn & Neilands, 1987). A volume of 0.5 ml of blue CAS solution was added to 0.5 ml of culture supernatant. Reactions were considered positive after monitoring changes of absorbance of the assay

reagent (from blue to orange) at 630nm. Absorbance differences (ΔOD) between the reference solution (uninoculated medium) and the culture supernatants were calculated.

Siderophore-mediated iron uptake

Iron uptake studies were performed using radiolabelled ferric siderophores. AcN14a strain was cultured in either iron and nitrogen free (-Fe-N), iron-sufficient nitrogen-sufficient (+Fe+N) or iron-free nitrogen-sufficient (-Fe+N) BAP medium. Twenty five days-old cells were harvested by centrifugation, washed once with distilled water, homogenized by repeated forced passages through a 21-g needle with a syringe and resuspended in (-Fe-N) BAP medium (incubation medium) at a 600 nm optical density of 0.33.

$^{59}Fe^{3+}$ uptake was tested in homologous conditions using *Frankia* supernatant as siderophore source and in heterologous conditions using purified rhizobactin1021 or other purified siderophores: desferral, enterobactin, rhizoferrin and different pyoverdines obtained from fluorescent *Pseudomonas* species: KT, 2440, PL9, W, 90.51, pf1 and ATCC 17400.

Label mix consisted of 5 µl of the commercial ^{59}Fe solution ($FeCl_3$ in 0.1 M HCl, specific activity 110 to 925 MBq/mg; Amersham) in 100 µl of water and 10 µl of a 6.5-mg/ml XAD-purified rhizobactin solution incubated 30 min at room temperature and then adjusted to 1 ml with incubation medium.

For uptake studies, 0.5 ml of label mix was added to 4.5 ml of bacterial suspension and incubated with gentle shaking at 30°C or 0°C. After 20, 60, 90 or 120 min, 1 ml sample was filtered through a 0.45µm nitrocellulose filter. Each filter was washed twice with 2 ml of incubation medium and wrapped in an aluminum foil for cell-associated radioactivity counting in a Gamma 4000 counter (Beckman). Medium radioactivity counting in the remaining 1 ml of filtrate allowed to determine the total amount of radioactivity present in the assay. A control assay was conducted with 4.5ml of uninoculated incubation medium to ensure that the siderophore content was high enough to reach a full complexation of the labeled iron.

$^{59}Fe^{3+}$ uptake due to indigenous *Frankia* siderophores was determined in the same way replacing purified siderophores of the label mix by 100 µl of unconcentrated culture supernatant.

IV.1.2. RESULTS

Nucleotide sequences and organization of HYD and DHB clusters

Analysis of the three sequenced genomes (ACN14a, EaN1pec and CcI3) revealed several gene clusters putatively involved in siderophore biosynthesis and transport. According

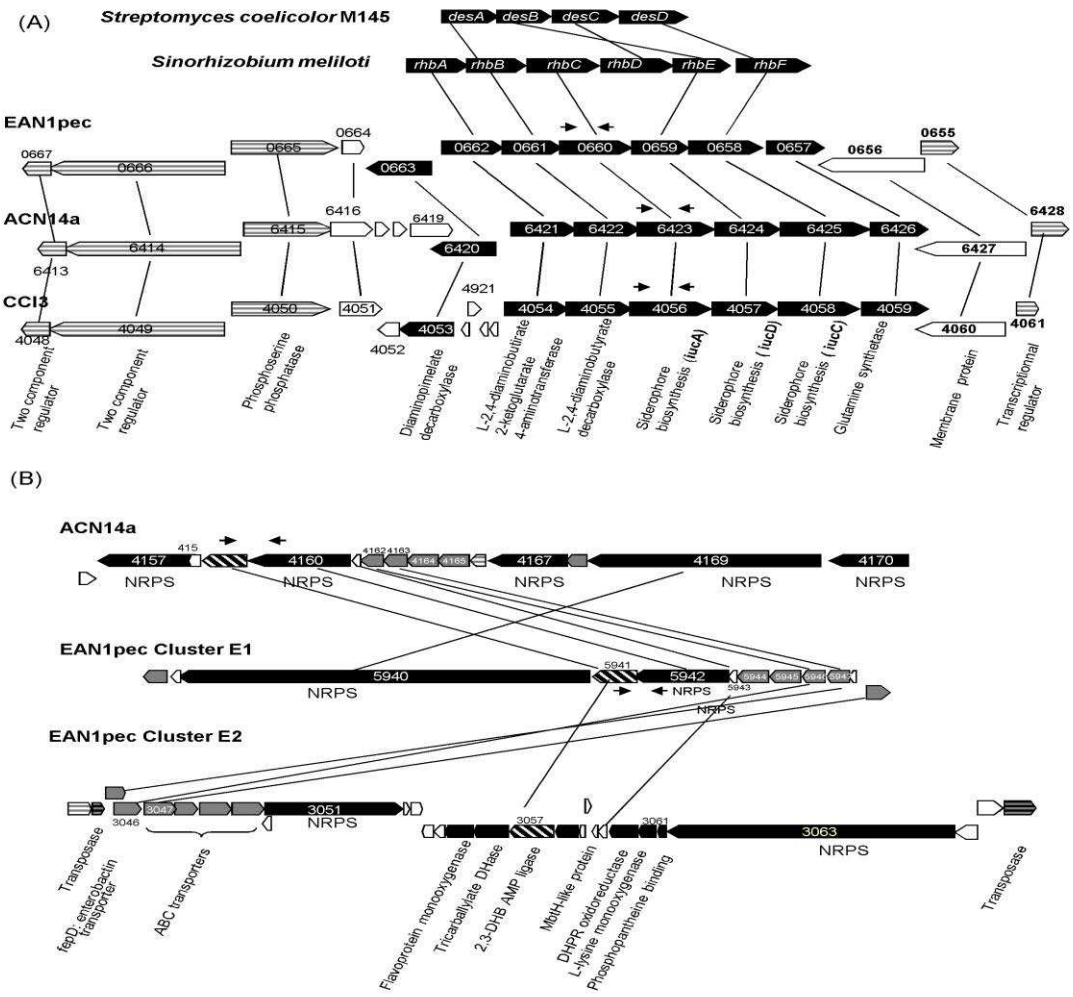

Figure 1 Genetic organization of the different biosynthesis gene clusters in *Frankia*. Arrows indicates primers position. (A)Hydroxamate cluster of *Frankia* compared with desferrioxamin (from *Streptomyces coelicolor* (Barona-Gomez *et al.*, 2004) and rhizobactin1021 (*Sinorhizobium meliloti*) biosynthesis genes, (B) Dihydroxybenzoate clusters. Grey: NRPS,

bold striped: 2,3-dihydroxybenzoate AMP ligase, black: structural genes, grey: siderophore uptake and siderophore ABC transporters, white: hypothetical proteins, horizontal grey striped: regulators and associated genes

to the strain, clusters involved in hydroxamate (HYD) and/or dihydroxybenzoate (DHB) (belonging to the catecholate family) siderophores were detected.

Gene clusters involved in HYD siderophores biosynthesis (Fig. 1A) were detected in the three sequenced strains, with a high synteny. For each strain, the cluster is composed of (i) a two component system which may be involved in regulation, (ii) genes homologous to rhizobactin1021 (a *Sinorhizobium meliloti* siderophore) or aerobactin (from *E. coli*) biosynthesis genes. The products of the 5 genes, FRAAL 6421, 6422, 6423, 6424 et 6425 (homologous to Francci3_ 4054, 4055, 4056, 4057, 4058, and Franean1_ 0662, 0661, 0660, 0659, 0658) are homologous to products of the 5 genes *rhbABCEF* involved in rhizobactin1021 biosynthesis, and the last 3 genes are homologous to *iucADC* genes involved in aerobactin biosynthesis. Additional genes coding for glutamine synthetase (FRAAL6426, Franean1_0657, Francci3_4059) and diaminopimelate decarboxymase (FRAAL6420, Franean1_0663, Francci3_4053), absent in the rhizobactin1021 operon, are present in the 3 strains in a divergon.

Gene clusters involved in DHB siderophores are strain specific and not syntenic (Figure 1B). No DHB siderophore cluster was detected in CcI3 strain as compared with one in ACN14a (between FRAAL4154 and 4170) and two in EaN1pec strain [between Franean1_5938 and Franean1_5948 (clusterE1) and between Franean1_3044 and Franean1_3064 (clusterE2)]. Each cluster is composed of a 2, 3-dihydroxybenzoate AMP ligase gene, and of the other genetic determinants necessary for the biosynthesis (NRPS, phenyloxazolin synthetase), and transport of siderophores with at least 4 genes homologous to putative ABC siderophores uptake and transport systems. The 2,3-dihydroxybenzoate AMP ligase sequence analysis revealed one adenylation domain in ACN14a cluster for activation of one dihydroxybenzoate moiety whereas 2 adenylation domains were detected in the 2 EaN1pec clusters, for activation and incorporation of 2 similar substrates (2,3 DHB or homolog) in the final product. Furthermore, the ACN14a DHB cluster comprises four NRPS (FRAAL 4160, 4167, 4169 and 4170) whereas EaN1pec clusterE1 and EaN1pec clusterE2 contain 2 NRPS-coding genes (Franean1_ 5940 and 5942) and (Franean1_ 3051 and 3063) respectively.

PCR amplifications

Hydroxamate and dihydroxybenzoate biosynthesis genes distribution was studied on the 39 *Frankia* strains by PCR amplification using the two sets of primers, HYD and DHB. PCR results are summarized in Table1. All strains tested but one, the frankobactin producer CeSI5, yielded positive results with either DHB or HYD or with both primers.

PCR amplifications with HYD primers yielded positive responses for 37 strains among the 39 tested. The phylogenetically divergent Ag21D1 strain (unpublished data) and CeSI5 yielded no HYD amplification. The PCR amplification generated a unique amplicon of approximately 650bp for strains from clusters 1a and 1b and 580bp for strains from cluster 3. These fragment sizes were similar to those calculated from the genomes sequence analysis: 657bp for ACN14a, 653 for CcI3 and 555 for EaN1pec.

No amplification was obtained for *Casuarinaceae* strains (cluster1b) with DHB primers. Among the 16 strains tested belonging to cluster1a, amplification was obtained only for 4 strains: ACN14a, AVCI1, Ag21D1, and M16467 with fragment size varying from 620 (Ag21D1) to 750pb (ACN14a). Eleven *Elaeagnaceae* strains among the 13 tested yielded an amplicon with DHB primers. Eight strains gave an amplicon of about 600pb whereas 3 strains ($Ea35_4$, $Ea35_2$ and $Ea8_4$) gave a shorter amplicon (between 530 and 400pb) and two strains ($Ea48_4$ and G2) did not yield amplification. G2 is considered an atypical strain since it is isolated from *Casuarina equisetifolia* but can only nodulate *Elaeagnaceae* strains. It harbored the *Elaeagnaceae* HYD pattern (a fragment of 570bp) but was similar to *Casuarina* strains by the absence of DHB cluster.

Phylogenetic analysis

Two sets of phylogenetic trees were constructed for HYD and DHB sequences based on partial amino acid sequences derived from PCR-amplified fragments (data not shown) and total sequences obtained from the three sequenced genomes respectively (Fig. 2).The topologies yielded by the two sets of sequences were conserved, better statistical support was though obtained when the complete sequences were used, therefore they were retained for further analysis below.

With both complete and partial HYD sequences, *Frankia* strains are grouped into a unique clade supported by high (between 99 and 100%) bootstrap values. Within this clade, the strains are grouped according to their respective infectivity groups (*Alnus-Myrica*, *Casuarina* and *Elaeagnaceae* strains) or 16S clustering (cluster 1a, 1b and 3), whatever the

variability of the amplicon size (partial sequences, data not shown). Neither Gram positive, nor Gram negative bacteria form consistently supported groups. Similarly the major branches, as defined by the 16S rRNA-based phylogeny (α, β, γ- Proteobacteria and Firmicutes), are not

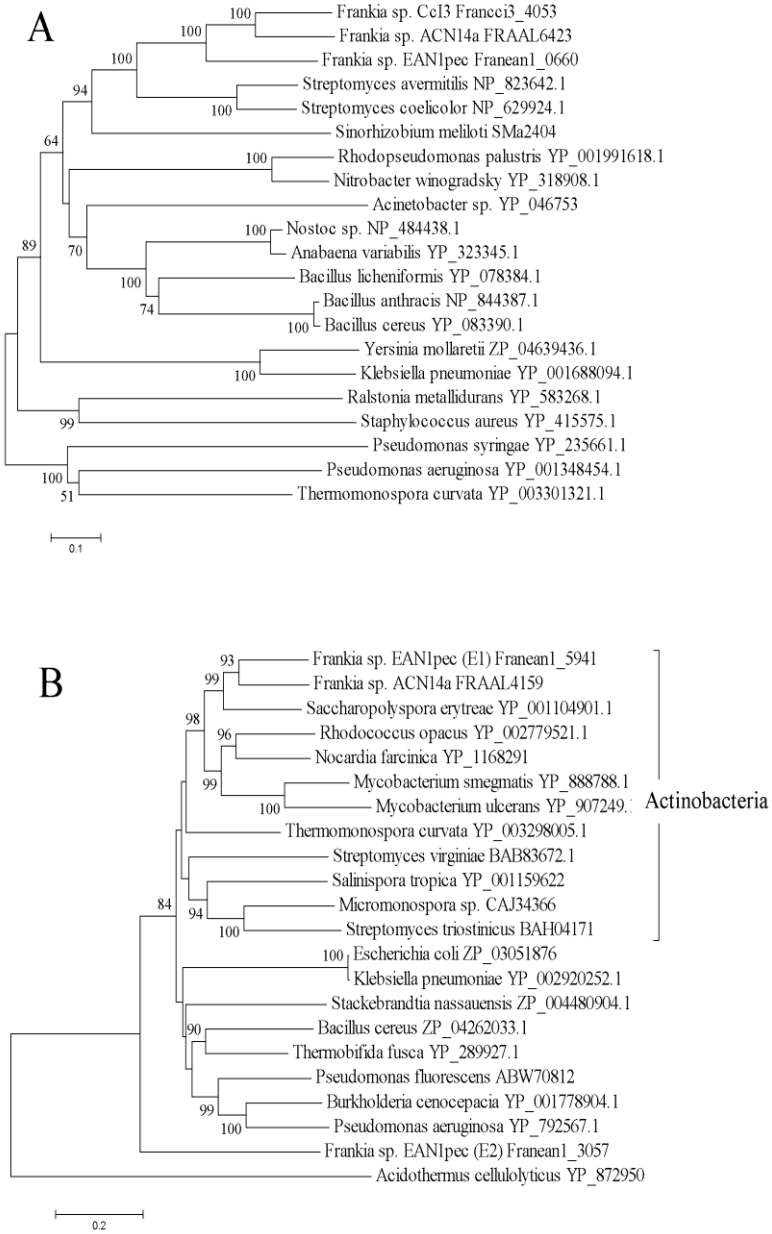

Figure 2 Phylogenetic tree generated by neighbour joining using the complete amino acid sequences of (A) *rhbC* homologues (HYD primers) and (B) 2,3-dihydroxybenzoate AMP ligase (DHB primers) . Bootstrap values were calculated from 1000 replicates, only values

above 50% are represented. Taxa sequences are identified by their NCBI accession number with the exception of *Frankia* sequences corresponding to MAGE platform nomenclature as in the Figure 1.

supported. Remarkably, the α-proteobacterium *Sinorhizobium meliloti* is positioned at the base of the actinobacteria branch with 94% bootstrap support.

The occurrence of DHB (2,3-dihydroxybenzoate AMP ligase) gene has been found in numerous actinobacterial genera including most of the closest *Frankia* neighbours, as determined by phylogenetic analysis of genome sequences (Wu *et al.*, 2009), like *Salinispora* sp., *Stackebrandtia* sp., *Thermobifida* sp., *Micromonospora* sp. and *Thermomonospora* sp. Remarkably DHB gene was absent in *Geodermatophilus*. The divergent DHB gene found in *Acidothermus cellulolyticus*, the closest *Frankia* neighbour in 16S rRNA phylogeny, likely not involved in siderophore synthesis, was used as an outgroup. The tree topology did not reflect 16S rRNA phylogeny indicating the probability of horizontal gene transfers.

According to the partial and the complete gene sequences, *Frankia* DHB genes from ACN14a and EaN1pec (cluster E1) are closely grouped and form, with other Actinobacteria, a significantly supported clade (98% bootstrap value), whereas DHB from EaN1pec (cluster E2) diverged markedly.

CAS ASSAY

The CAS assay was used to test the ability of 13 representative *Frankia* strains to produce siderophores under iron-limiting conditions. The detergent HDTMA contained in the CAS reagent appeared toxic for *Frankia* cells, as for many other bacteria (Milagres *et al.*, 1999) thus the test was applied on culture supernatants instead of agar plate cultures.

On the basis of the results summarized in figure 3, two groups of strains were found CAS-positive (CAS+) and CAS-negative (CAS-). CAS assay is usually considered positive when ΔOD is larger than 0.1 as was found for only three strains (Ea112, CpI1 and M16467). However, two other strains (ACN14 and G2), with ΔOD of respectively 0.08 and 0.09, were considered positive since they presented significant OD differences between 0µM and 20µM Fe treatments. These five positive strains belong to *Frankia* clusters 1a and 3. All the tested *Casuarina*–infective strains, showed a negative-CAS response at the three iron concentrations used, with Cj1-82 giving an intermediate result. The optimal siderophore production occurred, according to the strains, at 0 µM (Ea112, CpI1 and M16467) or at 0.5µM Fe (ACN14 and G2) in culture medium. No siderophore production was detected at 20µM of iron whatever the strain.

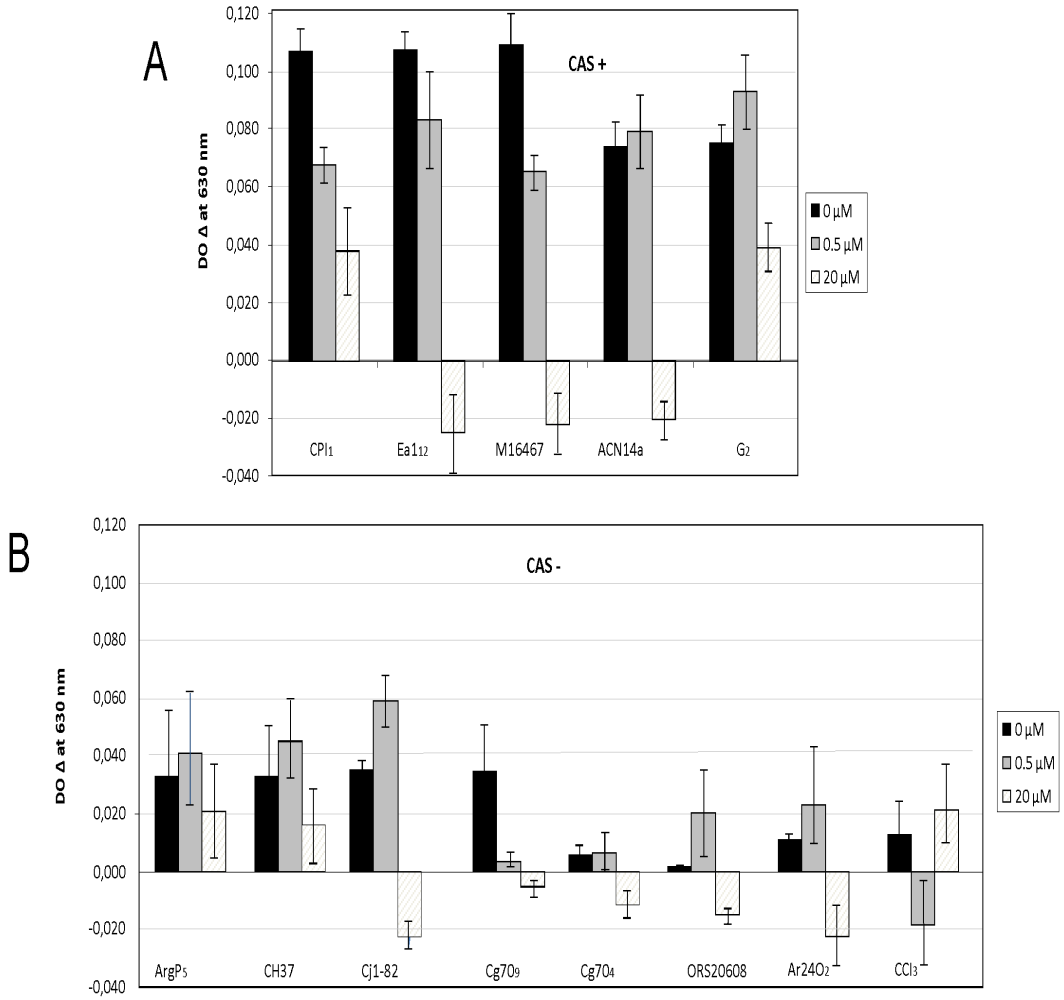

Figure 3 Iron-binding activity of different *Frankia* unconcentrated supernatants as assessed by the CAS test at three iron concentrations (0, 0.5 and 20 □M) in BAP medium. A: CAS-positive strains, positive responses correspond to a □OD at 630nm around or higher 0,1 and /or with large differences between 0mM and 20mM assay. B, CAS-negative strains

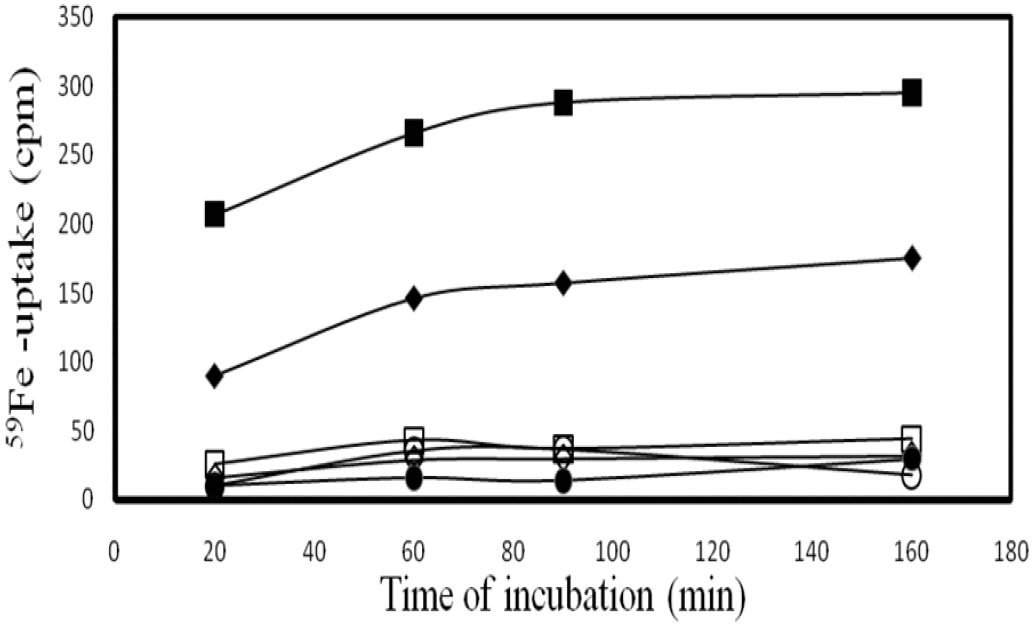

Figure 4 (^{59}Fe)-siderophore mediated uptake by AcN14a 25 days-old in presence or absence of iron and nitrogen in the medium and using either purified rhizobactin1021 or *Frankia* culture supernanants obtained from ACN14a iron depleted cultures as siderophore source. Black squares, supernatant on –Fe+N cultured cells, open squares, supernatant on +Fe+N cultured cells, black diamond, rhizobactin1021 on –Fe+N cultured cells, open diamond, rhizobactin1021 on +Fe+N cultured cells, black circles, rhizobactin1021 on -Fe-N cells, open circles, supernatant -Fe-N vs cultured cells

Siderophore-mediated iron uptake

$^{59}Fe^{3+}$ uptake was tested in homologous (*Frankia* supernatant) and heterologous (purified siderophore-mediated uptake) conditions using the CAS+ strain ACN14a. Only *Frankia* supernatant and rhizobactin1021 assays gave positive results, none of the other purified siderophore tested mediated $^{59}Fe^{3+}$ incorporation (data not shown). Incubation times lower than 20 minutes did not lead to $^{59}Fe^{3+}$ uptake (data not shown). The kinetic of ^{59}Fe-siderophore uptake are displayed in Fig. 4. Homologous and heterologous (rhizobactin1021) assays gave similar results with (i) linear $^{59}Fe^{3+}$ incorporation in iron free (-Fe+N) medium up to 90 min (ii) no $^{59}Fe^{3+}$ uptake in iron and nitrogen sufficient (+Fe+N) medium (iii) no $^{59}Fe^{3+}$ uptake in (-Fe-N) medium (iv) no $^{59}Fe^{3+}$ uptake at 0°C (data not shown), temperature at which the membrane transport systems dependent on energy are inactive (v)$^{59}Fe^{3+}$ uptake was higher with 25 days old cells compared with 31 days old cells. Iron incorporation was twice higher with *Frankia* supernatant compared with the rhizobactin1021 mediated assay.

IV.1.3. DISCUSSION

This work was aimed at analyzing siderophore biosynthesis determinants and their distribution among *Frankia* strains belonging to different infectivity groups, and to test at what extent N-fixing and non fixing conditions affected their iron-chelating activity. The functionality of the siderophore produced by the ACN14a strain was analysed by $^{59}Fe^{3+}$ uptake.

We concluded from these studies that 2 kinds of siderophores, belonging to two structural groups, hydroxamates (NRPS-independent synthesis) and dihydroxybenzoates (NRPS synthesis), could be synthesized by *Frankia*. All strains tested but one have the genetic complement to synthesize at least one kind of siderophore (Table1) and most of them could synthesize siderophores from different structural groups like in *Streptomyces* sp. (Bentley *et al.*, 2002; Challis & Ravel, 2000; Fiedler *et al.*, 2001) or in rhizobia (De *et al.*, 2003; Khandelwal *et al.*, 2002; Lynch *et al.*, 2001).

Analysis of the three sequenced genome revealed that only one (HYD) cluster is present in CcI3 as compared with 2 (HYD and DHB) in ACN14a and 3 (1 HYD and 2 different DHB) in EaN1pec (Fig.1B). Accordingly, the genetic survey among 39 reference strains confirmed the large occurrence of the HYD amplified gene within *Frankia* genus and the absence of DHB gene in *Casuarina*-nodulating strains suggesting that this kind of siderophore could not be synthesized by cluster 1b strains.

About 39% of *Frankia* strains tested had genes possibly involved in the synthesis of DHB siderophores (Table1). Phylogenetic analysis of the complete 2, 3-DHB AMP ligase gene sequences closely grouped ACN14a and EaN1pecE1 genes (Fig. 2B). We can hypothesize that a first lateral transfer may have occurred before the emergence of the 3 main *Frankia* lineages. According to this evolutionary scenario, the CcI3 DHB cluster may have been lost and the second EaN1pec cluster (E2), flanked by 2 transposases, may have been acquired recently (Fig. 1B).

Despite their different organization, the three DHB-type Frankia clusters are all composed of ABC siderophore transport and uptake systems, 2, 3-dihydroxybenzoate AMP ligase, different NRPS and other specific genes. The 2, 3-dihydroxybenzoate AMP ligase (entE homolog) is a key enzyme for DHB siderophore biosynthesis which catalyzes, via an ATP-dependent PPi exchange reaction, formation of (2, 3-dihydroxybenzoyl) adenylate from 2, 3-dihydroxy-benzoate (2, 3-DHB). 2, 3 DHB also acts as a virulence factor in Nocardia asteroides (Feistner & Beaman, 1987). Two other 2, 3 DHB AMP ligases were detected in the ACN14a and EaN1pec genomes, clustering with genes involved in tryptophan biosynthesis, thus probably not involved in siderophore metabolism. Many DHB siderophores are synthesized by members of the NRPS multienzyme family (Crosa & Walsh, 2002) which is also responsible for the biosynthesis of the majority of microbial peptide antibiotics. NRPS adenylation domains, which are the central point of action of NRPS, activate and incorporate amino acids in a final peptidic product. Their special organization defines the order of the incorporated residues in the final product (Marahiel *et al.*, 1997). Consequently, siderophore structure prediction can be made from 2, 3 dihydroxybenzoate AMP ligase and NRPS analysis (Fig. 5). According to the colinearity rules of NRPS, ACN14a cluster probably yields a peptidic product aggregating successively a dihydroxybenzoate, a cyclized cysteine, an unknown amino acid X1, a threonine, two other unknown amino acids X2 and X3 (similar to X1), a putative serine and a putative asparagine. EaN1pec cluster E1 permits the assembly of a product comprising a glycine, a threonine, a putative second glycine, a serine, a tryptophan, 2 dihydroxybenzoates and a cyclized cysteine. For the second EaN1pec cluster (E2), a serine, 2 glycines, 2 dihydroxybenzoates, and 3 serines are successively incorporated. The cyclization of cystein produces a thiazoline ring incorporated in ACN14a and E1 siderophore whereas in E2 cluster, a phenyloxazoline ring is probably obtained from the serine cyclization catalyzed by the phenyoxazoline synthase (Franean1_3062). Frankobactin also contains a phenyl oxazoline group linked with 2 hydroxamate moieties, justifying its classification in the

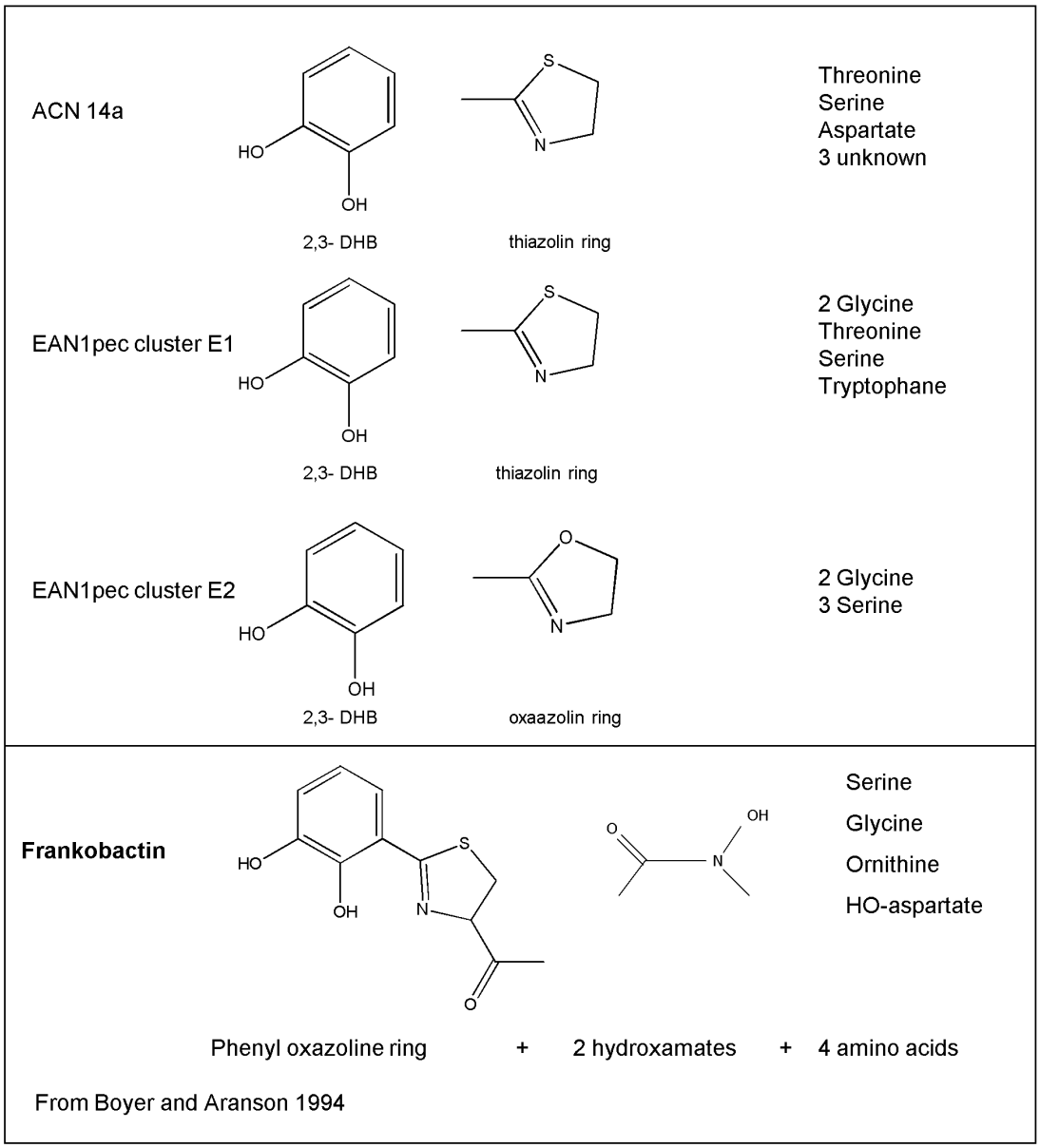

Figure 5 Structural components of the three putative DHB siderophores synthesized by ACN14a and EAN1pec strains compared with Frankobactin as proposed by Boyer and Aronson (1994)

hydroxamate group. Oxazoline rings are a common structure of several siderophores including agrobactin (Ong *et al.*, 1979), parabactin and vibriobactin (Griffiths *et al.*, 1984), where it may be derived from either serine or threonine. Analogous thiazolin ring are also encountered in anguibactin (Vibrio anguillarum), yersinibactin (Yersinia) or pyochelin (Pseudomonas aeruginosa).

Gene clusters involved in HYD siderophore biosynthesis were detected in a large majority of Frankia strains (38 among the 39 tested) and their characteristics (PCR fragment size and sequence) allow to differentiate cluster1 strains (infective on Alnus and Casuarina) from cluster3 (infective on Elaeagnaceae) (Table1, Fig. 2A). These two groups being consistent with 16S rRNA-based phylogeny and most of the polyphasic approaches on Frankia biodiversity, the acquisition of HYD gene could has been acquired long before emergence of the three Frankia lineages. Moreover, the 3 Frankia clusters are highly syntenic (Fig. 1A) indicating that few recombination events may have affected HYD siderophore biosynthesis genes, and thus illustrating its selective value. The synteny observed with the rhizobactin1021 biosynthesis cluster of Sinorhizobium meliloti, and the consistently supported grouping of this strain with actinobacteria branch in the phylogenetic tree based on rhbC homolog sequence suggest that those genes were probably obtained by S. meliloti's ancestor from an Actinobacterium via horizontal transfer during evolution.

Three Frankia genes in HYD cluster are highly homologous with E. coli iuc genes involved in aerobactin biosynthesis. Rhizobactin1021 and aerobactin, like many other hydroxamate siderophores are assembled by the NRPS-independent siderophore (NIS) pathway (Challis, 2005). Both aerobactin and rhizobactin1021 are citrate-based hydroxamate siderophores with similar structures, and strongly homologous receptor proteins (Lynch *et al.*, 2001). Five genes involved in rhizobactin1021-like siderophore biosynthesis were detected with a high synteny in the 3 Frankia genomes. The additional glutamine synthetase would catalyze formation of glutamic acid, a compound needed in large amounts to act as cofactor of RhbA protein and inducing, in the presence of the RhbB protein, 1,3 diaminopropane biosynthesis in rhizobactin1021 pathway (Lynch *et al.*, 2001). Homologs of rhbD and rhbG involved in acylation of siderophores (Challis, 2005; Lynch *et al.*, 2001)are not present. In a divergon, a diaminopimelate decarboxylase involved in lysine or cadaverin biosynthesis (Dewey, 1954; Dewey *et al.*, 1954) could produce the putative substrate for aerobactin biosynthesis. This divergon have been found involved in additional siderophore homologous

	Nucleotide position																		
	-50	-49	-48	-47	-46	-45	-44	-43	-42	-41	-40	-39	-38	-37	-36	-35	-34	-33	-32
AcN14a FRAAL6421	T	T	A	G	G	T	T	A	G	G	C	T	T	C	C	C	T	G	A
EAN1pec Franean1_0662	T	T	A	A	G	T	T	A	G	C	C	T	T	C	C	T	T	G	A
Ccl3 Francci3_4054	T	T	A	A	G	T	T	A	G	G	C	T	T	C	C	C	T	C	A
DtxR consensus binding site	T	.	A	G	G	.	.	A	G	.	C	T	.	.	C	C	T	.	A

Figure 6 Sequence alignment (5'-3') of the DtxR binding sites of *Frankia* upstream from the *rhbA* gene homologs compared with the minimal consensus sequence (Tao & Murphy, 1994). Nucleotide position is given for *Frankia* strains from the *rhbA* transcriptional start site.

to aerobactin. A siderophore highly similar to rhizobactin1021 and/or homologous to aerobactin could thus be synthetized by Frankia.

Using a heterologous iron uptake scheme with purified siderophores, we observed that, contrary to the other purified siderophore tested, exogenously supplied rhizobactin1021 can be directly used as iron source by AcN14a Frankia strain. However the rate of $^{59}Fe^{3+}$ uptake in the homologous experiment (with Frankia siderophore-containing supernatant) was representatively higher. Then due to the relatively tight specificity that exists between a siderophore and its cognate receptor (Hohnadel & Meyer, 1988), the respective iron acquisition system of Frankia and S. meliloti seem to have evolved similarly leading to the conservation of the transport system. These results corroborate that the iron-chelating compound synthesized by AcN14a strain is structurally similar to rhizobactin1021. The capacity to internalize rhizobactin1021 may signify that, in iron-limited environments, Frankia strains can utilize siderophores they do not synthesize themselves. This selective advantage may increase their competitivity during saprotrophic stage in the absence of the host plant.

Although genetic determinants of siderophore biosynthesis are present in all *Frankia* strains tested, siderophore expression needs to be evaluated. Taking into account the diversity of chemical structures predicted by the genetic analysis, we used the CAS assay described as a universal method to detect any strong iron binding compound, regardless of its chemical nature. CAS responses on unconcentrated supernatants were weak (ΔOD close or less than 0.1). This difficulty to detect significant siderophore production is often encountered on slow growing bacteria (Guerinot, 1991) and particularly in *Frankia,* leading to use highly concentrated samples. However, this procedure is responsible for known false positive responses (Winkelmann, 1991), due to the increase of phosphates, chelating agents other than siderophores and iron reductants, all interfering with the CAS assay. Weak supernatant concentrations can also be found for some amphiphilic siderophores, i.e rhizobactin1021 and mycobactin (Lynch *et al.*, 2001), which can be anchored to the bacteria through the lipid moiety of their molecule, thus being less released in the medium. However, *rhbG* homolog likely involved in the incorporation of the lipidic moiety in rhizobactin1021 molecule is absent in *Frankia* ACN14a.

Among the five CAS-positive strains (G2, Ea112, ACN, CpI1 and M16467), three have both genes involved in HYD and DHB siderophore biosynthesis. Whereas all CAS-negative strains (Ar24O2, $ARgP_5$, CH37, $Cg70_9$, $Cg70_4$, ORS020608, CcI3, Cj1-82) have

only siderophore biosynthesis HYD cluster. It is worth noting that all *Casuarina*-infective strains known for their genetic and physiological homogeneity also shared a unique siderophore amplicon pattern, and tested negative for siderophore production even though HYD cluster is likely present in all of them. This result could suggest a poor expression of HYD operon in this group of strains in the conditions used. This conclusion is supported by Aronson et al. (1994) on CcI3 strain and by Arahou et al. (1998) study where seven *Casuarina* strains were tested negative with the Csaky (hydroxamates specific) assay. However, strain Cj1-82 that gave an intermediate result in our study, tested positive in the study of Arahou et al (1998).

Regulation mechanisms of siderophore biosynthesis have never been described in *Frankia*. *fur* genes, first identified in *E. coli* (Hantke, 1981), mediate negative transcriptional regulation of siderophore production in response to iron and the associated oxidative stress. Fur-encoding gene homologs were detected in the 3 *Frankia* sequenced strains (FRAAL3168, Franean1_1532, and Francci3_3112) but no Fur binding sites were found in the siderophore clusters. DtxR is another negative iron regulator, first described in Actinobacteria, which regulates a set of genes similar to those regulated by *fur* in many Proteobacteria (Dussurget *et al.*, 1996). Gene *dtxR* is also present in the 3 *Frankia* strains (FRAAL 0908, Franean1_6206, Francci3_0424) and DtxR consensus binding site sequences (Tao & Murphy, 1994) were found in the three sequenced strains, with one or two mutations upstream *rhbA* homolog genes (Fig 6). Interestingly, *dtxR* gene is localized in an operon encoding enzymes involved in nitrogen fixation suggesting a probable link between iron regulation and nitrogen fixation.

In nitrogen fixing microorganisms, this link between nitrogen fixation activity and siderophore production is expected since the nitrogenase complex and the cytochroms needed for its functioning imply iron-rich molecules. *Frankia* is able to fix nitrogen in culture in N-deficient conditions. In this study, *Frankia* ACN14a cells were grown in an iron and nitrogen deficient medium in order to test the effect of the nitrogen fixation process on the siderophore production. No ^{59}Fe (either rhizobactin- or *Frankia* supernatant-mediated) uptake was detected suggesting that, in nitrogen-fixing conditions, *Frankia* strain AcN14a favors the nitrogen fixation process over the siderophore production. Similarly, an analysis with production-siderophore mutants in *Sinorhizobium meliloti* demonstrated a diminution in the ability to fix nitrogen *in planta* (Gill & Neilands, 1989) and rhizobactin1021 biosynthesis and transport genes are not expressed when nitrogenase genes are actively expressed (Lynch *et al.*, 2001). On the contrary, *Frankia* CeSI5 and 52065 strains exhibited linear (^{59}Fe)-frankobactin

uptake when they were cultured in an iron free and N-free as well as N-sufficient medium (Boyer *et al.*, 1999).

As shown on the large diversity of strains used in this work, siderophore production is widespread in *Frankia* genus and siderophore biosynthesis determinants present are coherent with phylogenetic positioning of the strains. Based on the present genomic analysis, each *Frankia* strain could produce from one to 4 (one hydroxamate- and 3 catechol-type) structurally different siderophores one of them being structurally close to and probably sharing strongly homologous transport system with rhizobactin1021. Further studies are required to identify this rhizobactin1021-like siderophore and confirm the predictions about DHB siderophore structure. More investigation will be necessary to assess in what conditions these genes would be expressed and their role during the symbiosis.

Acknowledgements – Thanks are expressed to G. Vansuyt (INRA, Dijon, France) for technical advices about siderophore detection and to P. Normand for fruitful discussions. Hector Bautista had a grant from CONACYT N° 48044 and M. Antonia Cruz was supported by a Mexico-France ECOS program.

References:

1. Appleby CA (1984) Leghemoglobin and *Rhizobium* Respiration. Annu Rev Plant Physiol, 35: 443-478

2. Arahou M, Diem HG and Sasson A (1998) Influence of iron depletion on growth and production of catechol siderophores by different *Frankia* strains. J Microbiol Biotechnol, 14: 31-36

3. Aronson D and Boyer G (1992) *Frankia* produces a hydroxamate siderophore under iron limitation. J Plant Nutrition 15: 2193-2201

4. Aronson D and Boyer G (1994) Growth and siderophore formation in six iron limited strains of *Frankia*. Soil Biol Biochem 26: 561-567

5. Bachhawat AK and Ghosh S (1987) Isolation and characterization of the outer membrane proteins of *Azospirillum brasilense*. J Gen Microbiol 133: 1751-1758

6. Bagnarol E, Popovici J, Alloisio N, Maréchal J, Pujic P, Normand P and Fernandez M (2007) Differential *Frankia* protein patterns induced by phenolic extracts from Myricaceae seeds. Physiol Plant 130: 380-390

7. Baker D and Torrey JG (1980) Characterization of an effective actinorhizal microsymbiont, *Frankia* sp. AvcI1 (Actinomycetales). Can J Microbiol 26: 1066-1071

8. Barona-Gomez F, Wong U, Giannakopulos AE, Derrick PJ and Challis GL (2004) Identification of a cluster of genes that directs desferrioxamine biosynthesis in *Streptomyces coelicolor* M145. Journal of the American Chemical Society 126: (50) 16282-16283

9. Barton LL, Johnson GV, Schitoskey K and Wertz M (1996) Siderophore-mediated iron metabolism in growth and nitrogen fixation by alfalfa nodulated with *Rhizobium meliloti.* J Plant Nutr, 19: (8) 1201-1210

10. Barton LL, Vester CR, Gill PR and Neilands JB (1994) The role of siderophores in symbiotic nitrogen-fixation by root nodule bacteria. In: Manthey JA, Crowley DE and Luster DG (ed) *T*he biochemistry of metal micronutrients in the rhizosphere. Lewis Publishers, Boca Raton, Fla. pp 1994 54-66

11. Benson DR, Vanden Heuvel BD and Potter D (2004) Actinorhizal Symbioses: Diversity and Biogeography. In: Gillings M (ed) Plant Microbiology Oxford: BIOS Scientific Publishers Ltd, pp 97-127

12. Benson HP, Boncompagni E and Guerinot ML (2005) An iron uptake operon required for proper nodule development in the *Bradyrhizobium japonicum*-soybean symbiosis. Mol Plant Microbe Interact 18: (9) 950-959

13. Bentley SD, Chater KF, Cerdeno-Tarraga AM, Challis GL, Thomson NR, James KD, Harris DE, Quail MA, Kieser H, Harper D, Bateman A, Brown S, Chandra G, Chen CW,

Collins M, Cronin A, Fraser A, Goble A, Hidalgo J, Hornsby T, Howarth S, Huang CH, Kieser T, Larke L, Murphy L, Oliver K, O'Neil S, Rabbinowitsch E, Rajandream MA, Rutherford K, Rutter S, Seeger K, Saunders D, Sharp S, Squares R, Squares S, Taylor K, Warren T, Wietzorrek A, Woodward J, Barrell BG, Parkhill J and Hopwood DA (2002) Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). Nature 417: (6885) 141-147.

14. Boyer GL and Aronson DB (1994) Iron Uptake and Siderophore Formation in the Actinorhizal Symbiont *Frankia*. In: Manthey JA, Crowley DE and Luster DG (ed) *T*he biochemistry of metal micronutrients in the rhizosphere. Lewis Publishers, Boca Raton, Fla. pp 41-54

15. Boyer GL, Guillam A and Trick CG (1987) Iron chelation and Uptake. In: Edited by Fay P and Van Baalen C (ed) *Cyanobacteria.* Elsevier Biomedical, London pp 415-436

16. Boyer GL, Kane SA, Alexander JA and Aronson DB (1999) Siderophore formation in iron-limited cultures of *Frankia* sp. strain 52065 and *Frankia* sp. strain CeSI5. Can J Bot 77: 1316-1320

17. Burggraaf A and Shipton W (1983) Studies on the growth of *Frankia* isolates in relation to infectivity and nitrogen fixation (acetylene reduction). Can J Bot 61: 2774-2782

18. Callaham D, Del Tredici P and Torrey J (1978) Isolation and cultivation *in vitro* of the actinomycete causing root nodulation in *Comptonia*. Science 199: 899-902

19. Challis GL (2005) A widely distributed bacterial pathway for siderophore biosynthesis independent of nonribosomal peptide synthetases. Chembiochem 6: (4) 601-611

20. Challis GL and Ravel J (2000) Coelichelin, a new peptide siderophore encoded by the *Streptomyces coelicolor* genome: structure prediction from the sequence of its non-ribosomal peptide synthetase. FEMS Microbiol Lett 187: (2) 111-114

21. Clawson M and Benson D (1999) Natural diversity of *Frankia* strains in actinorhizal root nodules from promiscuous hosts in the family Myricaceae. Appl Environ Microbiol 65: 4521-4527

22. Crosa JH and Walsh CT (2002) Genetics and assembly line enzymology of siderophore biosynthesis in bacteria. Microbiol Mol Biol Rev 66: (2) 223-249

23. De M, Basu M and Chakrabartty PK (2003) Increased synthesis of dihydroxybenzoic acid in the presence of aluminum by *Rhizobium* MO1. Acta Microbiologica Polonica 52: (2) 195-199

24. Dewey DL (1954) The distibution of diaminopimelic acid decarboxylase among some organisms of the coli-aerogenes group and certain other bacteria. J Gen Microbiol 11: 307-312

25. Dewey DL, Hoare DS and Work E (1954) Diaminopimelic acid decarboxylase in cells and extracts of *Escherichia coli* and *Aerobacter aerogenes*. Biochem J 58: 523-531

26. Diem H, Gauthier D and Dommergues Y (1982) Isolation of *Frankia* from nodules of *Casuarina equisetifolia*. Can J Microbiol 28: 526-530

27. Diem H, Gauthier D and Dommergues Y (1983) An effective strain of *Frankia* from *Casuarina* sp. Can J Bot 61: 2815-2821

28. Dussurget O, Rodriguez M and Smith I (1996) An *ideR* mutant of *Mycobacterium smegmatis* has derepressed siderophore production and an altered oxidative-stress response. Mol Microbiol 22: (3) 535-544

29. Feistner GJ and Beaman BL (1987) Characterization of 2,3-dihydroxybenzoic acid from *Nocardia asteroides* GUH-2. J Bacteriol 169: (9) 3982-3987

30. Fernandez M, Meugnier H, Grimont P and Bardin R (1989) Deoxyribonucleic acid relatedness among members of the genus *Frankia*. Int J Syst Bacteriol, 39: 424-429

31. Fiedler HP, Krastel P, Muller J, Gebhardt K and Zeeck A (2001) Enterobactin: the characteristic catecholate siderophore of Enterobacteriaceae is produced by S*treptomyces species*. FEMS Microbiol Lett 196: (2) 147-151

32. Gill PR, Jr. and Neilands JB (1989) Cloning a genomic region required for a high-affinity iron-uptake system in *Rhizobium meliloti* 1021. Mol Microbiol 3: (9) 1183-1189

33. Griffiths GL, Sigel SP, Payne SM and Neilands JB (1984) Vibriobactin, a siderophore from *Vibrio cholerae*. J Biochem biol chem 259: (1) 383-385

34. Guerinot ML (1991) Iron uptake and metabolism in the rhizobia/legume symbioses. Plant and Soil 130: 199-209

35. Guerinot ML and Yi Y (1994) Iron: Nutritious, Noxious, and Not Readily Available. Plant Physiol 104: (3) 815-820

36. Hantke K (1981) Regulation of ferric iron transport in *Escherichia coli* K12: isolation of a constitutive mutant. Mol Gen Genet 182: (2) 288-292

37. Higgins DG and Sharp PM (1988) CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. Gene 73: (1) 237-244

38. Hohnadel D and Meyer JM (1988) Specificity of pyoverdine-mediated iron uptake among fluorescent *Pseudomonas strains*. J Bacteriol 170: (10) 4865-4873

39. Jamann S, Fernandez M and Moiroud A (1992) Genetic diversity of Elaeagnaceae-infective *Frankia* strains isolated from various soils. Acta Oecolog 13: 395-405

40. Khandelwal SR, Manwar AV, Chaudhari BL and Chincholkar SB (2002) Siderophoregenic Bradyrhizobia boost yield of soybean. Appl Biochem Biotech 102-103: (1-6) 155-168

41. Langman L, Young IG, Frost GE, Rosenberg H and Gibson F (1972) Enterochelin system of iron transport in *Escherichia coli*: mutations affecting ferric-enterochelin esterase. J Bacteriol 112: (3) 1142-1149

42. Lechevalier M (1986) Catalog of *Frankia* strains. The Actinomycetes, 19: 131-162

43.Lynch D, O'Brien J, Welch T, Clarke P, Cuiv PO, Crosa JH and O'Connell M (2001) Genetic organization of the region encoding regulation, biosynthesis, and transport of rhizobactin 1021, a siderophore produced by *Sinorhizobium meliloti*. J Bacteriol 183: (8) 2576-2585

44. Marahiel MA, Stachelhaus T and Mootz HD (1997) Modular Peptide Synthetases Involved in Nonribosomal Peptide Synthesis. Chem Rev 97: (7) 2651-2674

45. Milagres AM, Machuca A and Napoleao D (1999) Detection of siderophore production from several fungi and bacteria by a modification of chrome azurol S (CAS) agar plate assay. J Microbiol Methods 37: (1) 1-6

46. Murry M, Fontaine M and Torrey J (1984) Growth kinetics and nitrogenase induction in *Frankia* sp. HFPArI3 grown in batch culture. Plant Soil, 78: 61-78

47. Neilands JB (1981) Iron absorption and transport in microorganisms. Annu Rev Nutr 1: 27-46

48. Neilands JB (1981) Microbial iron compounds. Annu Rev Biochem 50: 715-731

49.Neilands JB (1982) Microbial envelope proteins related to iron. Annu Rev Microbiol 36: 285-309

50. Nicholas DJD (1952) The use of fungy for determining trace metals in biological material. Analyst, 77: 629-641

51. Normand P and Lalonde M (1982) Evaluation of *Frankia* strains isolated from provenances of two *Alnus* species. Can J Microbiol 28: 1133-1142

52. Normand P and Lalonde M (1986) The genetics of actinorhizal *Frankia*: a review. Plant Soil, 90: 429-453

53.Normand P, Lapierre P, Tisa L, Gogarten JP, Alloisio N, Bagnarol E, Bassi CA, Berry AM, Bickhard DM, Choisne N, Couloux A, Cournoyer B, Cruveiller S, Daubin V, Demange N, Francino MP, Goltsman E, Huang Y, Kopp O, Labarre L, Lapidus A, Lavire C, Marechal J, Martinez M, Mastronunzio JE, Mullin BC, Niemann J, Pujic P, Rawnsley T, Rouy Z, Schenowitz C, Sellstedt A, Tavares F, Tomkins JP, Vallenet D, Valverde C, Wall L, Wang Y,

Medigue C and Benson DR (2007) Genome characteristics of facultatively symbiotic *Frankia* sp. strains reflect host range and host plant biogeography. Genome Research, 17: 7 - 15

54.Normand P, Orso S, Cournoyer B, Jeannin P, Chapelon C, Dawson J, Evtushenko L and Misra AK (1996) Molecular phylogeny of the genus *Frankia* and related genera and emendation of the family Frankiaceae. Int J Syst Bacteriol 46: (1) 1-9

55. Normand P, Queiroux C, Tisa L, Benson D, Cruveiller S, Rouy Z and Medigue C (2007) Exploring the genomes of *Frankia* sp. Physiol Plantarum 13: 331-343

56. Ong SA, Peterson T and Neilands JB (1979) Agrobactin, a siderophore from *Agrobacterium tumefaciens*. J Biol Chem 254: (6) 1860-1865

57. Puppo A, Dimitrijevic L, Diem H and Dommergues Y (1985) Homogeneity of superoxide dismutase patterns in *Frankia* strains from Casuarinaceae. FEMS Microbiol Lett 30: 43-46

58. Saitou N and Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4: (4) 406-425.

59. Sangwan I and O'Brian M R (1992) Characterization of delta-Aminolevulinic Acid Formation in Soybean Root Nodules. Plant Physiol 98: (3) 1074-1079

60. Schwyn B and Neilands JB (1987) Universal chemical assay for the detection and determination of siderophores. Anal Biochem 160: (1) 47-56

61. Simonet P, Thi LN, Moiroud A and Bardin R (1989) Diversity of *Frankia* strains isolated from a single alder stand. Plant Soil, 118: 13-22

62. Simpson FB and Neilands JB (1976) Siderochromes in cyanophyaceae: isolation and characterization of schizokinen from Anabaena sp. J Phycol, 12: 44-48

63. Stevens JB, Carter RA, Hussain H, Carson KC, Dilworth MJ and Johnston AW (1999) The *fhu* genes of *Rhizobium leguminosarum*, specifying siderophore uptake proteins: *fhuDCB* are adjacent to a pseudogene version of *fhuA*. Microbiol 14: 593-601

64. Tao X and Murphy JR (1994) Determination of the minimal essential nucleotide sequence for diphtheria tox repressor binding by in vitro affinity selection. PNAS 91: (20) 9646-9650

65. Winkelman G E (1991) Handbook of microbial iron chelates. CRC Press, Boca raton, Floride.

66.Wu D, Hugenholtz P, Mavromatis K, Pukall R, Dalin E, Ivanova NN, Kunin V, Goodwin L, Wu M, Tindall BJ, Hooper SD, Pati A, Lykidis A, Spring S, Anderson IJ, D'Haeseleer P, Zemla A, Singer M, Lapidus A, Nolan M, Copeland A, Han C, Chen F, Cheng JF, Lucas S, Kerfeld C, Lang E, Gronow S, Chain P, Bruce D, Rubin EM, Kyrpides NC, Klenk HP and Eisen JA (2009) A phylogeny-driven genomic encyclopaedia of Bacteria and Archaea. Nature 462: (7276) 1056-1060

67. Zhang Z, Lopez MF and Torrey JG (1984) A comparison of cultural characteristics and infectivity of *Frankia* isolates from root nodules of *Casuarina* species. Plant and Soil, 78:79-90.

IV.2. Genetic occurrence and molecular phylogeny of the sodF gene in Frankia

Introduction

Oxidative stress to cells results from the presence of radical-forming oxygen species. The first radical produced in the ROS (reactive oxygen species) cascade is the superoxide (O2-) which is toxic to the biological systems in general (Apel & Hirt, 2004) because all the major biological molecules can be damaged by oxygen derived-radicals. The ROS inactivates the enzymes by amino acid oxidation and by attacking the active centers of metalloenzymes (Santos *et al.*, 2000).

The accumulation of a large quantity of ROS, also called "oxidative burst", has been reported as a common response from the hosts toward a pathogen attack (Lamb & Dixon, 1997). Knowing that the first stages in the symbiosis establishment imply the host invasion by the microsymbiont, it is expected the host elicits also a defensive response. Hence microsymbionts, like *Frankia*, have developed different mechanisms to counteract the accumulation of ROS, and thus the presence of antioxidant defense is one of the evolutionary adaptations to the symbiotic lifestyle (Glyan'ko & Vasil'eva, 2010).

The superoxide dismutases are the first line of defense against the ROS (O2.-) since they are metalloenzymes involved in detoxification of reactive oxygen by dismutation of the superoxide radical anion (Fridovich, 1995). The superoxide dismutases are also considered as essential since SOD-deficient mutants in prokaryotes are hypersensitive to oxygen and moreover in organisms lacking other antioxidant enzymes, such as catalase and peroxidase, SOD is expressed without exception (Asada *et al.*, 1977; Richier *et al.*, 2005).

In agreement with the metal present in the active site, four types of SOD have been identified. The first consists of SOD with Cu(II) plus Zn(II), another with Mn(III), a third type with Fe(III) and a fourth with Ni(II/III). The FeSOD and MnSOD types are found predominantly in prokaryotes and mitochondria, they are closely related with regard to three-dimensional structure and amino acid sequence, and close phylogenetic relationships between these two kinds of enzymes suggest a common origin. Moreover the FeSOD is called also cambialistic SOD thanks to its ability of using Fe or MN depending on environmental metal availability (Seib *et al.*, 2006).

In the research of phylogenetic markers more informative that the 16S the *sodA* has proposed to carry out this kind of analysis. Its election was not hazardous, because in human

pathogens this function is considered as virulence factor and so it is directly implicated in host-pathogen interactions (Roggenkamp *et al.*, 1997; Tsolis *et al.*, 1995).

Hence, *sodA* has been used successfully to improve diagnostics for the *Streptococcus*, *Enterococcus*, and *Pasteurella* species (Gautier *et al.*, 2005; Nomoto *et al.*, 2008; Poyart *et al.*, 2000). In addition such gene has become one of the reference markers in *Nocardia* species definition (Rodriguez-Nava *et al.*, 2007) In all cases, this gene sequence exhibited a strong phylogenetic signal, displaying topologies similar to those obtained with the 16S but with a higher discrimination level, achieving the differentiation between closely related strains at fine scale.

Frankia SOD activity levels is among the highest reported in prokaryotes (Steele & Stowers, 1986). It is not surprising because similar mechanisms have been encountered between the host-pathogen interactions and the symbiotic establishment. Thus at root penetration and invasion stages *Frankia* confronts the oxidative response triggered by the host plant. Moreover is has been reported that the first response of *Frankia* to the plant exudates is a classical stress response to toxic root products. *Frankia alni* cultured in presence of *A. glutinosa* exudates showed an overexpression of a ferric superoxide dismutase coded by a *sodA* orthologous, *sodF* (Bagnarol *et al.*, 2007). After those initial stages high levels of SOD are also required for the efficacy of nodule morphogenesis and for the nitrogen fixation.

Against to that expected during symbiosis, no correlation between the SOD (*sodF*) activity and the nitrogen fixation rate has been found. The levels of *sodF* expression are constitutive and there is not significant variation in its production (Alskog & Huss-Danell, 1997). Hence, it has been hypothesized the ferric superoxidase is involved in the symbiosis establishment rather than in the nitrogen fixation. Moreover in *Rhizobium* reports presented evidence that bacterial SOD effects extend beyond protection of the nitrogenase complex since SOD-defective mutants display undesirable effects at all steps of symbiosis, including infection, nodulation and bacteroid differentiation (Santos *et al.*, 2000).

All this suggest that *sodF* gene is directly related with symbiosis establishment and that host plant exercises selective pressure over this function. Thus in this work the occurrence of *sodF* gene was studied in a large panel of *Frankia* strains belonging to different phylogenetic groups. Then a comparative *sodF* sequence analysis was developed in order to construct a phylogeny of the genus. Moreover this work planned to contributes to the *Frankia*-Actinorhizal plant symbiosis initial stage study.

IV.2.1. Material and Methods

Frankia strains and DNA extraction

In this study were employed cultured *Frankia* strains, their host plant and the geographic origin are listed in the Table 1. The strains were grown at 28°C in F medium (*Elaeagnus*-infective strains) (Simonet *et al.*, 1985), F medium containing Tween 80 (*Alnus*-infective strains) or BAP medium (other strains) (Murry *et al.*, 1984c) with weekly manual shaken.

Genomic DNA from cell cultures was extracted from 30ml of fresh culture with the DNeasy Plant Mini kit (QIAGEN) according to the manufacturer's instructions.

The primers design was based on the sequence coding for the FeSOD (Franean1_3497, Francci3_2817 and FRAAL4337) retrieved in the three genomic sequences of *Frankia* until now available. The primers F6116-sodFfdT7F (5' TAATACGACTCACTATAGGGTGCACCACGACAAGCACCA 3') and F6117-sodFrvT3R (5'ATTAACCCTCACTAAAGGGAGACGTTCTTGTACTGCAGGTA 3') flank a 400pb fragment that comprises the positions 115 to 514 of the sodF coding sequence.

PCR reactions were carried out in a final volume of 100 µl. Each PCR reaction contained 5 µl template ADN (50-200ng), 10 µl of 10 X PCR buffer, 5 µl of each primer (10 mM), 10 µl of a dNTP mix (2.5 mM), 4µl of $MgCl_2$ (25 mM), 5 µl DMSO, 2.5 U Taq DNA polymerase and 61µl of sterile MiliQ water. The reaction conditions were initial denaturation at 95°C for 5 min, followed by 30 cycles of the denaturation at 95°C for 1 min, annealing for 1 min at 58 °C, and extension at 72°C for 1 min. A final extension step was performed at 72°C for 10 min. Amplified products were resolved on a 1.0% agarosa gel.

Sequence analysis:

The editing of sequences was developed in BioEdit (Hall, 1999) and the alignments were performed with the Clustal W as implemented in BioEdit (Thompson *et al.*, 1994). Statistics such as the number and proportion of polymorphic sites, mean G+C content and ratio of mean synonymous substitutions per synonymous site/mean nonsynonymous substitutions per nonsynonymous site (ds/dn) were calculated using the START2 program (Jolley *et al.*, 2001). Haplotype numbers of indels were determined by using DnaSP5. ML analyses utilized TREEFINDER (Jobb *et al.*, 2004). Following the Akaike Information Criterion (AIC) (Posada & Buckley, 2004) computed with the "Propose Model" option of

Table 1 Liste of *Frankia* strains and actinorhizal nodules

Strain designation	Trivial designation	Source host	Geographical origin	Reference
ULQ0132022024	ARgN22d	*Alnus rugosa*	Québec, Canada	(Normand & Lalonde, 1982)
HFP013103	ArI3	*Alnus rubra*	Oregon, USA	(Normand & Lalonde, 1982)
	$Ac2_{18}$	*Alnus cordata*	Miribel, France	Unpublished
	$Ac23_{23}$	*Alnus cordata*	Saou (France)	Unpublished
	Ar24H5	*Alnus rubra*	Orléans, France	(Simonet *et al.*, 1989)
	$Ac24I_5$	*Alnus cordata*	Orléans, France	Unpublished
LLR01322	AirI1	*Alnus incana ssp. rugosa*	Vermont, USA	(Lechevalier, 1986)
ULF0107024251	$Ag24_{251}$	*Alnus glutinosa*	Orléans, France	(Fernandez *et al.*, 1989a)
ULF010102171	AcVc1	*Alnus cordata*	Corsica, France	(Fernandez *et al.*, 1989a)
ULF01010244	AcoN24d	*Alnus cordata*	Orléans, France	(Simonet *et al.*, 1984)
DDB010110	AvcI1	*Alnus viridis ssp. crispa*	Ontario, Canada	(Baker & Torrey, 1980)
ULF014101715	AVN17o	*Alnus viridis*	La Toussuire, France	(Fernandez *et al.*, 1989a)
ULF014102203	AVL3	*Alnus viridis*	Lautaret, France	(Fernandez *et al.*, 1989a)
ULF010102340	$Ac23_{40}$	*Alnus crispa*	Orléans, France	(Fernandez *et al.*, 1989a)
ULQ010201401	ACN14a	*Alnus crispa*	Tadoussaq, Canada	(Normand & Lalonde, 1986)
ULQ0132105009	$ARgP5^{AG}$	*Alnus rugosa*	Québec, Canada	(Normand & Lalonde, 1986)
	Ag21D1	*Alnus glutinosa*	Corsica, France	Unpublished
	I38	*Alnus incana*	La pallud, France	Unpublished
	$Ai96_6$	*Alnus incana*	101 (Lacrans (01) Carrière Famy	Unpublished
	$Av59_7$	*Alnus viridis*	Alpe du Grand Serre, France	Unpublished
ULF01070602	$Mg60_2^{AG}$	*Alnus glutinosa*	Landes, France	(Fernandez *et al.*, 1989a)
	AJ01	*Alnus japonica*	Makabe, Ibaraki Pref., Japan	Unpublished
RBR162021	M16467	*Myrica pensilvanica*	New Jersey, USA	(Clawson & Benson, 1999)
RBR162013	M16477	*Myrica pensilvanica*	New Jersey, USA	Unpublished
HFP020203	CcI_3	*Casuarina cunninghamiana*	Florida, USA	(Zhang *et al.*, 1984)
ORS022602	TA	*Allocasuarina torulosa*	Australie	(Zhang *et al.*, 1984)
ORS020608	BR	*Casuarina equisetifolia*	Brazil	(Puppo *et al.*, 1985)
ORS021001	CjI-82	*Casuarina junghuniana*	Thailand	(Diem *et al.*, 1983)
	$Cg70_3$	*Casuarina glauca*	Inde	Unpublished
	$Cg70_4$	*Casuarina glauca*	Inde	Unpublished
HFP020202	CcI_2	*Casuarina cunninghamiana*	Florida, USA	(Zhang *et al.*, 1984)
ORS020604	G_2	*Casuarina equisetifolia*	Guadeloupe	(Diem *et al.*, 1982)
ULF130100112	Ea112	*Elaeagnus angustifolia*	Ecully, France	(Fernandez *et al.*, 1989a)
ULF130100303	$Ea36_7$	*Elaeagnus angustifolia*	Ecully, France	(Fernandez *et al.*, 1989a)
ULQ130100144	EAN1pec	*Elaeagnus angustifolia*	Illinois (U.S.A)	(Lalonde *et al.*, 1981)
ULF130103502	$Ea35_2$	*Elaeagnus angustifolia*	Sutri, Italy	(Jamann *et al.*, 1992a)
	$Ea48_1$	*Elaeagnus angustifolia*	St Etienne de Tinée, France	Unpublished
ULQ132500106	EUN1f	*Elaeagnus umbellata*	Illinois, USA	(Lalonde *et al.*, 1981)
	$Ea48_4$	*Elaeagnus angustifolia*	St Etienne de Tinée, France	(Jamann *et al.*, 1992a)
UFL130100804	Ea84	*Elaeagnus angustifolia*	Pont en Royan, France	(Jamann *et al.*, 1992a)
UFL130100701	$Ea7_1$	*Elaeagnus angustifolia*	Toulon, France	(Jamann *et al.*, 1992a)

	BMG5.3	*Elaeagnus angustifolia*	Tunisia	(Gtari *et al.*, 2004)
ORS060501		*Colletia spinossisima*	Argentina	(Gauthier *et al.*, 1984)
	$Hr75_2$	*Hippophaë rhmanoides*	Station 7 N75, France	Unpublished
	ChI_7	*Colletia hystrix*	Chile	(Caru, 1993)
ORS140102	CH37	*Hippophaë rhmanoides*	Nogent sur Marne, France	(Prin *et al.*, 1991)
Nodules				
	Dryas	Canada		
	Av200	*Alnus viridis*	Bionassay, France	
	Av201	*Alnus viridis*	Bionassay, France	
	Coriaria R	*Coriaria ruscifolia*	Mexique	
	Coriaria M	*Coriaria myrtifolia*	France	
	Datisca	*Datisca glomerata*	Argentine	

TREEFINDER, the TVM [Optimum,Empirical]:G[Optimum]:5 was employed as the best-fitting model of nucleotide substitution. The node reliability was assessed using approximate bootstrap tests (LR-ELW edge support: the expected-likelihood weights applied to local rearrangements of tree topology (Strimmer & Rambaut, 2002) with 1000 replications.

IV.2.2. Results

PCR amplification

An internal fragment of the Fe-superoxide dismutase was amplified by PCR. Fifty strains yielded unambiguous positive amplification and no size polymorphism was found. In all cases a single fragment was obtained. Only the group of Nod-Fix- strains yielded negative amplification reactions, repeated attempts to amplify the fragment including variations in the PCR conditions were unsuccessful. Besides, Blast researches were performed in the genome of the EuI1c strain (Nod-/Fix) and they did not retrieved positive results.

Sequence analysis

The sequence yielded by the set of primers employed represent 64% of the coding region of the *sodF* gene. The mean standard content of G+C (67 %) is consistent with that reported for (66-72%) the genus *Frankia*. Among the fifty three strains analyzed 31 haplotypes were detected and 38% of the sites were polymorphic. Two statistic test were performed in order to determine the selective pressure that are leading the evolution of the superoxide dismutase locus. First the Tajima's D test yielded a value of 0.90 and second the dN/dS ratio for the whole panel of strains was 0.89 both results suggest that this locus is under strong neutral selection pressure.

Phylogenetic analysis

The *sodF* gene is highly discriminative since an overall mean distance of 0.123 was exhibited. The dendrogram constructed by maximum likelihood method based on the *sodF* gene sequences revealed distinct clusters within the *Frankia* genus (Fig 1). The genus was splitted in two major groups without statistical support; the first one (supported LR-ELW value of 99) contains the *Alnus*, *Myrica* and *Casuarina* compatible strains as well as the *Elaeagnaceae* strains corresponding to a part of the genomospecies G6 (described by AFLP and MLSA), but interestingly it corresponds to genomospecies GS4 (described by Fernandez

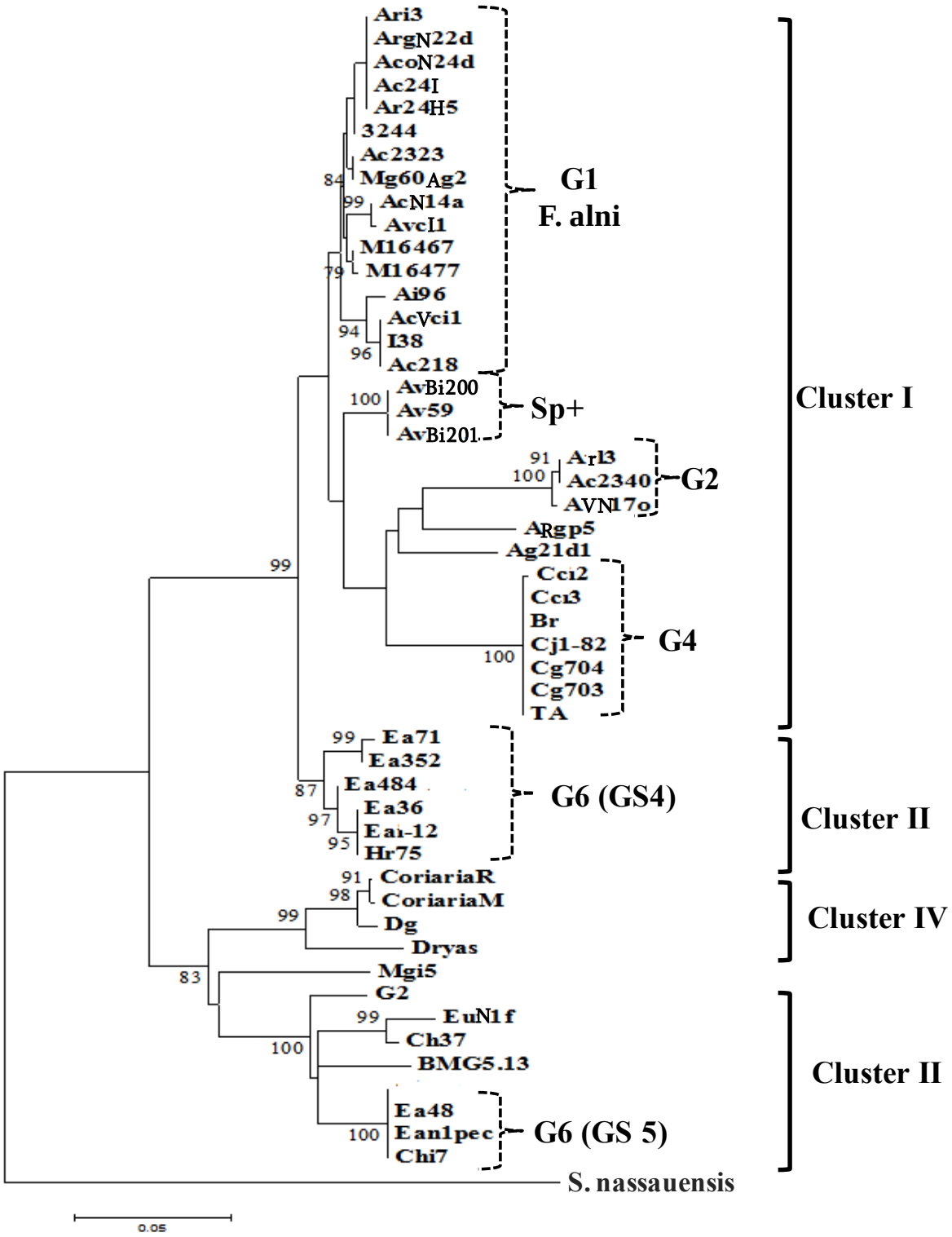

Figure 1 Maximum Likelihood tree of the genus *Frankia* based in the analysis of *sodF* partial sequence. The numbers at the nodes correspond to the LR-ELW (1000 replicates). Scale represents mutations per nucleotide. The genospecies and clusters are named in agreement with the nomenclature established with AFLP and MLSA studies.

et al., 1989). The second cluster (supported by a bootstrap value of 83) contains the group of non-isolated strains of *Dryas*, *Coriaria* and *Datisca* (boostrap of 99%) and a heterogeneous group of *Elaeagnaceae* compatible strains which includes the other part of the genomospecies G6 which correspond to GS5 described by Fernandez et al (1989). The most flagrant divergence found with respect phylogeny depicted by AFLP and MLSA data is the splitting of the *Elaeagnaceae* group. Some internal groups were well supported, as is the case of *Casuarina* strains (G4) and G2, both supported by LR-ELW values of 100. Besides, the *sodF* marker does not support *F. alni,* within this group subclustering we evidenced with the generation of coherent groups i.e. *A. viridis* Sp+ strains.

IV.2.3. Discussion

At present bacterial systematic relies in the nucleotide or aminoacids sequence comparisons, because is relatively simple and rapid (Hanage *et al.*, 2006). In this work the phylogeny of a relevant ecological marker was made in order to unveal the relationships among the *Frankia* strains. The *sodF* gene codes for a Fe-superoxide dismutase which is involved in the first line of defense against reactive oxygen species ROS. The parallelism that exist between the early steps of the intracellular interactions between *Frankia*-AP and between AP-phytopathogens suggest that *Frankia* is confronted to ROS when it enters the root (Tavares *et al.*, 2007).

In *Frankia* the superoxide dismutase genes have been poorly studied, a lack of correlation between FeSOD activity and nitrogen fixation activity nitrogenase was suggested since any clear relationship was found between ARA and SOD or even catalase activities per vesicle cluster protein (Alskog & Huss-Danell, 1997). In contrast, it was reported a FeSOD hyper expression, at a similar rate, among ACN14a, Ea1 12 and CcI3 when they were confronted to methanolyc extracts of *Myrica's* phenolic compounds (Bagnarol *et al.*, 2007). Suggesting that FeSOD is likely involved in the symbiosis establishment.

Genome mining allowed states that the three *Frankia* sequenced strains code for at least one superoxide dismutase. The three strains possess a gene coding for a FeSOD, besides the strain EAN1pec codes also for a NiSOD (Franean1_0977) and a Cu/Zn SOD (Franean1_4059). The CcI3 codes additionally for a Cu/Zn (Francci3_2876). It is interesting to note that the *Alnus* compatible strain ACN14a does not code for additional superoxide dismutases. The occurrence of *sodF* within the genus seems to be wide as was proved by the

positive PCR detection in the panel of strain here employed (with the exception of Nod-Fix- strains).

A *sodF* ortholog, *sodA* (Mn-superoxide dismutase) has been included in multigenic approaches in phylogenetic trials studies of some other genera of actinobacteria (Devulder *et al.*, 2005; Jurado *et al.*, 2008; Rodriguez-Nava *et al.*, 2007). In all cases these gene sequence displayed a strong phylogenetic signal, obtaining phylogenies similar to those obtained with the 16S but with higher discrimination level, achieving the differentiation between closely related strains at fine scale. In addition to those characteristics the *sodF* sequence might be subjected to purifying selection, therefore a based on this gene phylogeny was constructed.

In summarizing, all the data until now available, support the formation of three major clusters (Normand *et al.*, 1996), however with *sodF,* only two were depicted and no statistical support was found for the genus conception. These results disagree with the phylogeny constructed by MLSA of housekeeping genes (See chapter II). And thus we conclude that *sodF* must not be included in this kind of analysis.

Nonetheless, FeSOD coding gene may be useful to differentiate strains at a finer level. Since it has efficiently splitted a group that have been described as homogenous by other genotyping methods, since they have exhibited 49% of DNA relatedness (Fernandez *et al.*, 1989a); they can not be differentiated by the 16S-23S IGS (Nazaret *et al.*, 1991) and blurred *nif D-K* polymorphism has been yielded (Lumini & Bosco, 1999). With MLSA both genomospecies were clustered in a coherent group and they rather represented subspecies. Since the most evident difference between these groups is their geographic site of origin (The GS4 comprises only French and Italian isolates while the GS5 contains isolates from a wide geographical range) this marker might give insights into the ecological adaptation of each group of strains. In addition GS4 members variability did not seem to be affected by edaphic factors (Jamann *et al.*, 1992a).

To summarize, the *sodF* sequence display a non canonical *Frankia* phylogeny and interesting groupings were displayed, enabling the discrimination a group closely related strains and allowed to emit hypothesis about host influence on *Frankia* diversification processes.

Due to its conflictive phylogenetic signal we do not advise to use of *sodF* gene for the phylogenetic inference of the genus *Frankia*, although it can be performed to distinguish between close genomic groups of *Elaeagnaceae* strains.

The cluster II of Nod-/Fix- strains lack *sodF* function indirectly reinforces the hypothesis that this gene is involved in the symbiosis establishment.

Bibliography

1. **Alskog, G., and K. Huss-Danell.** 1997. SOD, catalase and N2ase activities of symbiotic Frankia (Alnus incana) in response to different oxygen tensions. Physiol. Plant. **99:**286-292.
2. **Apel, K., and H. Hirt.** 2004. Reactive oxygen species: metabolism, oxidative stress, and signal transduction. Annu Rev Plant Biol **55:**373-99.
3. **Asada, K., S. Kanematsu, and K. Uchida.** 1977. Superoxide dismutases in photosynthetic organisms: absence of the cuprozinc enzyme in eukaryotic algae. Arch Biochem Biophys **179:**243-56.
4. **Bagnarol, E., J. Popovici, N. Alloisio, J. Maréchal, P. Pujic, P. Normand, and M. Fernandez.** 2007. Differential Frankia protein patterns induced by phenolic extracts from Myricaceae seeds. Physiol. Plant.
5. **Baker, D.** 1987. Relationships among pure-cultured strains of *Frankia* based on host specificity. Physiol Plant **70:**245-248.
6. **Baker, D., and J. G. Torrey.** 1980. Characterization of an effective actinorhizal microsymbiont, *Frankia* sp. AvcI1 (Actinomycetales). Canadian Journal of Microbiology **26:**1066-1071.
7. **Burggraaf, A., and W. Shipton.** 1983. Studies on the growth of *Frankia* isolates in relation to infectivity and nitrogen fixation (acetylene reduction). Can J Bot **61:**2774-2782.
8. **Caru, M.** 1993. Characterization of native Frankia strains isolated from chilean shrubs (Rhamnaceae). Plant and Soil **157:**137-145.
9. **Clawson, M., and D. Benson.** 1999. Natural diversity of Frankia strains in actinorhizal root nodules from promiscuous hosts in the family Myricaceae. Appl Environ Microbiol **65:**4521-4527.
10. **Devulder, G., M. Perouse de Montclos, and J. P. Flandrois.** 2005. A multigene approach to phylogenetic analysis using the genus Mycobacterium as a model. Int J Syst Evol Microbiol **55:**293-302.
11. **Diem, H., D. Gauthier, and Y. Dommergues.** 1983. An effective strain of *Frankia* from *Casuarina* sp. Can J Bot **61:**2815-2821.
12. **Diem, H., D. Gauthier, and Y. Dommergues.** 1982. Isolation of *Frankia* from nodules of *Casuarina equisetifolia*. Can J Microbiol **28:**526-530.
13. **Fernandez, M., H. Meugnier, P. Grimont, and R. Bardin.** 1989. Deoxyribonucleic acid relatedness among members of the genus *Frankia*. Int J Syst Bacteriol **39:**424-429.
14. **Fridovich, I.** 1995. Superoxide radical and superoxide dismutases. Annu Rev Biochem **64:**97-112.
15. **Gauthier, D., L. Frioni, H. Diem, and Y. Dommergues.** 1984. The *Colletia spinosissima-Frankia* symbiosis. Oecolog Plant **5:**231-239.
16. **Gautier, A., D. Dubois, F. Escande, J. Avril, P. Trieu-Cuot, and O. Gaillot.** 2005. Rapid and accurate identification of human isolates of Pasteurella and related species by sequencing the sodA gene. J Clin Microbiol **43:**2307-2314.
17. **Glyan'ko, A. K., and G. G. Vasil'eva.** 2010. Reactive oxygen and nitrogen species in Legume-Rhizobial Symbiosis: A review. Applied Biochemistry and Microbiology **46:**15-22.

18. **Gtari, M., L. Brusetti, G. Skander, D. Mora, A. Boudabous, and D. Daffonchio.** 2004. Isolation of Elaeagnus-compatible Frankia from soils collected in Tunisia. FEMS Microbiol Lett **234:**349-55.
19. **Hall, T. A.** 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symposium Series **41:**95-98.
20. **Hanage, W. P., C. Fraser, and B. G. Spratt.** 2006. Sequences, sequence clusters and bacterial species. Philos Trans R Soc Lond B Biol Sci **361:**1917-27.
21. **Jamann, S., M. Fernandez, and A. Moiroud.** 1992. Genetic diversity of Elaeagnaceae-infective *Frankia* strains isolated from various soils. Acta Oecolog **13:**395-405.
22. **Jobb, G., A. von Haeseler, and K. Strimmer.** 2004. TREEFINDER: a powerful graphical analysis environment for molecular phylogenetics. BMC Evol Biol **4:**18.
23. **Jolley, K. A., E. J. Feil, M. S. Chan, and M. C. Maiden.** 2001. Sequence type analysis and recombinational tests (START). Bioinformatics **17:**1230-1.
24. **Jurado, V., P. Boiron, R. M. Kroppenstedt, F. Laurent, A. Couble, L. Laiz, H. P. Klenk, J. M. Gonzalez, C. Saiz-Jimenez, D. Mouniee, E. Bergeron, and V. Rodriguez-Nava.** 2008. Nocardia altamirensis sp. nov., isolated from Altamira cave, Cantabria, Spain. Int J Syst Evol Microbiol **58:**2210-4.
25. **Lalonde, M., H. Calvert, and S. Pine.** 1981. Isolation and use of *Frankia* strains in actinorhizae formation, p. 296-299. *In* W. N. AH Gibson (ed.), Current Perspectives in Nitrogen Fixation. Australian Academy of Science, Canberra.
26. **Lamb, C., and R. A. Dixon.** 1997. The Oxidative Burst in Plant Disease Resistance. Annu Rev Plant Physiol Plant Mol Biol **48:**251-275.
27. **Lechevalier, M.** 1986. Catalog of *Frankia* strains. The Actinomycetes **19:**131-162.
28. **Lumini, E., and M. Bosco.** 1999. Polymerase chain reaction - restriction fragment length polymorphisms for assessing and increasing biodiversity of *Frankia* culture collections. Can. J. Bot. **77:**1261-1269.
29. **Mirza, M., D. Hahn, and A. Akkermans.** 1992. Characterization of *Frankia* strains from *Coriaria nepalensis*. Syst Appl Microbiol **15:**289-295.
30. **Murry, M. M., M. S. Fontaine, and J. G. Torrey.** 1984. Growth kinetics and nitrogenase induction in Frankia sp. HFP ArI3 grown in batch culture. Plant and Soil **78:**61-78.
31. **Nazaret, S., B. Cournoyer, P. Normand, and P. Simonet.** 1991. Phylogenetic relationships among *Frankia* genomic species determined by use of amplified 16S rRNA sequences. J. Bacteriol. **173:**4072-4078.
32. **Nomoto, R., H. Kagawa, and T. Yoshida.** 2008. Partial sequencing of sodA gene and its application to identification of Streptococcus dysgalactiae subsp. dysgalactiae isolated from farmed fish. Lett Appl Microbiol **46:**95-100.
33. **Normand, P., and M. Lalonde.** 1982. Evaluation of *Frankia* strains isolated from provenances of two *Alnus* species. Can. J. Microbiol. **28:**1133-1142.
34. **Normand, P., and M. Lalonde.** 1986. The genetics of actinorhizal Frankia: a review. Plant Soil **90:**429-453.
35. **Normand, P., S. Orso, B. Cournoyer, P. Jeannin, C. Chapelon, J. Dawson, L. Evtushenko, and A. K. Misra.** 1996. Molecular phylogeny of the genus *Frankia* and related genera and emendation of the family Frankiaceae. Int. J. Syst. Bacteriol. **46:**1-9.
36. **Posada, D., and T. R. Buckley.** 2004. Model selection and model averaging in phylogenetics: advantages of akaike information criterion and bayesian approaches over likelihood ratio tests. Syst Biol **53:**793-808.

37. **Poyart, C., G. Quesnes, and P. Trieu-Cuot.** 2000. Sequencing the gene encoding manganese-dependent superoxide dismutase for rapid species identification of enterococci. J Clin Microbiol **38:**415-8.
38. **Prin, Y., L. Maggia, B. Picard, H. Diem, and P. Goullet.** 1991. Electrophoretic comparison of enzymes from 22 single-spore cultures obtained from *Frankia* strain ORS140102. FEMS Microbiol Lett **77:**223-228.
39. **Puppo, A., L. Dimitrijevic, H. Diem, and Y. Dommergues.** 1985. Homogeneity of superoxide dismutase patterns in *Frankia* strains from Casuarinaceae. FEMS Microbiol Lett **30:**43-46.
40. **Richier, S., P. Furla, A. Plantivaux, P. L. Merle, and D. Allemand.** 2005. Symbiosis-induced adaptation to oxidative stress. J Exp Biol **208:**277-85.
41. **Rodriguez-Nava, V., Z. U. Khan, G. Potter, R. M. Kroppenstedt, P. Boiron, and F. Laurent.** 2007. Nocardia coubleae sp. nov., isolated from oil-contaminated Kuwaiti soil. Int J Syst Evol Microbiol **57:**1482-6.
42. **Roggenkamp, A., T. Bittner, L. Leitritz, A. Sing, and J. Heesemann.** 1997. Contribution of the Mn-cofactored superoxide dismutase (SodA) to the virulence of Yersinia enterocolitica serotype O8. Infect Immun **65:**4705-10.
43. **Santos, R., D. Herouart, A. Puppo, and D. Touati.** 2000. Critical protective role of bacterial superoxide dismutase in rhizobium-legume symbiosis. Mol Microbiol **38:**750-9.
44. **Seib, K. L., H. J. Wu, S. P. Kidd, M. A. Apicella, M. P. Jennings, and A. G. McEwan.** 2006. Defenses against oxidative stress in Neisseria gonorrhoeae: a system tailored for a challenging environment. Microbiol Mol Biol Rev **70:**344-61.
45. **Simonet, P., A. Capellano, E. Navarro, R. Bardin, and A. Moiroud.** 1984. An improved method for lysis of *Frankia* with achromopeptidase allows detection of new plasmids. Can J Microbiol **30:**1292-1295.
46. **Simonet, P., P. Normand, A. Moiroud, and M. Lalonde.** 1985. Restriction enzyme digestion patterns of Frankia plasmids. Plant and Soil **87:**49-60.
47. **Simonet, P., L. N. Thi, A. Moiroud, and R. Bardin.** 1989. Diversity of *Frankia* strains isolated from a single alder stand. Plant Soil **118:**13-22.
48. **Steele, D., and M. Stowers.** 1986. Superoxide dismutase and catalase in *Frankia*. Can J Microbiol **32:**409-413.
49. **Strimmer, K., and A. Rambaut.** 2002. Inferring confidence sets of possibly misspecified gene trees. Proc Biol Sci **269:**137-42.
50. **Tavares, F., C. L. Santos, and A. Sellstedt.** 2007. Reactive oxygen species in legume and actinorhizal nitrogen-fixing symbioses: te mycrosymbiont's responses to un unfriendly reception. Physiol Plant **130:**344-356.
51. **Thompson, J. D., D. G. Higgins, and T. J. Gibson.** 1994. Clustal W - Improving the sensitivity of progressive multiple alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Research **22:**4673-4680.
52. **Tsolis, R. M., A. J. Baumler, and F. Heffron.** 1995. Role of Salmonella typhimurium Mn-superoxide dismutase (SodA) in protection against early killing by J774 macrophages. Infect Immun **63:**1739-44.
53. **Zhang, Z., M. F. Lopez, and J. G. Torrey.** 1984. A comparison of cultural characteristics and infectivity of *Frankia* isolates from root nodules of *Casuarina* species. Plant and Soil **78:**79-90.

CHAPTER V: GENERAL DISCUSSION

The genus *Frankia* encompasses filamentous diazotrophic actinobacteria, defined by their ability to induce N2-fixing root nodules on a broad range of angiosperms belonging to eight botanic families called "actinorhizal plants"(Benson & Silvester, 1993). Actinorhizal symbiosis is established between *Frankia* and about 200 species of plants that inhabit widespread terrestrial ecosystems all around the world. (Dawson, 2007a). The tightness of symbiotic association and the variety of inhabited ecosystems modeled actinorhizal symbiosis diversity as a mosaic of biogeographic patterns (Benson & Dawson, 2007a; Thompson, 1999). Thus *Frankia* diversity generation and maintenance are the resultants of the convergence of several factors i.e. host distribution, host specificity and environmental factors.

To identify the processes generating *Frankia* diversity and guiding its evolution, we must understand 1) the extent of *Frankia* diversity; it means the potential richness that is comprised within the genus and 2) the cohesive forces that assemble the genus. For the first point, is necessary to establish an efficient taxonomical classification and an accurate strain identification and for the second one, a reliable phylogeny depiction of the genus is necessary. Both elements are controversial in *Frankia* because of *Frankia* fastidious nature and the lack of consensus and homogeneity in the techniques before applied to study the genus diversity.

Impact of host plant on Frankia diversification and symbiosis biogeography

In this work, we followed the recommendations of the *ad hoc* committee for the re-evaluation of species definition in bacteriology (Stackebrandt *et al.*, 2002b) and thus we choose two of the approaches they proposed -a random genomic analysis called AFLP (Amplified Fragment Length Polymorphism) and a Multilocus Sequence Analysis for phylogenetic reconstruction (MLSA)- to analyze *Frankia*'s phylogeny and taxonomical classification. The addition of the two approaches efficiently covered the maximum of the phylogenetic signal (from major clusters to below of genomospecies level) within the *Frankia* genus. Besides, a large collection of about 100 strains representative of geographic, genetic and ecological genus diversity were included in the analysis, in addition to non-cultivable strains, particularly those harboring a Sp+ phenotype (sporulation *in planta*) which had never been extensively studied before.

The resulting phylogeny based in the overall work of this thesis (Fig. 1) confirmed the division of the genus into four major clades with the following statements:

1) The sister relationships of Clusters I and II corresponding respectively to the *Alnus-Myrica-Casuarina* and to the *Elaeagnaceae-Rhamnaceae* to strains infective

2) The cluster III, comprising non-infective and/or non effective strains, is the closest neighbor of the former.

3) The cluster IV corresponds to the non-cultivable strains and occupies the basal position of the phylogeny, likely representing the most ancient branching of the genus.

This genus structure and the relative correlation of *Frankia* clusters and the host infectivity groups, already reported, was here for the first time strongly confirmed thanks to the extensive set of strains analyzed and to the test performed (Benson & Dawson, 2007a; Clawson *et al.*, 2004a; Normand *et al.*, 1996). Thus, the topology displayed clearly demonstrates *Frankia* grouping on the basis of the host plant infectivity, allowing us to emit the conclusion that biogeographical distribution of the actinorhizal symbiosis exists and it might be the result of diverse evolutionary mechanisms i.e. host plant distribution, allopatry and coevolution events (Ramette & Tiedje, 2007).

Even if globally, numerous incongruence exist between phylogeny some examples show high impact of host plant on phylogeny at a low level (tips of the tree).

- The most striking example of host plant influence is the case of *Casuarina* infective strains. This group here renamed as genomospecies G4 contains typical *Casuarina* strains that are undistinguishable by MLSA analysis (See Chapter II). This group of strains seems to represent a recent lineage that emerged from the cluster I; the isolates there included are from diver geographic sites where *Casuarina* is growing as an introduced species (Simonet *et al.*, 1999). The high genomic homology of this strains (as proved by AFLP, MLSA and by the ecological marker *sodF*) and the strong specificity towards their host, let us suggest that the latter is involved in the dispersion and restricted diversity maintenance within this group.
- Aside, other cluster which dispersion and variability is subjected to host plant impact is the group of *Myrica pensylvanica* isolates belonging to *F. alni*. *Myrica* has been considered as a promiscuous host, since in greenhouse experiments they can be infected by *Frankia* strains belonging to different host-infectivity groups; however it demonstrated that in the nature some *Myrica*, like *M. gale* displayed specific

associations only with strains of *Alnus* group (Huguet *et al.*, 2001a). It seems to be the case for the three *Myrica pensylvanica* strains because they tightly grouped in a coherent group without the inclusion of other non *M. pensylvanica strains*.

- In the three studied *Alnus* species the Sp+ strains exhibited a remarkable intra-cluster genetic homogeneity and the phylogenetic structuration of these strains is strongly dependant of the host, with two very close groups corresponding to *A. incana* and *A. viridis* strains and a divergent group corresponding to *A. glutinosa*. The group of *A. glutinosa Frankia* Sp+ represents an early branching within the cluster I. In contrast, the Sp+ strains of *A. incana* and *A. viridis* exhibit the same characteristics of genetic homogeneity and high host affinity however they belong to *F. alni* from which they diverged as an early branching occurring within *F. alni*. The clear genetic separation between *A. incana* and *A. viridis* Sp+ seemed to have happened at a recent evolutionary scale since they displayed minimal differences.

In the cluster IV we also have evidence of host affinity guiding the diversity patterns of *Frankia*, since in this cluster were grouped a *Coriaria myrtifolia* and a *Coriaria ruscifolia* endophytes, coming from France and Mexico, respectively, displayed remarkable genomic similarity suggesting that host select this genotype in different geographical sites.

Impact of environmental factors

The impact of environmental factors on the evolution and diversification among the set of strains here employed, has not been systematically determined. The group most widely studied in this regard was the G6, when the impact of soil pH upon strain genetic variability was tested (Jamann *et al.*, 1992a). The conclusions of this work did give evidence that the low pH soils exerted a selective pressure and consequently diminishing the strains variability. With our data this group of strains clustered in a coherent group, suggesting low genetic variability.

Frankia clusters evolution

The phylogeny reported in this study lead us to wonder about the major cluster's ontogeny. The strains of basal cluster IV are still refractory to *in vitro* the isolation and thus they might exhibit an elevated dependence on their host. They displayed characteristics morphological

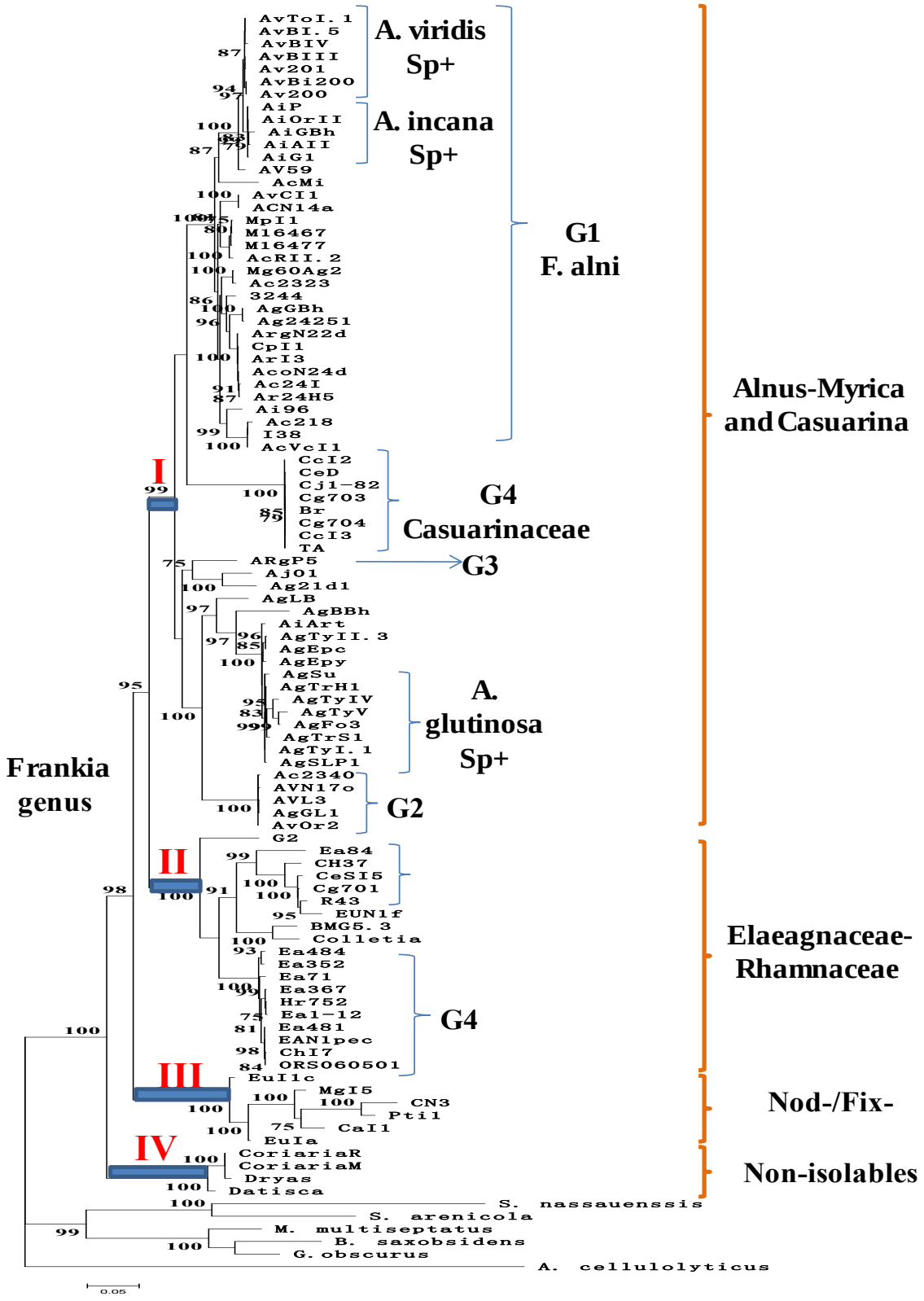

Fig. 1 Current phylogeny of the *Frankia* genus depicted by MLSA. Dendrogram constructed by Maximum Likelihood method with GTR+Γ as a model of substitution. The phylogeny is based on the concatenation of partial nucleotide sequences of five loci (*atpD*, *dnaA*, *ftsZ*, *pgk* and *rpoB*) retrieved from one hundred *Frankia* strains. Numbers at the branches indicate the LR-ELW values (1000 replicates).

features in the nodule, confirming the divergence with respect to other clusters. The phylogenetic topology might suggest this cluster was generated by the loss of saprophytic capacity. The last idea must be confirmed because we do not still know if strains of this cluster are capable to inhabit as saprophytes in the soil.

The other interesting question raises with respect to the cluster III ontogeny, because its members have loss the ability of nodulate and to fix nitrogen. Contrary to Rhizobia, *nod* genes have not been defined in *Frankia*, though highly homologous sequences have been found not in a plasmid a but in the genome making the their lateral transfer less likely. The loss of nodulation and nitrogen fixation seems to have marked the divergence of this cluster then. The marker *sodF* was not found within this group, there is strong evidence that this marker is implicated at the initial steps of symbiosis establishment. All this could indicate that the lost of nodulation capacity implied the loss of several portions of the genome.
Besides, it is worth to noting that these group exhibited long branching patterns, indicating that the evolutionary rate of this specimens might be different with respect to others groups.

Perspectives

This study has open several questions about *Frankia* diversification processes; therefore there is a list of issues that it would be interesting to manage in future studies.

Study of Frankia diversity in endemic and in insular sites

It has been reported that diversity of *Frankia* strains at the native habitat of the actinorhizal plants is higher. Thus, when *Casuarina* and *Allocasuarina* genus are introduced in foreign habitats do not nodulate unless *Frankia* is introduced with the host seedling and as demonstrated in the phylogeny here proposed these plants are infected by a unique genotype (Simonet *et al.*, 1999). Conversely, the symbionts of these plants at their native habitat exhibited considerable diversity. Thus more studies must be carried out at sites of origin of actinorhizal plants in order to unveil the diversity of their endemic compatible strain.
In the *Casuarinaceae* the *Gymnostoma* has a natural habitat restricted to the Malaysian-Australian region and it was found its compatible strains do not cluster with the other *Casuarinaceae*- typical strains but with the *Elaeagnaceae*-infective (Navarro *et al.*, 1997).

The correct placement of these *Frankia* strains with our MLSA approach would give insights into the evolutionary processes such strains are immersed.

Geographical isolated ecosystems as the insular places have also exhibited characteristics *Frankia* diversity patterns harboring new genotypes (Huguet *et al.*, 2005b). In this study the divergent clustering of the (Huguet *et al.*, 2005b) strain Ag21D1 isolated at the island of Corsica supported the idea that new genotypes and a unknown diversity may exists at such geographical isolated places. In this

Many species of *Alnus* genus were originated in the Asian continent, and others own as a native place insular sites as is the case of *A. cordata* (Corsica) in both cases the Auln species were then dispersed to a large variety of temperate ecosystems. It would be interesting to design the patterns of dispersion of such species by analyzing simultaneously the population of *Frankia* that infect them. The result expected would be a gradient of diversity that can lead us to understand the mechanisms of generation and maintenance of diversity at different special scales.

Sp+ strains as "cheaters"

Here we demonstrated that Sp+ formed divergent phylogenetic groups within the genus, moreover they exhibit remarkable genetic homogeneity and high host specificity. Besides, there is report that these strains exhibit also high infective power and they are capable to displace other *Frankia* populations inhabiting at the same stand. In contrast Sp+ strains

are refractive to the isolation as IV cluster strains so we can imagine that similar mechanisms are leading their evolution toward obligate symbiosis.

However Sp+ strains present some features that allow to different hypothesis as well. First, Sp+ strains exhibit a high infective power toward their host and an efficiently colonization, however it does not seem completely benefic for their partner, because it has been reported that Sp+ strains yield lower rates of nitrogen fixation compared with some Sp-, and so host nodulated by these strains do not displayed significant improvements in their growing.

The Sp+ strain behaviour could indicates that they are microorganisms seeking to obtain the maximum of fitness without reciprocating it host in the magnitude agreed for a symbiotic relationship. This behaviour has been also described in some Rhizobia species (Provorov & Vorobyov, 2008; Rankin *et al.*, 2007), the strains nodulent their host and then they seem to "forget" their obligations to contribute in an efficiently manner to the symbiosis sometimes

the host plant apply sanctions to this type of profiteer organisms. Thus, the Sp+ strains are likely a good model to study the "host-sanction" theory and to prove they actually act as cheaters (the term employed for those organisms). However we do not know how much this kind of studies will contribute to clarify the evolutionary relevance of the *in planta* sporulation trait.

CHAPTER VI: GENERAL BIBLIOGRAPHY[1]

[1] This list contains all the references which are not included in the precedents articles

Achtman, M. (2008). Evolution, population structure, and phylogeography of genetically monomorphic bacterial pathogens. *Annual review of microbiology* **62**, 53-70.

Achtman, M. & Wagner, M. (2008). Microbial diversity and the genetic nature of microbial species. *Nature reviews* **6**, 431-440.

Akimov, V. & Dobritsa, S. (1992a). Grouping of *Frankia* strains on the basis of DNA relatedness. *Systematic and applied microbiology* **15**, 372-379.

Akimov, V. N. & Dobritsa, S. V. (1992b). Grouping of *Frankia* strains on the basis of DNA relatedness. *Syst Appl Microbiol* **15**, 372-379.

Alskog, G. & Huss-Danell, K. (1997). SOD, catalase and N2ase activities of symbiotic Frankia (Alnus incana) in response to different oxygen tensions. *Physiol Plant* **99**, 286-292.

Alvarez-Matínez, E. R., Valverde, A., Ramírez-Bahena, M. E. & other authors (2009). The analysis of core and symbiotic genes of rhizobia nodulating Vicia from different continents reveals their common phylogenetic origin and suggests the distribution of Rhizobium leguminosarum strains together with Vicia seeds. *Archives of microbiology* **191**, 659-668.

An, C., Riggsby, W. & Mullin, B. (1985a). Relationships of *Frankia* isolates based on deoxyribonucleic acid homology studies. *International journal of systematic bacteriology* **35**, 140-146.

An, C. S., Riggsby, W. S. & Mullin, B. C. (1985b). Relationships of *Frankia* isolates based on deoxyribonucleic acid homology studies. *Int J Syst Bacteriol* **35**, 140-146.

Anderson, M. D., Ruess, R. W., Myrold, D. D. & Taylor, D. L. (2009). Host species and habitat affect nodulation by specific Frankia genotypes in two species of Alnus in interior Alaska. *Oecologia* **160**, 619-630.

Apel, K. & Hirt, H. (2004). Reactive oxygen species: metabolism, oxidative stress, and signal transduction. *Annual review of plant biology* **55**, 373-399.

Appleby, C. A. (1984). Leghemoglobin and Rhizobium Respiration. *Annu Rev Plant Physiol* **35**, 443-478.

Arahou, M., Diem, H. G. & Sasson, A. (1998). Influence of iron depletion on growth and production of catechol siderophores by different Frankia strains. *J Microbiol Biotechnol* **14**, 31-36.

Aronson, D. & Boyer, G. (1992). *Frankia* produces a hydroxamate siderophore under iron limitation. *J Plant Nutrition* **15**, 2193-2201.

Aronson, D. & Boyer, G. (1994). Growth and siderophore formation in six iron limited strains of *Frankia*. *Soil Biol Biochem* **26**, 561-567.

Asada, K., Kanematsu, S. & Uchida, K. (1977). Superoxide dismutases in photosynthetic organisms: absence of the cuprozinc enzyme in eukaryotic algae. *Archives of biochemistry and biophysics* **179**, 243-256.

Bachhawat, A. K. & Ghosh, S. (1987). Isolation and characterization of the outer membrane proteins of Azospirillum brasilense. *J Gen Microbiol* **133**, 1751-1758.

Bagnarol, E., Popovici, J., Alloisio, N., Maréchal, J., Pujic, P., Normand, P. & Fernandez, M. (2007). Differential Frankia protein patterns induced by phenolic extracts from Myricaceae seeds. *Physiol Plant*.

Baker, D. & Torrey, J. G. (1980). Characterization of an effective actinorhizal microsymbiont, *Frankia* sp. AvcI1 (Actinomycetales). *Canadian Journal of Microbiology* **26**, 1066-1071.

Baker, D. & Mullin, B. (1994). Diversity of *Frankia* nodule endophytes of the actinorhizal shrub *Ceanothus* as assessed by RFLP patterns from single nodule lobes. *Soil Biol Biochem* **26**, 547-552.

Barona-Gomez, F., Wong, U., Giannakopulos, A. E., Derrick, P. J. & Challis, G. L. (2004). Identification of a cluster of genes that directs desferrioxamine biosynthesis in Streptomyces coelicolor M145. *Journal of the American Chemical Society* **126**, 16282-16283.

Barton, L. L., Vester, C. R., Gill, P. R. & Neilands, J. B. (1994). The role of siderophores in symbiotic nitrogen-fixation by root nodule bacteria. In *The biochemistry of metal micronutrients in the rhizosphere*, pp. 54-66. Edited by J. A. Manthey, D. E. Crowley & D. G. Luster. Boca Raton, Fla: Lewis Publishers.

Barton, L. L., Johnson, G. V., Schitoskey, K. & Wertz, M. (1996). Siderophore-mediated iron metabolism in growth and nitrogen fixation by alfalfa nodulated with Rhizobium meliloti. *J Plant Nutr* **19**, 1201-1210.

Becking, J. (1970a). Plant-endophyte symbiosis in non-leguminous plants. *Plant Soil* **32**, 611-654.

Becking, J. (1974). Frankiaceae. In *Bergey's Manual of Determinative Bacteriology*, pp. 701-706. Edited by N. G. RE Buchanan. Baltimore: MD, Williams and Wilkins.

Becking, J. H. (1970b). Frankiaceae *fam. nov.* (Actinomycetales) with one new combination and six new species of the genus *Frankia* Brunchorst 1886, 174. *International journal of systematic bacteriology* **20**, 201-220.

Benson, D., Buchholz, S. & Hanna, D. (1984). Identification of *Frankia* strains by two-dimensional polyacrylamide gel electrophoresis. *Appl Environ Microbiol* **47**, 489-494.

Benson, D., Stephens, D., Clawson, M. & Silvester, W. (1996). Amplification of 16s rrna genes from Frankia strains in root nodules of Ceanothus griseus, Coriaria arborea, Coriaria plumosa, Discaria toumatou and Purshia tridentata. *Appl Environ Microbiol* **62**, 2904-2909.

Benson, D. R. & Silvester, W. B. (1993). Biology of *Frankia* strains, actinomycete symbionts of actinorhizal plants. *Microbiol Rev* **57**, 293-319.

Benson, D. R., Vanden Heuvel, B. D. & Potter, D. (2004). Actinorhizal Symbioses: Diversity and Biogeography. In *Plant Microbiology*, pp. 97-127. Edited by M. Gillings. Oxford: BIOS Scientific Publishers Ltd.

Benson, D. R. & Dawson, J. O. (2007a). Recent advances in the biogeography and genecology of symbiotic Frankia and its host plants. *Physiologia Plantarum* **130**, 318-330.

Benson, D. R. & Dawson, J. O. (2007b). Recent advances in the biogeography and genecology of symbiotic *Frankia* and its host plants. *Physiol Plant* **130**, 318-330.

Benson, H. P., Boncompagni, E. & Guerinot, M. L. (2005). An iron uptake operon required for proper nodule development in the Bradyrhizobium japonicum-soybean symbiosis. *Mol Plant Microbe Interact* **18**, 950-959.

Bentley, S. D., Chater, K. F., Cerdeno-Tarraga, A. M. & other authors (2002). Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2). *Nature* **417**, 141-147.

Bergsten, J. (2005). A review of long-branch attraction. *Cladistics* **21**, 163-193.

Berry, A. & Torrey, J. (1983). Root hair deformation in the infection process of *Alnus rubra*. *Can J Bot* **61**, 2863-2876.

Bishop, C. J., Aanensen, D. M., Jordan, G. E., Kilian, M., Hanage, W. P. & Spratt, B. G. (2009a). Assigning strains to bacterial species via the internet. *BMC Biol* **7**, 3.

Bishop, C. J., Aanensen, D. M., Jordan, G. E., Kilian, M., Hanage, W. P. & Spratt, B. G. (2009b). Assigning strains to bacterial species via the internet. *BMC Biol* **7**, 3.

Blears, M. J., De Grandis, S. A., Lee, H. & Trevors, J. T. (1998). Amplified Fragment Length polymorphism (AFLP): a review of the procedure and its applications. *Journal of Industrial Microbiology and Biotechnology* **21**, 99-114.

Bloom, R. A., Mullin, B. C. & Tate, R. L. (1989a). DNA restriction patterns and DNA-DNA solution hybridization studies of *Frankia* isolates from *Myrica pensylvanica* (bayberry). *App Environ Microbiol* **55**, 2155-2160.

Bloom, R. A., Mullin, B. C. & Tate, R. L. d. (1989b). DNA restriction patterns and DNA-DNA solution hybridization studies of Frankia isolates from Myrica pennsylvanica (bayberry). *Applied And Environmental Microbiology* **55**, 2155-2160.

Bolhuis, H., Severin, I., Confurius-Guns, V., Wollenzien, U. I. & Stal, L. J. Horizontal transfer of the nitrogen fixation gene cluster in the cyanobacterium Microcoleus chthonoplastes. *The ISME journal* **4**, 121-130.

Bontemps, C., Elliott, G. N., Simon, M. F. & other authors (2009). Burkholderia species are ancient symbionts of legumes. *Molecular ecology* **19**, 44-52.

Bosco, M., Fernandez, M., Simonet, P., Materassi, R. & Normand, P. (1992). Evidence that some *Frankia* sp. strains are able to cross boundaries between *Alnus* and *Elaeagnus* host specificity groups. *Appl Environ Microbiol* **58**, 1569-1576.

Boyer, G. L., Guillam, A. & Trick, C. G. (1987). Iron chelation and Uptake. In *Cyanobacteria* pp. 415-436. Edited by P. Fay & C. Van Baalen. London: Elsevier Biomedical.

Boyer, G. L. & Aronson, D. B. (1994). Iron Uptake and Siderophore Formation in the Actinorhizal Symbiont *Frankia*. In *Biochemistry of Metal Micronutrients in the Rhizosphere*, pp. 41-54. Edited by J. A. Manthey, Crowley, D.E., Luster D.G. Boca Raton: Lewis - CRC - Taylor & Francis.

Boyer, G. L., Kane, S. A., Alexander, J. A. & Aronson, D. B. (1999). Siderophore formation in iron-limited cultures of *Frankia* sp. strain 52065 and *Frankia* sp. strain CeSI5. *Can J Bot* **77**, 1316-1320.

Bruen, T. C., Philippe, H. & Bryant, D. (2006). A simple and robust statistical test for detecting the presence of recombination. *Genetics* **172**, 2665-2681.

Burggraaf, A. & Shipton, W. (1983). Studies on the growth of *Frankia* isolates in relation to infectivity and nitrogen fixation (acetylene reduction). *Can J Bot* **61**, 2774-2782.

Burleigh, S. & Dawson, J. (1994). Occurrence of *Myrica*-nodulating *Frankia* in Hawaiian volcanic soils. *Plant and Soil* **164**, 283-289.

Bussel, J. D., Waycott, M. & Chappill, J. A. (2005). Arbitrarily amplified DNA markers as characters for phylogenetic inference. *Perspect Plant Ecol Evol Syst* **7**, 3-26.

Callaham, D., Del Tredici, P. & Torrey, J. (1978). Isolation and cultivation *in vitro* of the actinomycete causing root nodulation in *Comptonia*. *Science* **199**, 899-902.

Caru, M. (1993). Characterization of native Frankia strains isolated from chilean shrubs (Rhamnaceae). *Plant and Soil* **157**, 137-145.

Challis, G. L. & Ravel, J. (2000). Coelichelin, a new peptide siderophore encoded by the Streptomyces coelicolor genome: structure prediction from the sequence of its non-ribosomal peptide synthetase. *FEMS microbiology letters* **187**, 111-114.

Challis, G. L. (2005). A widely distributed bacterial pathway for siderophore biosynthesis independent of nonribosomal peptide synthetases. *Chembiochem* **6**, 601-611.

Cilia, V., Lafay, B. & Christen, R. (1996). Sequence heterogeneities among 16S ribosomal RNA sequences, and their effect on phylogenetic analyses at the species level. *Molecular biology and evolution* **13**, 451-461.

Clawson, M. & Benson, D. (1999). Natural diversity of Frankia strains in actinorhizal root nodules from promiscuous hosts in the family Myricaceae. *Appl Environ Microbiol* **65**, 4521-4527.

Clawson, M., Gawronski, J. & Benson, D. R. (1999). Dominance of Frankia strains in stands of Alnus incana subsp. rugosa and Myrica pensylvanica. *Can J Bot* **77**, 1203-1207.

Clawson, M. L., Caru, M. & Benson, D. R. (1998). Diversity of frankia strains in root nodules of plants from the families elaeagnaceae and rhamnaceae. *Appl Environ Microbiol* **64**, 3539-3543.

Clawson, M. L., Bourret, A. & Benson, D. R. (2004a). Assessing the phylogeny of Frankia-actinorhizal plant nitrogen-fixing root nodule symbioses with Frankia 16S rRNA and glutamine synthetase gene sequences. *Molecular phylogenetics and evolution* **31**, 131-138.

Clawson, M. L., Bourret, A. & Benson, D. R. (2004b). Assessing the phylogeny of *Frankia*-actinorhizal plant nitrogen-fixing root nodule symbioses with *Frankia* 16S rRNA and glutamine synthetase gene sequences. *Mol Phylogenet Evol* **31**, 131-138.

Clerc, A., Manceau, C. & Nesme, X. (1998a). Comparison of Randomly Amplified Polymorphic DNA with Amplified Fragment Length Polymorphism To Assess Genetic Diversity and Genetic Relatedness within Genospecies III of Pseudomonas syringae. *Appl Environ Microbiol* **64**, 1180-1187.

Clerc, A., Manceau, C. & Nesme, X. (1998b). Comparison of Randomly Amplified Polymorphic DNA with Amplified Fragment Length Polymorphism To Assess Genetic Diversity and Genetic Relatedness within Genospecies III of *Pseudomonas syringae*. *Appl Environ Microbiol* **64**, 1180-1187.

Coenye, T., Schouls, L. M., Govan, J. R., Kersters, K. & Vandamme, P. (1999). Identification of *Burkholderia* species and genomovars from cystic fibrosis patients by AFLP fingerprinting. *Int J Syst Bacteriol* **49 Pt 4**, 1657-1666.

Coenye, T., Gevers, D., Van de Peer, Y., Vandamme, P. & Swings, J. (2005). Towards a prokaryotic genomic taxonomy. *FEMS microbiology reviews* **29**, 147-167.

Cohan, F. M. & Perry, E. B. (2007). A systematics for discovering the fundamental units of bacterial diversity. *Curr Biol* **17**, R373-386.

Cohan, F. M. & Koeppel, A. F. (2008). The origins of ecological diversity in prokaryotes. *Curr Biol* **18**, R1024-1034.

Costechareyre, D., Rhouma, A., Lavire, C., Portier, P., Chapulliot, D., Bertolla, F., Boubaker, A., Dessaux, Y. & Nesme, X. (2010). Rapid and Efficient Identification of Agrobacterium Species by recA Allele Analysis : Agrobacterium recA Diversity. *Microb Ecol*.

Cournoyer, B. & Lavire, C. (1999). Analysis of *Frankia* evolutionary radiation using glnII sequences. *FEMS microbiology letters* **177**, 29-34.

Cronquist, A. (1981). *An integrated system of classification of flowering plants.* New York: Columbia Univ. Press.

Crosa, J. H. & Walsh, C. T. (2002). Genetics and assembly line enzymology of siderophore biosynthesis in bacteria. *Microbiol Mol Biol Rev* **66**, 223-249.

Dawson, J. O. (2007a). Ecology of actinorhizal plants. *Nitrogen-fixing Actinorhizal Symbiosis In K Pawlowski and W E Newton (Eds)*, 199-227.

Dawson, J. O. (2007b). Ecology of actinorhizal plants. In *Nitrogen-fixing Actinorhizal Symbioses*, pp. 199-227. Edited by K. P. a. W. E. Newton. Dordrecht, The Netherlands: Springer.

De, M., Basu, M. & Chakrabartty, P. K. (2003). Increased synthesis of dihydroxybenzoic acid in the presence of aluminum by Rhizobium MO1. *Acta microbiologica Polonica* **52**, 195-199.

de Queiroz, K. (2005). Ernst Mayr and the modern concept of species. *Proceedings of the National Academy of Sciences of the United States of America* **102 Suppl 1**, 6600-6607.

Deletoile, A., Passet, V., Aires, J., Chambaud, I., Butel, M. J., Smokvina, T. & Brisse, S. Species delineation and clonal diversity in four Bifidobacterium species as revealed by multilocus sequencing. *Research in microbiology* **161**, 82-90.

Devulder, G., Perouse de Montclos, M. & Flandrois, J. P. (2005). A multigene approach to phylogenetic analysis using the genus Mycobacterium as a model. *International journal of systematic and evolutionary microbiology* **55**, 293-302.

Dewey, D. L. (1954). The distibution of diaminopimelic acid decarboxylase among some organisms of the coli-aerogenes group and certain other bacteria. *J Gen Microbiol* **11**, 307-312.

Dewey, D. L., Hoare, D. S. & Work, E. (1954). Diaminopimelic acid decarboxylase in cells and extracts of Escherichia coli and Aerobacter aerogenes. *Biochem J* **58**, 523-531.

Diem, H., Gauthier, D. & Dommergues, Y. (1982). Isolation of *Frankia* from nodules of *Casuarina equisetifolia*. *Can J Microbiol* **28**, 526-530.

Diem, H., Gauthier, D. & Dommergues, Y. (1983). An effective strain of *Frankia* from *Casuarina* sp. *Can J Bot* **61**, 2815-2821.

Dobritsa, S. V. (1998a). Grouping of *Frankia* strains on the basis of susceptibility to antibiotics, pigment production and host specificity. *Int J Syst Bacteriol* **48**, 1265–1275.

Dobritsa, S. V. (1998b). Grouping of *Frankia* strains on the basis of susceptibility to antibiotics, pigment production and host specificity. *International journal of systematic bacteriology* **48**, 1265–1275.

Dussurget, O., Rodriguez, M. & Smith, I. (1996). An ideR mutant of Mycobacterium smegmatis has derepressed siderophore production and an altered oxidative-stress response. *Molecular microbiology* **22**, 535-544.

Edwards, U., Rogall, T., Blocker, H., Emde, M. & Bottger, E. C. (1989). Isolation and direct complete nucleotide determination of entire genes. Characterization of a gene coding for 16S ribosomal RNA. *Nucleic acids research* **17**, 7843-7853.

Faten, G., Nouioui, I., Chair, M., Boudabous, A. & Gtari, M. (2010). 16S-23S rRNA Intergenic spacer region variability in the genus *Frankia*. *Microb Ecol* **In press**.

Fay, M. F., Cowan, R. S. & Leitch, I. J. (2005). The effects of nuclear DNA content (C-value) on the quality and utility of AFLP fingerprints. *Ann Bot* **95**, 237-246.

Feistner, G. J. & Beaman, B. L. (1987). Characterization of 2,3-dihydroxybenzoic acid from Nocardia asteroides GUH-2. *Journal of bacteriology* **169**, 3982-3987.

Fernandez, M., Meugnier, H., Grimont, P. & Bardin, R. (1989a). Deoxyribonucleic acid relatedness among members of the genus *Frankia*. *International journal of systematic bacteriology* **39**, 424-429.

Fernandez, M. P., Meugnier, H., Grimont, P. A. D. & Bardin, R. (1989b). Deoxyribonucleic acid relatedness among members of the genus *Frankia*. *Int J Syst Bacteriol* **39**, 424-429.

Fiedler, H. P., Krastel, P., Muller, J., Gebhardt, K. & Zeeck, A. (2001). Enterobactin: the characteristic catecholate siderophore of Enterobacteriaceae is produced by Streptomyces species.(1). *FEMS microbiology letters* **196**, 147-151.

Fraser, C., Alm, E. J., Polz, M. F., Spratt, B. G. & Hanage, W. P. (2010). The bacterial species challenge: Making sense of genetic and ecological diversity. *Science* **323**, 741-746.

Fridovich, I. (1995). Superoxide radical and superoxide dismutases. *Annual review of biochemistry* **64**, 97-112.

Garcia-Pereira, M. J., Caballero, A. & Quesada, H. (2010). Evaluating the relationship between evolutionary divergence and phylogenetic accuracy in AFLP data sets. *Mol Biol Evol* **27**, 988-1000.

Gardes, M., Bousquet, J. & Lalonde, M. (1987). Isozyme variation among 40 *Frankia* strains. *Appl Environ Microbiol* **53**, 1596-1603.

Gauthier, D., Frioni, L., Diem, H. & Dommergues, Y. (1984). The *Colletia spinosissima-Frankia* symbiosis. *Oecolog Plant* **5**, 231-239.

Gautier, A., Dubois, D., Escande, F., Avril, J., Trieu-Cuot, P. & Gaillot, O. (2005). Rapid and accurate identification of human isolates of Pasteurella and related species by sequencing the sodA gene. *Journal of clinical microbiology* **43**, 2307-2314.

Gevers, D., Vandepoele, K., Simillon, C. & Van de Peer, Y. (2004). Gene duplication and biased functional retention of paralogs in bacterial genomes. *Trends in microbiology* **12**, 148-154.

Gevers, D., Cohan, F. M., Lawrece, J. G. & authors., o. (2005a). Opinion: Re-evaluating prokaryotic species. *Nature reviews* **3**, 733-739.

Gevers, D., Cohan, F. M., Lawrence, J. G. & other authors (2005b). Opinion: Re-evaluating prokaryotic species. *Nat Rev Microbiol* **3**, 733-739.

Gill, P. R., Jr. & Neilands, J. B. (1989). Cloning a genomic region required for a high-affinity iron-uptake system in Rhizobium meliloti 1021. *Molecular microbiology* **3**, 1183-1189.

Glyan'ko, A. K. & Vasil'eva, G. G. (2010). Reactive oxygen and nitrogen species in Legume-Rhizobial Symbiosis: A review. *Applied Biochemistry and Microbiology* **46**, 15-22.

Gogarten, J. P. & Townsend, J. P. (2005). Horizontal gene transfer, genome innovation and evolution. *Nature reviews* **3**, 679-687.

Grantcharova, N., Lustig, U. & Flardh, K. (2005). Dynamics of FtsZ assembly during sporulation in Streptomyces coelicolor A3(2). *Journal of bacteriology* **187**, 3227-3237.

Griffiths, G. L., Sigel, S. P., Payne, S. M. & Neilands, J. B. (1984). Vibriobactin, a siderophore from Vibrio cholerae. *The Journal of biological chemistry* **259**, 383-385.

Gtari, M., Brusetti, L., Skander, G., Mora, D., Boudabous, A. & Daffonchio, D. (2004). Isolation of Elaeagnus-compatible Frankia from soils collected in Tunisia. *FEMS microbiology letters* **234**, 349-355.

Gtari, M., Brusetti, L., Hassen, A., Mora, D., Daffonchio, D. & Boudabous, A. (2007). Genetic diversity among *Elaeagnus* compatible *Frankia* strains and sympatric-related nitrogen-fixing actinobacteria revealed by *nifH* sequence analysis. *Soil Biol Biochem* **39**, 372-377.

Guerinot, M. L. (1991). Iron uptake and metabolism in the rhizobia/legume symbioses. *Plant and Soil* **130**, 199-209.

Guerinot, M. L. & Yi, Y. (1994). Iron: Nutritious, Noxious, and Not Readily Available. *Plant physiology* **104**, 815-820.

Gurtler, V. & Stanisich, V. A. (1996). New approaches to typing and identification of bacteria using the 16S-23S rDNA spacer region. *Microbiology (Reading, England)* **142**.

Gurtler, V. & Mayall, B. C. (2001). Genomic approaches to typing, taxonomy and evolution of bacterial isolates. *International journal of systematic and evolutionary microbiology* **51**, 3-16.

Hahn, D., Nickel, A. & Dawson, J. (1999). Assessing Frankia populations in plants and soil using molecular methods. *FEMS Microbiol Ecol* **29**, 215-227.

Hahn, D. (2007). Polyphasic taxonomy of the genus Frankia. *Nitrogen-fixing Actinorhizal Symbiosis In K Pawlowski and W E Newton (Eds)*, 25-47.

Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* **41**, 95-98.

Hanage, W. P., Fraser, C. & Spratt, B. G. (2006). Sequences, sequence clusters and bacterial species. *Philosophical transactions of the Royal Society of London* **361**, 1917-1927.

Hantke, K. (1981). Regulation of ferric iron transport in Escherichia coli K12: isolation of a constitutive mutant. *Mol Gen Genet* **182**, 288-292.

Harriott, O. T. & Bourret, A. (2003). Improving dispersed growth of Frankia using carbopol. *Plant and Soil* **254**, 69-74.

Higgins, D. G. & Sharp, P. M. (1988). CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. *Gene* **73**, 237-244.

Hohnadel, D. & Meyer, J. M. (1988). Specificity of pyoverdine-mediated iron uptake among fluorescent Pseudomonas strains. *Journal of bacteriology* **170**, 4865-4873.

Hosted, T., Rochefort, D. & Benson, D. (1993). Close linkage of genes encoding glutamine synthetase I and II in *Frankia alni* CpI1. *J Bacteriology* **175**, 3679-3684.

Hotchkiss, R. D. (1948). A microchemical reaction resulting in the staining of polysaccharide structures in fixed tissue preparations. *Archives of biochemistry* **16**, 131-141.

Hughet, V., MacCray Batzli, J., Zimpfer, J. F., Normand, P., Dawson, J. O. & Fernandez, M. P. (2001). Diversity and Specificity of Frankia Strains in Nodules of Sympatric Myrica gale, Alnus incana, and Shepherdia canadensis determined by rrs gene polymorphism. *Applied and Environmental Microbiology* **67**, 2116-2122.

Huguet, V., Batzli, J. M., Zimpfer, J. F., Normand, P., Dawson, J. O. & Fernandez, M. P. (2001a). Diversity and specificity of *Frankia* strains in nodules of sympatric *Myrica gale, Alnus incana*, and *Shepherdia canadensis* determined by rrs gene polymorphism. *Appl Environ Microbiol* **67**, 2116-2122.

Huguet, V., Batzli, J. M., Zimpfer, J. F., Normand, P., Dawson, J. O. & Fernandez, M. P. (2001b). Diversity and specificity of *Frankia* strains in nodules of sympatric *Myrica gale, Alnus incana*, and *Shepherdia canadensis* determined by rrs gene polymorphism. *Appl Environ Microbiol* **67**, 2116-2122.

Huguet, V., Mergeay, M., Cervantes, E. & Fernandez, M. P. (2004a). Diversity of *Frankia* strains associated to *Myrica gale* in Western Europe: impact of host plant (*Myrica* vs. *Alnus*) and of edaphic factors. *Environ Microbiol* **6**, 1032-1041.

Huguet, V., Mergeay, M., Cervantes, E. & Fernandez, M. P. (2004b). Diversity of *Frankia* strains associated to *Myrica gale* in Western Europe: impact of host plant (*Myrica* vs. *Alnus*) and of edaphic factors. *Environ Microbiol* **6**, 1032-1041.

Huguet, V., Gouy, M., Normand, P., Zimpfer, J. F. & Fernandez, M. P. (2005a). Molecular phylogeny of Myricaceae: a reexamination of host-symbiont specificity. *Molecular phylogenetics and evolution* **34**, 557-568.

Huguet, V., Land, E. O., Casanova, J. G., Zimpfer, J. F. & Fernandez, M. P. (2005b). Genetic diversity of *Frankia* microsymbionts from the relict species *Myrica faya* (Ait.) and *Myrica rivas-martinezii* (S.) in Canary Islands and Hawaii. *Microb Ecol* **49**, 617-625.

Hulton, C. S., Higgins, C. F. & Sharp, P. M. (1991). ERIC sequences: a novel family of repetitive elements in the genomes of Escherichia coli, Salmonella typhimurium and other enterobacteria. *Molecular microbiology* **5**, 825-834.

Huson, D. H. & Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. *Molecular biology and evolution* **23**, 254-267.

Huss-Danell, K. (1997). Actinorhizal symbioses and their N2-fixation. *New Phytol* **136**, 375-405.

Jamann, S., Fernandez, M. & Moiroud, A. (1992a). Genetic diversity of Elaeagnaceae-infective *Frankia* strains isolated from various soils. *Acta Oecolog* **13**, 395-405.

Jamann, S., Fernandez, M. P. & Moiroud, A. (1992b). Genetic diversity of Elaeagnaceae-infective *Frankia* strains isolated from various soils. *Acta Oecolog* **13**, 395-405.

Janssen, P., Coopman, R., Huys, G., Swings, J., Bleeker, M., Vos, P., Zabeau, M. & Kersters, K. (1996). Evaluation of the DNA fingerprinting method AFLP as a new tool in bacterial taxonomy. *Microbiology* **142**, 1881-1893.

Janssen, P., Maquelin, K., Coopman, R., Tjernberg, I., Bouvet, P., Kersters, K. & Dijkshoorn, L. (1997). Discrimination of *Acinetobacter* genomic species by AFLP fingerprinting. *Int J Syst Bacteriol* **47**, 1179-1187.

Jeong, S.-C. & Myrold, D. D. (1999). Genomic fingerprinting of *Frankia* microsymbionts from *Ceanothus* copopulations using repetitive sequences and polymerase chain reactions. *Can J Bot* **77**, 1220-1230.

Jeong, S., Ritchie, N. & Myrold, D. (1999). Molecular phylogenies of plants and Frankia support multiple origins of
actinorhizal symbioses. *Mol Phylogenet Evol* **13**, 493-503.

Jobb, G., von Haeseler, A. & Strimmer, K. (2004). TREEFINDER: a powerful graphical analysis environment for molecular phylogenetics. *BMC evolutionary biology* **4**, 18.

Jobidon, R. & Thibault, J. R. (1980). Optimal surface sterilization of Alnus crispa seeds with merthiolate. *FOR SCI* **26**, 483-484.

Jolley, K. A., Feil, E. J., Chan, M. S. & Maiden, M. C. (2001). Sequence type analysis and recombinational tests (START). *Bioinformatics (Oxford, England)* **17**, 1230-1231.

Jurado, V., Boiron, P., Kroppenstedt, R. M. & other authors (2008). Nocardia altamirensis sp. nov., isolated from Altamira cave, Cantabria, Spain. *International journal of systematic and evolutionary microbiology* **58**, 2210-2214.

Khandelwal, S. R., Manwar, A. V., Chaudhari, B. L. & Chincholkar, S. B. (2002). Siderophoregenic Bradyrhizobia boost yield of soybean. *Applied biochemistry and biotechnology* **102-103**, 155-168.

Kim, B. J., Lee, J. H., Lyu, M. A., Kim, S. J., Bai, G. H., Chae, G. T., Kim, E. C., Cha, C. Y. & Kook, Y. H. (1999). Identificación of mycobacterial species by comparative analysis of the RNA polymerase gene (rpoB). *Journal of clinical microbiology* **37**, 1714-1720.

Koopman, W. J. (2005). Phylogenetic signal in AFLP data sets. *Syst Biol* **54**, 197-217.

Krauss, S. L. (2000). Accurate gene diversity estimates from amplified length polymorphism markers. *Molecular ecology* **9**, 1241-1245.

Kurdali, F., Rinaudo, G., Moiroud, A. & Domenach, A. (1990). Competition for nodulation and $^{15}N_2$-fixation between a Sp^+ and a Sp^- Frankia strain from *Alnus incana*. *Soil Biol Biochem* **22**, 57-64.

Laguerre, G., Nour, S. M., Macheret, V., Sanjuan, J., Drouin, P. & Amarger, N. (2001). Classification of rhizobia based on nodC and nifH gene analysis reveals a close phylogenetic relationship among Phaseolus vulgaris symbionts. *Microbiology (Reading, England)* **147**, 981-993.

Laine, A. L. (2009). Role of coevolution in generating biological diversity: spatially divergent selection trajectories. *Journal of experimental botany* **60**, 2957-2970.

Lalonde, M., Calvert, H. & Pine, S. (1981). Isolation and use of *Frankia* strains in actinorhizae formation. In *Current Perspectives in Nitrogen Fixation*, pp. 296-299. Edited by W. N. AH Gibson. Canberra: Australian Academy of Science.

Lamb, C. & Dixon, R. A. (1997). The Oxidative Burst in Plant Disease Resistance. *Annual review of plant physiology and plant molecular biology* **48**, 251-275.

Langman, L., Young, I. G., Frost, G. E., Rosenberg, H. & Gibson, F. (1972). Enterochelin system of iron transport in Escherichia coli: mutations affecting ferric-enterochelin esterase. *Journal of bacteriology* **112**, 1142-1149.

Lawrence, J. G. (2002). Gene transfer in bacteria: speciation without species? *Theor Popul Biol* **61**, 449-460.

Lechevalier, M., Stern, A. & Lechevalier, H. (1981). Phospholipids in the taxonomy of actinomycetes. In *Actinomycetes Proceedings of the Fourth International Symposium on Actinomycete Biology*, pp. 111-116. Edited by G. P. KP Schaal. Stuttgart: Gustav Fischer Verlag.

Lechevalier, M. (1984a). The taxonomy of the genus *Frankia*. *Plant Soil* **78**, 1-6.

Lechevalier, M. (1986). Catalog of *Frankia* strains. *The Actinomycetes* **19**, 131-162.

Lechevalier, M. (1994). Taxonomy of the genus *Frankia* (Actinomycetales). *International journal of systematic bacteriology* **44**, 1-8.

Lechevalier, M. P. (1984b). The taxonomy of the genus *Frankia*. *Plant Soil* **78**, 1-6.

Librado, P. & Rozas, J. (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics (Oxford, England)* **25**, 1451-1452.

Lindstedt, B. A., Heir, E., Vardund, T. & Kapperud, G. (2000). Fluorescent amplified-fragment length polymorphism genotyping of Salmonella enterica subsp. enterica serovars and comparison with pulsed-field gel electrophoresis typing. *J Clin Microbiol* **38**, 1623-1627.

Liu, Q.-Q. & Berry, A. (1991). The infection process and nodule initiation in the *Frankia-Ceanothus* root nodule symbiosis: a structural and histochemical study. *Protoplasma* **163**, 82-92.

Ludwig, W. & Klenk, H. P. (2005). Overview: a phylogenetic backbone and taxonomic framework for prokaryotic systematic. In Garrity, G.M., Bean, D.J., Krieg, N.R., Staley, J.T. (Eds.), Bergey's manual of Systematic Bacteriology. The Proteobacteria. Part A Introductory essays. Springer. **2**, 49-65.

Lumini, E. & Bosco, M. (1996). PCR-restriction fragment length polymorphism identification and host range of single-spore isolates of the flexible *Frankia* sp. strain UFI 132715. *Appl Environ Microbiol* **62**, 3026-3029.

Lumini, E. & Bosco, M. (1999). Polymerase chain reaction - restriction fragment length polymorphisms for assessing and increasing biodiversity of *Frankia* culture collections. *Can J Bot* **77**, 1261-1269.

Lynch, D., O'Brien, J., Welch, T., Clarke, P., Cuiv, P. O., Crosa, J. H. & O'Connell, M. (2001). Genetic organization of the region encoding regulation, biosynthesis, and transport of rhizobactin 1021, a siderophore produced by Sinorhizobium meliloti. *Journal of bacteriology* **183**, 2576-2585.

Maggia, L. & Bousquet, J. (1994). Molecular phylogeny of the actinorhizal Hamamelidae and relationships with host promiscuity towards *Frankia*. *Molecular ecology* **3**, 459-467.

Marahiel, M. A., Stachelhaus, T. & Mootz, H. D. (1997). Modular Peptide Synthetases Involved in Nonribosomal Peptide Synthesis. *Chemical reviews* **97**, 2651-2674.

Marechal, J., Clement, B., Nalin, R., Gandon, C., Orso, S., Cvejic, J. H., Bruneteau, M., Berry, A. & Normand, P. (2000). A *rec*A gene phylogenetic analysis confirms the close proximity of *Frankia* to *Acidothermus*. *Int J Syst Evol Microbiol* **50**, 781-785.

Martens, M., Dawyndt, P., Coopman, R., Gillis, M., De Vos, P. & Willems, A. (2008). Advantages of multilocus sequence analysis for taxonomic studies: a case study using 10 housekeeping genes in the genus Ensifer (including former Sinorhizobium). *Int J Syst Evol Microbiol* **58**, 200-214.

Martin, B., Humbert, O., Camara, M. & other authors (1992). A highly conserved repeated DNA element located in the chromosome of Streptococcus pneumoniae. *Nucleic acids research* **20**, 3479-3483.

Maunuksela, L., Zepp, K., Koivula, T., Zeyer, J., Haahtela, K. & Hahn, D. (1999). Analysis of Frankia populations in three soils devoid of actinorhizal plants. *FEMS Microbiol Ecol* **28**, 11-21.

McCormick, J. R., Su, E. P., Driks, A. & Losick, R. (1994). Growth and viability of Streptomyces coelicolor mutant for the cell division gene ftsZ. *Molecular microbiology* **14**, 243-254.

Meudt, H. M. & Clarke, A. C. (2007a). Almost fogotten or latest practice? AFLP applications, analyses and advances. *Trends Plant Sci* **12**, 106-117.

Meudt, H. M. & Clarke, A. C. (2007b). Almost fogotten or latest practice? AFLP applications, analyses and advances. *Trends in plant science* **12**, 106-117.

Mierzwa, B., Wdowiak-Wrobel, S., Kalita, M., Gnat, S. & Malek, W. (2010). Insight into the evolutionary history of symbiotic genes of Robinia pseudoacacia rhizobia deriving from Poland and Japan. *Archives of microbiology* **192**, 341-350.

Milagres, A. M., Machuca, A. & Napoleao, D. (1999). Detection of siderophore production from several fungi and bacteria by a modification of chrome azurol S (CAS) agar plate assay. *Journal of microbiological methods* **37**, 1-6.

Mirza, B. S., Welsh, A., Rieder, J. P., Paschke, M. W. & Hahn, D. (2009a). Diversity of frankiae in soils from five continents. *Syst Appl Microbiol* **32**, 558-570.

Mirza, B. S., Welsh, A., Rieder, J. P., Paschke, M. W. & Hahn, D. (2009b). Diversity of frankiae in soils from five continents. *Systematic and applied microbiology* **32**, 558-570.

Mirza, M., Pawlowski, K., Hafeez, F., Chaudhary, A. & Akkermans, A. (1994). Ultrastructure of the endophyte and localization of *nifH* transcripts in root nodules of *Coriaria nepalensis* Wall by *in situ* hybridization. *The New phytologist* **126**, 131-136.

Monz, C. & Schwintzer, C. (1989). The physiology of spore-negative and spore-positive nodules of *Myrica gale*. *Plant Soil* **118**, 75-87.

Mortimer, P. & Arnold, C. (2001a). FAFLP: last word in microbial genotyping ? *J Med Microbiol* **50**, 393-395.

Mortimer, P. & Arnold, C. (2001b). FAFLP: last word in microbial genotyping ? *J Med Microbiol* **50**, 393-395.

Mougel, C., Thioulouse, J., Perriere, G. & Nesme, X. (2002a). A mathematical method for determining genome divergence and species delineation using AFLP. *International journal of systematic and evolutionary microbiology* **52**, 573-586.

Mougel, C., Thioulouse, J., Perrière, G. & Nesme, X. (2002b). A mathematical method for determining genome divergence and species delineation using AFLP. *International journal of systematic and evolutionary microbiology* **52**, 573-586.

Moulin, L., Bena, G., Boivin-Masson, C. & Stepkowski, T. (2004). Phylogenetic analyses of symbiotic nodulation genes support vertical and lateral gene co-transfer within the Bradyrhizobium genus. *Molecular phylogenetics and evolution* **30**, 720-732.

Murry, M., Fontaine, M. & Torrey, J. (1984a). Growth kinetics and nitrogenase induction in *Frankia* sp. HFPArI3 grown in batch culture. *Plant Soil* **78**, 61-78.

Murry, M., Konopka, A., Pratt, S. & Vandergon, T. (1997). The use of PCR-based typing methods to asses the diversity and phylogenic relationships of *Frankia* nodule endophytes of the actinorhizal shrub *Ceanothus*. *Physiologia Plantarum*.

Murry, M. A., Fontaine, M. S. & Torrey, J. G. (1984b). Growth kinetics and nitrogenase induction in *Frankia* sp. HFPArI3 grown in batch culture. *Plant Soil* **78**, 61-78.

Murry, M. A., Zhang, D., Schneider, M. & de Bruijn, F. J. (1995). Use of repetitive sequences and the polymerase chain reaction to fingerprint the genomes of *Frankia* isolates. *Symbiosis* **19**, 223-240.

Murry, M. M., Fontaine, M. S. & Torrey, J. G. (1984c). Growth kinetics and nitrogenase induction in Frankia sp. HFP ArI3 grown in batch culture. *Plant and Soil* **78**, 61-78.

Nakata, N., Tobe, T., Fukuda, I., Suzuki, T., Komatsu, K., Yoshikawa, M. & Sasakawa, C. (1993). The absence of a surface protease, OmpT, determines the intercellular spreading ability of Shigella: the relationship between the ompT and kcpA loci. *Molecular microbiology* **9**, 459-468.

Nalin, R., Domenach, A.-M. & Normand, P. (1995). Molecular structure of the *Frankia* spp. *nifD-nifK* intergenic spacer and design of Frankia genus compatible primer. *Mol Ecol* **4**, 483-491.

Navarro, E., Nalin, R., Gauthier, D. & Normand, P. (1997). The nodular microsymbionts of *Gymnostoma* spp. are *Elaeagnus*-infective strains. *Appl Environ Microbiol* **63**, 1610-1616.

Navarro, E., Bousquet, J., Moiroud, A., Munive, A., Piou, D. & Normand, P. (2003). Molecular Phylogeny of Alnus (Betulaceae), inferred from Nuclear Ribosomal DNA ITS sequences. *Plant & Soil* **00**, 000-000.

Nazaret, S., Cournoyer, B., Normand, P. & Simonet, P. (1991). Phylogenetic relationships among *Frankia* genomic species determined by use of amplified 16S rDNA sequences. *J Bacteriol* **173**, 4072-4078.

Neilands, J. B. (1981a). Microbial iron compounds. *Annual review of biochemistry* **50**, 715-731.

Neilands, J. B. (1981b). Iron absorption and transport in microorganisms. *Annual review of nutrition* **1**, 27-46.

Neilands, J. B. (1982). Microbial envelope proteins related to iron. *Annual review of microbiology* **36**, 285-309.

Nesme, X., Normand, P., Tremblay, F. & Lalonde, M. (1985). Nodulation speed of *Frankia* sp. on *Alnus glutinosa*, *Alnus crispa*, and *Myrica gale*. *Can J Bot* **63**, 1292-1295.

Newcomb, W., Callaham, D., Torrey, J. & Peterson, R. (1979). Morphogenesis and fine structure of the actinomycetous endophyte of nitrogen-fixing root nodules of *Comptonia peregrina*. *Bot Gaz* **140(S)**, S22-S34.

Nicholas, D. J. D. (1952). The use of fungy for determining trace metals in biological material. *Analyst* **77**, 629-641.

Nichols, R. (2001). Gene trees and species trees are not the same. *Trends Ecol Evol* **16**, 358-364.

Nittayajarn, A., Mullin, B. & Baker, D. (1990). Screening of symbiotic frankiae for host specificity using restriction fragment length polymorphism analysis. *Appl Environ Microbiol* **56**, 1172-1174.

Nomoto, R., Kagawa, H. & Yoshida, T. (2008). Partial sequencing of sodA gene and its application to identification of Streptococcus dysgalactiae subsp. dysgalactiae isolated from farmed fish. *Letters in applied microbiology* **46**, 95-100.

Normand, P. & Lalonde, M. (1982). Evaluation of *Frankia* strains isolated from provenances of two *Alnus* species. *Can J Microbiol* **28**, 1133-1142.

Normand, P. & Lalonde, M. (1986). The genetics of actinorhizal Frankia: a review. *Plant Soil* **90**, 429-453.

Normand, P., Gouy, M., Cournoyer, B. & Simonet, P. (1992). Nucleotide sequence of *nifD* from *Frankia alni* strain ArI3: Phylogenetical inferences. *Mol Biol Evol* **9**, 495-506.

Normand, P., Orso, S., Cournoyer, B., Jeannin, P., Chapelon, C., Dawson, J., Evtushenko, L. & Misra, A. K. (1996). Molecular phylogeny of the genus *Frankia* and related genera and emendation of the family Frankiaceae. *Int J Syst Bacteriol* **46**, 1-9.

Normand, P., Lapierre, P., Tisa, L. & other authors (2007a). Genome characteristics of facultatively symbiotic *Frankia* sp. strains reflect host range and host plant biogeography. *Genome Research* **17**, 7 - 15.

Normand, P., Lapierre, P., Tisa, L. & other authors (2007b). Genome characteristics of facultatively symbiotic *Frankia* sp. strains reflect host range and host plant biogeography. *Genome Res* **17**, 7 - 15.

Normand, P., Queiroux, C., Tisa, L., Benson, D., Cruveiller, S., Rouy, Z. & Medigue, C. (2007c). Exploring the genomes of Frankia sp. *Physiologia Plantarum* **13**, 331-343.

Ochman, H., Lawrence, J. G. & Groisman, E. A. (2000). Lateral gene transfer and the nature of bacterial innovation. *Nature* **405**, 299-304.

Ochman, H. & Moran, N. A. (2001). Genes lost and genes found: evolution of bacterial pathogenesis and symbiosis. *Science* **292**, 1096-1099.

Olive, D. M. & Bean, P. (1999). Principles and applications of methods for DNA-based typing of microbial organisms. *Journal of clinical microbiology* **37**, 1661-1669.

Ong, S. A., Peterson, T. & Neilands, J. B. (1979). Agrobactin, a siderophore from Agrobacterium tumefaciens. *The Journal of biological chemistry* **254**, 1860-1865.

Paschke, M., Dawson, J. & David, M. (1989). Soil nitrogen mineralization in plantations of *Juglans nigra* interplanted with actinorhizal *Elaeagnus umbellata* or *Alnus glutinosa*. *Plant Soil* **118**, 33-42.

Paschke, M. & Dawson, J. (1993). Avian dispersal of *Frankia*. *Canadian Journal of Botany* **71**, 1128-1131.

Pépin, R. & Boumendil, J. (1982). Préservation de l'ultrastructure du sclérote de Sclerotina tuberosa (champignon discomycètes); un modèle pour le préparation des échantillons imperméables et hétérogènes. *Cytologia* **47**, 359-377.

Pérez, N.-O., Olivera, H., Vásquez, L. & Valdés, M. (1999). Genetic characterization of Mexican Frankia strains nodulating Casuarina equisetifolia. *Can J Bot* **77**, 1214-1219.

Pompanon, F., Bonin, A., Bellemain, E. & Taberlet, P. (2005). Genotyping errors: causes, consequences and solutions. *Nat Rev Genet* **6**, 847-859.

Portier, P., Fischer-Le Saux, M., Mougel, C., Lerondelle, C., Chapulliot, D., Thioulouse, J. & Nesme, X. (2006a). Identificationof genomic species in *Agrobacterium* Biovar 1 by AFLP Genomic Markers. *Appl Environ Microbiol* **72**, 7123-7131.

Portier, P., Fischer-Le Saux, M., Mougel, C., Lerondelle, C., Chapulliot, D., Thioulouse, J. & Nesme, X. (2006b). Identificationof genomic species in *Agrobacterium* Biovar 1 by AFLP Genomic Markers. *Applied and Environmental Microbiology* **72**, 7123-7131.

Portier, P., Costechareyre, D., Muller, D., Chapulliot, D., Oger, C., Perriére, G., Guéguen, L., Daubin, V. & Nesme, X. (2008). Démasquage des gènes spécifiques d'une espèce génomique du complexe Agrobacterium tumefaciens par AFLP et multicapteur à ADN. *Les Actes du BRG* **7**, 351-371.

Posada, D. & Buckley, T. R. (2004). Model selection and model averaging in phylogenetics: advantages of akaike information criterion and bayesian approaches over likelihood ratio tests. *Systematic biology* **53**, 793-808.

Poyart, C., Quesnes, G. & Trieu-Cuot, P. (2000). Sequencing the gene encoding manganese-dependent superoxide dismutase for rapid species identification of enterococci. *Journal of clinical microbiology* **38**, 415-418.

Prin, Y., Maggia, L., Picard, B., Diem, H. & Goullet, P. (1991). Electrophoretic comparison of enzymes from 22 single-spore cultures obtained from *Frankia* strain ORS140102. *FEMS microbiology letters* **77**, 223-228.

Provorov, N. A. & Vorobyov, N. I. (2008). Equilibrium between the "genuine mutualists" and "symbiotic cheaters" in the bacterial population co-evolving with plants in a facultative symbiosis. *Theoretical population biology* **74**, 345-355.

Puppo, A., Dimitrijevic, L., Diem, H. & Dommergues, Y. (1985). Homogeneity of superoxide dismutase patterns in *Frankia* strains from Casuarinaceae. *FEMS microbiology letters* **30**, 43-46.

Rademaker, J. L. W., Hoste, B., Louws, F. J., Kersters, K., Swings, J., Vauterin, L., Vauterin, P. & De Bruijn, F. J. (2000a). Comparison of AFLP and rep-PCR genomic fingerprinting with DNA-DNA homology studies: *Xanthomonas* as a model system. *Int J Syst Evol Microbiol* **50**, 665-677.

Rademaker, J. L. W., Hoste, B., Louws, F. J., Kersters, K., Swings, J., Vauterin, L., Vauterin, P. & De Bruijn, F. J. (2000b). Comparison of AFLP and rep-PCR genomic fingerprinting with DNA-DNA homology studies: *Xanthomonas* as a model system. *International journal of systematic and evolutionary microbiology* **50**, 665-677.

Ramette, A. & Tiedje, J. M. (2007). Biogeography: an emerging cornerstone for understanding prokaryotic diversity, ecology, and evolution. *Microb Ecol* **53**, 197-207.

Rankin, D. J., Bargum, K. & Kokko, H. (2007). The tragedy of the commons in evolutionary biology. *Trends in ecology & evolution (Personal edition)* **22**, 643-651.

Richier, S., Furla, P., Plantivaux, A., Merle, P. L. & Allemand, D. (2005). Symbiosis-induced adaptation to oxidative stress. *The Journal of experimental biology* **208**, 277-285.

Rodriguez-Nava, V., Khan, Z. U., Potter, G., Kroppenstedt, R. M., Boiron, P. & Laurent, F. (2007). Nocardia coubleae sp. nov., isolated from oil-contaminated Kuwaiti soil. *International journal of systematic and evolutionary microbiology* **57**, 1482-1486.

Roggenkamp, A., Bittner, T., Leitritz, L., Sing, A. & Heesemann, J. (1997). Contribution of the Mn-cofactored superoxide dismutase (SodA) to the virulence of Yersinia enterocolitica serotype O8. *Infection and immunity* **65**, 4705-4710.

Rombauts, S., Van de Peer, Y. & Rouzé, P. (2003). AFLPinSilico, simulating AFLP fingerprints. *Bioinformatics* **19**, 776-777.

Rossello-Mora, R. & Amann, R. (2001). The species concept for prokaryotes *FEMS Microbiol Rev* **25**, 39-67.

Rouvier, C., Nazaret, S., Fernandez, M., Picard, B., Simonet, P. & Normand, P. (1992). rrn and nif intergenic spacers and isoenzyme patterns as tools to characterize Casuarina-infective Frankia strains. *Acta Oecologica* **13**, 487-495.

Rouvier, C., Prin, Y., Reddell, P., Normand, P. & Simonet, P. (1996). Genetic diversity among *Frankia* strains nodulating members of the family Casuarinaceae in Australia revealed by PCR and restriction length polymorphism analysis with crushed root nodules. *Appl Environ Microbiol* **62**, 979-985.

Saitou, N. & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular biology and evolution* **4**, 406-425.

Sangwan, I. & O'Brian M, R. (1992). Characterization of delta-Aminolevulinic Acid Formation in Soybean Root Nodules. *Plant physiology* **98**, 1074-1079.

Santos, R., Herouart, D., Puppo, A. & Touati, D. (2000). Critical protective role of bacterial superoxide dismutase in rhizobium-legume symbiosis. *Molecular microbiology* **38**, 750-759.

Savelkoul, P. H. M., Aarts, H. J. M., De Haas, J., Dijkshoorn, L., Duim, B., Otsen, M., Rademaker, J. L. W., Schouls, L. & Lenstra, J. A. (1999). Amplified-Fragment Length Polymorphism Analysis: the State of an Art. *Journal of clinical microbiology* **37**, 3083-3091.

Sawabe, T., Kita-Tsukamoto, K. & Thompson, F. L. (2007). Inferring the evolutionary history of vibrios by means of multilocus sequence analysis. *J Bacteriol* **189**, 7932-7936.

Schumpp, O. & Deakin, W. J. (2010). How inefficient rhizobia prolong their existence within nodules. *Trends in plant science* **15**, 189-195.

Schwintzer, C. (1990a). Spore-positive and spore-negative nodules. In *The Biology of Frankia and Actinorhizal Plants*, pp. 177-193. Edited by J. T. CR Schwintzer. New York, NY: Academic Press.

Schwintzer, C. R. (1990b). Spore-positive and spore-negative nodules. In *The biology of Frankia and actinorhizal plants*, pp. 177-193. Edited by C. R. a. T. Schwintzer, J.D. San Diego: Academic Press, Inc.

Schwyn, B. & Neilands, J. B. (1987). Universal chemical assay for the detection and determination of siderophores. *Analytical biochemistry* **160**, 47-56.

Seib, K. L., Wu, H. J., Kidd, S. P., Apicella, M. A., Jennings, M. P. & McEwan, A. G. (2006). Defenses against oxidative stress in Neisseria gonorrhoeae: a system tailored for a challenging environment. *Microbiol Mol Biol Rev* **70**, 344-361.

Shi, Y. & Ruan, J. (1992). DNA base composition and homology values in the classification of some *Frankia* strains. *Acta Microbiol Sin* **32**, 133-136.

Simonet, P., Capellano, A., Navarro, E., Bardin, R. & Moiroud, A. (1984). An improved method for lysis of *Frankia* with achromopeptidase allows detection of new plasmids. *Can J Microbiol* **30**, 1292-1295.

Simonet, P., Normand, P., Moiroud, A. & Lalonde, M. (1985). Restriction enzyme digestion patterns of Frankia plasmids. *Plant and Soil* **87**, 49-60.

Simonet, P., Thi, L. N., Moiroud, A. & Bardin, R. (1989). Diversity of *Frankia* strains isolated from a single alder stand. *Plant Soil* **118**, 13-22.

Simonet, P., Grosjean, M.-C., Misra, A., Nazaret, S., Cournoyer, B. & Normand, P. (1991). *Frankia* genus-specific characterization by polymerase chain reaction. *Appl Environ Microbiol* **57**, 3278-3286.

Simonet, P., Bosco, M., Chapelon, C., Moiroud, A. & Normand, P. (1994). Molecular characterization of *Frankia* microsymbionts from spore-positive and spore-negative nodules in a natural alder stand. *Appl Environ Microbiol* **60**, 1335-1341.

Simonet, P., Navarro, E., Rouvier, C. & other authors (1999). Co-evolution between *Frankia* populations and host plants in the family Casuarinaceae and consequent patterns of global dispersal. *Environ Microbiol* **1**, 525-533.

Simpson, F. B. & Neilands, J. B. (1976). Siderochromes in cyanophyaceae: isolation and characterization of schizokinen from Anabaena sp. *J Phycol* **12**, 44-48.

Soltis, D., Soltis, P., Morgan, D., Swensen, S., Mullin, B., Dowd, J. & Martin, P. (1995). Chloroplast gene sequence data suggest a single origin of the predisposition for symbiotic nitrogen fixation in angiosperms. *Proceedings National Academy Sciences USA* **92**, 2647-2651.

Soltis, D. E., Soltis, P. S., Chase, M. W., Mort, M. E., Albach, D. C. & et al. (2000). Anngiosperms Phylogeny inferred from 18S rDNA, rbcL, and atpB sequences. *Bot J Linn Soc* **133**.

St-Laurent, L., Bousquet, J., Simon, L. & Lalonde, M. (1987). Separation of various *Frankia* strains in the *Alnus* and *Elaeagnus* host specificity groups using sugar analysis. *Can J Microbiol* **33**, 764-772.

Stackebrandt, E. & Goebel, B. (1994). Taxonomic note: A place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. *International Journal Systematic Bacteriology* **44**, 846-849.

Stackebrandt, E., Frederiksen, W., Garrity, G. M. & other authors (2002a). Report of the ad hoc committee for the re-evaluation of the species definition in bacteriology. *Int J Syst Evol Microbiol* **52**, 1043-1047.

Stackebrandt, E., Frederiksen, W., Garrity, G. M. & other authors (2002b). Report of the ad hoc committee for the re-evaluation of the species definition in bacteriology. *International journal of systematic and evolutionary microbiology* **52**, 1043-1047.

Staley, J. T. (2006). The bacterial species dilemma and the genomic-phylogenetic species concept. *Philosophical transactions of the Royal Society of London* **361**, 1899-1909.

Steele, D. & Stowers, M. (1986). Superoxide dismutase and catalase in *Frankia*. *Can J Microbiol* **32**, 409-413.

Stern, M. J., Ames, G. F., Smith, N. H., Robinson, E. C. & Higgins, C. F. (1984). Repetitive extragenic palindromic sequences: a major component of the bacterial genome. *Cell* **37**, 1015-1026.

Stevens, J. B., Carter, R. A., Hussain, H., Carson, K. C., Dilworth, M. J. & Johnston, A. W. (1999). The fhu genes of Rhizobium leguminosarum, specifying siderophore uptake proteins: fhuDCB are adjacent to a pseudogene version of fhuA. *Microbiology (Reading, England)* **145 (Pt 3)**, 593-601.

Strimmer, K. & Rambaut, A. (2002). Inferring confidence sets of possibly misspecified gene trees. *Proceedings* **269**, 137-142.

Swensen, S. M. (1996). The evolution of actinorhizal symbioses: evidence for multiple origins of the symbiotic association. *AM J Bot* **83**, 1503-1512.

Swensen, S. M. & Mullin, B. C. (1997). Phylogenetic relationships among actinorhizal plants. The impact of molecular systematics and implications for the evolution of actinorhizal symbioses. *Physiol Plant* **99**, 565-573.

Tao, X. & Murphy, J. R. (1994). Determination of the minimal essential nucleotide sequence for diphtheria tox repressor binding by in vitro affinity selection. *Proceedings of the National Academy of Sciences of the United States of America* **91**, 9646-9650.

Tavares, F., Santos, C. L. & Sellstedt, A. (2007). Reactive oxygen species in legume and actinorhizal nitrogen-fixing symbioses: te mycrosymbiont's responses to un unfriendly reception. *Physiol Plant* **130**, 344-356.

Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). Clustal W - Improving the sensitivity of progressive multiple alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* **22**, 4673-4680.

Thompson, J. N. (1999). The evolution of species interactions. *Science* **284**, 2116-2118.

Torrey, J. (1987). Endophyte sporulation in root nodules of actinorhizal plants. *Physiol Plant* **70**, 279-288.

Traag, B. A. & van Wezel, G. P. (2008). The SsgA-like proteins in actinomycetes: small proteins up to a big task. *Antonie van Leeuwenhoek* **94**, 85-97.

Tsolis, R. M., Baumler, A. J. & Heffron, F. (1995). Role of Salmonella typhimurium Mn-superoxide dismutase (SodA) in protection against early killing by J774 macrophages. *Infection and immunity* **63**, 1739-1744.

Valdés, M. (2007a). *Frankia* Ecology. In *Nitrogen-fixing Actinorhizal Symbiosis*, pp. 49-71. Edited by K. P. a. W. E. Newton. Dordrecht, The Netherlands: Springer.

Valdés, M. (2007b). Frankia Ecology. *Nitrogen-fixing Actinorhizal Symbiosis In K Pawlowski and W E Newton (Eds)*, 49-71.

van den Braak, N., Simons, G., Gorkink, R. & other authors (2003). A new high-throughput AFLP approach for identification of new genetic polymorphism in the genome of the clonal microorganism Mycobacterium tuberculosis. *Journal of microbiological methods* **56**, 49-62.

van Dijk, C. & Merkus, E. (1976). A microscopical study of the development of a spore-like stage in the life cycle of the root nodule endophyte of *Alnus glutinosa* (L.) Gaertn. *The New phytologist* **77**, 73-91.

van Dijk, C. (1984). Ecological aspects of spore formation in *Frankia-Alnus* symbiosis: Leiden, The Netherlands, State University.

van Dijk, C., Sluimer, A. & Weber, A. (1988). Host range differentiation of spore-positive and spore-negative strain types of *Frankia* in stands of *Alnus glutinosa* and *Alnus incana* in Finland. *Physiol Plant* **72**, 349-358.

Vandamme, P., Pot, B., Gillis, M., de Vos, P., Kersters, K. & Swings, J. (1996). Polyphasic taxonomy, a consensus approach to bacterial systematics. *Microbiological reviews* **60**, 407-438.

VandenBosch, K. & Torrey, J. (1985). Development of endophytic *Frankia* sporangia in field- and laboratory-grown nodules of *Comptonia peregrina* and *Myrica gale*. *Amer J Bot* **72**, 99-100.

Vaneechoutte, M. (1996). DNA fingerprinting techniques for microorganisms. A proposal for classification and nomenclature. *Mol Biotechnol* **6**, 115-142.

Vinuesa, P., Silva, C., Lorite, M. J., Izaguirre-Mayoral, M. L., Bedmar, E. J. & Martinez-Romero, E. (2005). Molecular systematics of rhizobia based on maximum likelihood and Bayesian phylogenies inferred from rrs, atpD, recA and nifH sequences, and their use in the classification of Sesbania microsymbionts from Venezuelan wetlands. *Syst Appl Microbiol* **28**, 702-716.

Vitorino, L. R., Margos, G., Feil, E. J., Collares-Pereira, M., Ze-Ze, L. & Kurtenbach, K. (2008). Fine-scale phylogeographic structure of Borrelia lusitaniae revealed by multilocus sequence typing. *PloS one* **3**, e4002.

Vos, P., Hogers, R., Bleeker, M. & other authors (1995a). a new technique for DNA fingerprinting *Nucleic acids research* **23**, 4407-4414.

Vos, P., Hogers, R., Bleeker, M. & other authors (1995b). a new technique for DNA fingerprinting *Nucleic Acids Res* **23**, 4407-4414.

Vuylsteke, M., Peleman, J. D. & Van Eijk, M. J. (2007). AFLP technology for DNA fingerprinting. *Nature Protocols* **2**, 1387-1398.

Wayne, L. G., Brenner, D. J., Colwell, R. R. & other authors (1987). International Committee on Systematic Bacteriology. Report of the ad hoc committee on reconciliation of approaches to bacterial systematics. *International journal of systematic bacteriology* **37**, 463–464.

Welsh, A., Mirza, B. S., Rieder, J. P., Paschke, M. W. & Hahn, D. (2009a). Diversity of frankiae in root nodules of Morella pensylvanica grown in soils from five continents. *Systematic and applied microbiology* **32**, 201-210.

Welsh, A., Mirza, B. S., Rieder, J. P., Paschke, M. W. & Hahn, D. (2009b). Diversity of frankiae in root nodules of *Morella pensylvanica* grown in soils from five continents. *Syst Appl Microbiol* **32**, 201-210.

Welsh, A. K., Dawson, J. O., Gottfried, G. J. & Hahn, D. (2009c). Diversity of Frankia populations in root nodules of geographically isolated Arizona alder trees in central Arizona (United States). *Appl Environ Microbiol* **75**, 6913-6918.

Whitaker, R. J. (2006). Allopatric origins of microbial species. *Phil Trans R Soc B* **361**.

Winkelmann, G. e. (1991). *Handbook of Microbial Iron Chelates* Boca Raton, Floride: CRC Press.

Wu, D., Hugenholtz, P., Mavromatis, K. & other authors (2009). A phylogeny-driven genomic encyclopaedia of Bacteria and Archaea. *Nature* **462**, 1056-1060.

Yang, Z. (2006). *Computational molecular evolution*. Oxford: Oxford University Press.

Zehr, J. P., Jenkins, B. D., Short, S. M. & Steward, G. F. (2003). Nitrogenase gene diversity and microbial community structure: a cross-system comparison. *Environ Microbiol* **5**, 539-554.

Zeigler, D. R. (2003). Gene sequences useful for predicting relatedness of whole genomes in bacteria. *International journal of systematic and evolutionary microbiology* **53**, 1893-1900.

Zhang, W., Jayarao, B. M. & Knabel, S. (2004). Multi-virulence-locus-sequence-typing of Listeria Monocytogenes. *Applied and Environmental Microbiology* **70**, 913-920.

Zhang, Z., Lopez, M. F. & Torrey, J. G. (1984). A comparison of cultural characteristics and infectivity of *Frankia* isolates from root nodules of *Casuarina* species. *Plant and Soil* **78**, 79-90.

Zimpfer, J. F., Kennedy, G. J., Smyth, C. A., Hamelin, J., Navarro, E. & Dawson, J. O. (1999). *Casuarina*-infective *Frankia* occurs in restricted zones near planted *Casuarina cunninghamiana* trees in Jamaica. *Forest, Farmand community tree research reports* **4**, 17-27.