Systematic and Applied Microbiology 41 (2018) 311–323

Contents lists available at ScienceDirect

Systematic and Applied Microbiology

journal homepage: www.elsevier.de/syapm

Robust *Frankia* phylogeny, species delineation and intraspecies diversity based on Multi-Locus Sequence Analysis (MLSA) and Single-Locus Strain Typing (SLST) adapted to a large sample size

Adrien C. Pozzi[a,c,*], Hector H. Bautista-Guerrero[a,b], Sophie S. Abby[c,d], Aude Herrera-Belaroussi[a], Danis Abrouk[a], Philippe Normand[a], Frédéric Menu[c], Maria P. Fernandez[a]

[a] Université de Lyon, Université Lyon 1, CNRS, UMR 5557, INRA, UMR 1418, Laboratoire d'Ecologie Microbienne, 43 Boulevard du 11 novembre 1918, F-69622 Villeurbanne Cedex, France
[b] Laboratorio de Microbiologia Agricola, Escuela Nacional de Ciencias Biologicas, Instituto Politecnico Nacional, Prolongacion de Carpio y Plan de Ayala s/n, México D.F., Mexico
[c] Université de Lyon, Université Lyon 1, CNRS, UMR 5558, Laboratoire de Biométrie et Biologie Evolutive, 43 Boulevard du 11 novembre 1918, F-69622 Villeurbanne Cedex, France
[d] Universiteí Grenoble Alpes, CNRS, UMR 5525, VetAgroSup, INP Grenoble, Laboratoire Techniques de l'Ingeínierie Meídicale et de la Complexiteí - Informatique, Matheímatiques et Applications Grenoble (TIMC-IMAG), Domaine de la Merci, F-38706 La Tronche Cedex, France

ARTICLE INFO

Article history:
Received 12 September 2017
Received in revised form 8 March 2018
Accepted 10 March 2018

Keywords:
Frankia
MLSA phylogeny
Genomospecies
Taxonomy
Evolution
Diversity

ABSTRACT

Diazotrophic Actinobacteria of the genus *Frankia* represent a challenge to classical bacterial taxonomy as they include many unculturable strains. As a consequence, we still have a poor understanding of their diversity, evolution and biogeography. In this study, a Multi-Locus Sequence Analysis (MLSA) using *atpD*, *dnaA*, *ftsZ*, *pgk*, and *rpoB* loci was done on a large set of cultured and uncultured strains, compared to 16S rRNA and correlated to Average Nucleotide Identity (ANI) from available *Frankia* genomes. MLSA provided a robust resolution of *Frankia* genus phylogeny and clarified the status of unresolved species and complex of species.

The robustness of single-gene topologies and their congruence with the MLSA tree were tested. Lateral Gene Transfers (LGT) were few and scattered, suggesting they had no impact on the concatenate topology. The *pgk* marker – providing the longest sequence, highest mean genetic divergence and least occurrence of LGT – was used to survey an unequalled number of *Alnus*-infective *Frankia* — mainly uncultured strains from a broad range of host-species and geographic origins. This marker allowed reliable Single-Locus Strain Typing (SLST) below the species level, revealed an undiscovered taxonomical diversity, and highlighted the effect of cultivation, sporulation phenotype and host plant species on symbiont richness, diversity and phylogeny.

Introduction

Frankia sp. are nitrogen-fixing soil Actinobacteria. They establish root-nodule symbioses with a range of mostly woody Actinorhizal plants — distributed in eight families of pioneer dicots [12]. *Frankia* strains are widespread and able to thrive in diverse and extreme environments including areas devoid of compatible host plants [80] indicating their ability to durably survive both as endophytic symbionts and as saprobes. However, the non-culturability of numerous strains of *Frankia* makes difficult to study and to identify phenotypic bases for species description. The isolation of *Frankia* strains able to fulfill Koch's postulates [18] still cannot be achieved for several lineages, and circa 50% of *Frankia* strains remain uncultured to date [80], including all those sporulating *in-planta* (Sp+, spore-positive phenotype) [121]. As a consequence, the taxonomy of *Frankia* is still in its infancy with most putative species undescribed and several phylogenetic aspects unresolved.

Early attempts to cluster strains based on physiological markers were obviously possible only with easily culturable strains

* Corresponding author at: Department of Ecology, Evolution and Behavior, College of Biological Sciences, University of Minnesota-Twin Cities, 140 Gortner 6104A, 1479 Gortner Avenue St. Paul, MN 55108, USA.
E-mail addresses: adrien.c.pozzi@gmail.com, apozzi@umn.edu (A.C. Pozzi).

https://doi.org/10.1016/j.syapm.2018.03.002
0723-2020/

[5,11,37–38,62]. Genetic approaches based on PCR amplification of nodular DNA later allowed to overcome the barrier of *in-vitro* culturing [47,74,78,83,112]. Based on 16S rRNA, the *Frankia* genus appeared as a monophyletic group made of four main Clusters corresponding to host infection groups [83]. Yet, 16S rRNA-based phylogenies were shown to lack resolution and robustness within genus *Frankia* and to yield suboptimal topologies relative to other markers [108]. Other loci such as the protein coding genes *nifH, recA, glnA* and the Internal Transcribed Spacer (ITS) 16-23S rRNA provided a wealth of data and confirmed the four main 16S rRNA Clusters [25,39,57,67,77,112]. Nonetheless, single gene phylogenies are often not congruent with whole genome phylogenies due to recombination events and Lateral Gene Transfers (LGT) [79,122]. In *Frankiaceae*, this approach has often led to conflicting tree topologies regarding the relative position of the different Clusters [25,83,87] and the assignment of some strains to different Clusters.

At the division/class level, whole genome comparisons on a limited number of strains have permitted to estimate the number of LGT and to place *Frankia* at the root of the aerobic Actinobacteria clade [108]. At the species level, about 20 genomospecies have been reliably delineated within the *Frankia* genus by DNA-DNA hybridization (DDH) [2,14,34,109], considered a standard for bacterial species definition [125]. Less labor-intensive and less DNA-consuming but still based on whole genomes, Amplified Fragment Length Polymorphism (AFLP) was later proposed as an alternative method since it generates genomic distances that allow tree reconstruction [54–55,124] and are highly correlated to DDH [72,100], despite the fact AFLP-based tree reconstruction was shown to be biased by large differences in genome size between closely related strains [36,68], as reported for *Frankia* — 5–11 Mb [84,119–120]. Since the first paper reporting sequences of *Frankia* genomes [84] there has been a large increase in the number of genomes available, allowing the use of Average Nucleotide Identity (ANI) to explore *Frankia's* taxonomy [85–86,88–90,92,119–120]. Nevertheless, genome-wide approaches like sequencing, DDH or AFLP remain technically challenging when applied to uncultured endophytes and thus cannot be used to study the phylogeny of *Frankia*. Multi-Locus Sequence Analysis (MLSA) was recently proposed to describe the diversity of *Frankia* strains, revealing new phylogenetic clades and allowing to position uncultured strains [87,95] or to propose new species [85]. For Actinobacteria, MLSA was shown to be significantly better than other markers such as 16S [108]. Using a set of non-transferred or rarely transferred and rarely duplicated genes, ubiquitous within a taxon, it allows to buffer single-*locus* conflicting signals and the bias due to genome size differences. MLSA could thus represent a promising practical and portable approach to allow precise *Frankia* strain assignation to species, and to generate a robust genus phylogeny including uncultured strains.

We developed here a new MLSA scheme to address the need for a comprehensive phylogeny as a reference to delineate robust taxonomical units within the *Frankia* genus. We included reference cultured strains and uncultured endophytes from the four described Clusters, various actinorhizal plant genera and continents. The MLSA method was calibrated with previously described genomospecies (based on DDH and AFLP) and compared with 16S rRNA and genome (ANI) data. The occurrence of LGT was tested in order to assess the congruence of each single-gene phylogeny and the robustness of the phylogenetic signal, as well as to find a ubiquitous and reliable marker allowing accurate strain assignment to species. The relevance of such a rapid Single-Locus Strain Typing (SLST) was assessed on an unprecedented set of *Alnus*-infective (Cluster 1) strains — mostly uncultured, and whose *in-planta* sporulation phenotype was recorded. The SLST enabled studying intra-species diversity and the factors driving the evolution of these strains.

Materials and methods

Frankia *cultivation, nodule sampling, phenotypic characterization and DNA extraction*

To reconstruct the MLSA-based phylogeny of the *Frankia* genus, a total of 68 cultured *Frankia* strains and 7 uncultured endophytes from various host plants of far-ranging geographic origins were used (Table 1). The strains were grown at 28 °C in liquid F medium (Cluster 3 strains) [111], F medium containing Tween 80 (Cluster 1 *Alnus*-infective strains), or BAP medium (other strains) [3,73] with weekly manual shaking.

To deepen our understanding of Cluster 1 strain phylotaxonomy, a total of 577 root nodules and 38 isolated strains from 20 *Alnus* species (belonging to the 3 *Alnus* subgenera *Alnobetula, Alnus* and *Clethropsis*) and 3 Myricaceae species (*Comptonia peregrina, Myrica gale* and *Morella pensylvanica*) were included. In all, the 23 sampled host species covered Northern America, Western and Eastern Europe, Western and Eastern Asia (Supplementary information SI 1). All nodules were phenotyped for *in-planta* sporulation by microscopic observations of nodule hand sections stained with Lactophenol Blue (Réactifs RAL, Marcillac, France) as previously described [95].

Total DNA from pure cell cultures was extracted with the DNeasy Plant Mini kit (QIAGEN, Courtaboeuf, France) according to the manufacturer's instructions. Composite nodular DNA was extracted aseptically from a single lobe as previously described [95].

Gene loci *and primers*

In addition to the classical 16S rRNA gene, five independent *loci* that are present as single copy and spread throughout the genome of three sequenced reference strains [84] were studied. They represent different functional categories: energy production and ATP biosynthesis (*atpD*), chromosomal replication (*dnaA*), cell division (*ftsZ*), transcription and RNA metabolism (*rpoB*), and glycolysis (*pgk*). The same three reference genomic sequences [84] were used to design specific primers for the partial amplification of *atpD*, *dnaA*, *ftsZ*, and *pgk*, respectively (Supplementary information SI 2). We did not redesign primers as they were always specific to the genus *Frankia* and effective on each new genome published from that point. Primers used for *rpoB* amplification have been reported previously [58].

PCR amplification and sequencing

The five housekeeping *loci* were amplified by a simple PCR reaction in a final volume of 50 μL. Each PCR reaction contained 5 μL of template DNA (50–200 ng), 5 μL of 10× PCR buffer, 2.5 μL of each primer (10 mM), 5 μL of a dNTP mix (2 mM), 4 μL of MgCl2 (25 mM), 5 μL DMSO, 2.5 U of Taq DNA polymerase and 23 μL of sterile MilliQ water. The reaction conditions were: initial denaturation for 5 min at 94 °C, 30 cycles of denaturation for 1 min at 94 °C, annealing for 1 min at the primer-pair-specific annealing temperature (Supplementary information SI 2), and extension at 72 °C for 1 min. A final extension step was performed at 72 °C for 10 min.

16S rRNA genes were amplified with universal primers pA and pH [33]. Initial PCR amplification was achieved in a 100 μL mixture composed of 2 μL of template DNA (10 ng), 10 μL of 10× 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 MilliQ water. Amplification conditions were: initial denaturation for 5 min at 94 °C, 30 cycles of denaturation for 45 s at

Table 1
Host source and geographic origin of the 75 *Frankia* strains and actinorhizal nodules used in the Multi-Locus Sequence Analysis (MLSA) analysis.

Strain designation[a]	Source host	Geographical origin	Ref.
Ac2.18	*Alnus cordata*	Miribel, France	This study
Ac23.23	*Alnus cordata*	Saou, France	This study
Ac23.40	*Alnus cordata*	Saou, France	[34]
Ac24I.5	*Alnus cordata*	Orleans, France	This study
ACN14a	*Alnus alnobetula ssp. crispa*	Tadoussac, Canada	[82]
ACN1ag	*Alnus alnobetula ssp. crispa*	Tadoussac, Canada	[59]
AcoN24d	*Alnus cordata*	Orleans, France	[110]
AcVc1	*Alnus cordata*	Corsica, France	[34]
Ag21D1	*Alnus glutinosa*	Corsica, France	This study
Ag24.251	*Alnus glutinosa*	Orleans, France	[34]
Ai96.6	*Alnus incana ssp. incana*	Lacrans, France	This study
AirI1	*Alnus incana ssp. rugosa*	Vermont, USA	[61]
AJ01	*Alnus japonica*	Ibaraki, Japan	[130]
Allo2	*Allocasuarina verticillata*	Uruguay	[40]
Ar24H5	*Alnus rubra*	Orleans, France	[113]
ARgN22d	*Alnus incana ssp. rugosa*	Quebec, Canada	[81]
ARgP5	*Alnus incana ssp. rugosa*	Quebec, Canada	[82]
ArI3	*Alnus rubra*	Oregon, USA	[10]
Av200 nod	*Alnus alnobetula ssp. alnobetula*	Alps, France	[95]
Av201 nod	*Alnus alnobetula ssp. alnobetula*	Alps, France	[95]
Av59.7	*Alnus alnobetula ssp. alnobetula*	Alps, France	[95]
AvcI1	*Alnus alnobetula ssp. crispa*	Atikokan, Canada	[7]
AVL3	*Alnus alnobetula ssp. alnobetula*	Alps, France	[34]
AVN17o	*Alnus alnobetula ssp. alnobetula*	Alps, France	[34]
BCU110501	*Discaria trinevis*	Argentina	[20]
BMG5.1	*Coriaria japonica*	Japan	[43]
BMG5.12	*Elaeagnus angustifolia*	Tunisia	[42]
BMG5.23	*Casuarina glauca*	Tunisia	[39]
BMG5.3	*Elaeagnus angustifolia*	Tunisia	[42]
BR	*Casuarina equisetifolia*	Brazil	[98]
CaI1	*Ceanothus americanus*	Vermont, USA	[61]
CcI2	*Casuarina cunninghamiana*	Florida, USA	[132]
CcI3	*Casuarina cunninghamiana*	Florida, USA	[132]
CcI6	*Casuarina cunninghamiana*	Egypt	[65]
CeD	*Casuarina equisetifolia*	Senegal	[29]
CeSI5	*Casuarina equisetifolia*	USA	[17]
Cg70.1	*Casuarina glauca*	India	[87]
Cg70.3	*Casuarina glauca*	India	[8]
Cg70.4	*Casuarina glauca*	India	[8]
CH37	*Hippophaë rhamnoides*	Nogent sur Marne, France	[97]
ChI.7	*Colletia hystrix*	Chile	[19]
CjI-82	*Casuarina junghuniana*	Thailand	[29]
Cn3	*Coriaria nepalensis*	Pakistan	[70]
Colletia spinosissima nod	*Colletia spinosissima*	Argentina	This study
Coriaria myrtifolia nod	*Coriaria myrtifolia*	France	This study
Coriaria ruscifolia nod	*Coriaria ruscifolia*	Mexico	This study
CpI1	*Comptonia peregrina*	Massachusetts, USA	[18]
Dg1	*Datisca glomerata*	Pakistan	[92]
Dc12	*Datisca cannabina*	Pakistan	[50]
Dryas drumondii nod	*Dryas drummondii*	Gaspé, Canada	[16]
Ea1.12	*Elaeagnus angustifolia*	Ecully, France	[34]
Ea35.2	*Elaeagnus angustifolia*	Sutri, Italy	[53]
Ea36.7	*Elaeagnus angustifolia*	Ecully, France	[34]
Ea48.1	*Elaeagnus angustifolia*	St Etienne de Tinée, France	This study
Ea48.4	*Elaeagnus angustifolia*	St Etienne de Tinée, France	[53]
Ea7.1	*Elaeagnus angustifolia*	Toulon, France	[53]
Ea8.4	*Elaeagnus angustifolia*	Pont en Royan, France	[53]
EAN1pec	*Elaeagnus angustifolia*	Illinois, USA	[60]
EuI1a	*Elaeagnus umbellata*	Massachussetts, USA	[4]
EuI1c	*Elaeagnus umbellata*	Massachussetts, USA	[6]
EUN1f	*Elaeagnus umbellata*	Illinois, USA	[60]
G2 = DSM45899 = CECT9038	*Casuarina equisetifolia*	Guadeloupe, France	[30]
Hr75.2	*Hippophaë rhamnoides*	France	[8]
I.38	*Alnus incana ssp incana*	La Pallud, France	[8]
M16467	*Morella pensylvanica*	New Jersey, USA	[23]
M16477	*Morella pensylvanica*	New Jersey, USA	[8]
Mg60.2AG	*Alnus glutinosa*	Landes, France	[34]
MgI5	*Myrica gale*	New York, USA	[61]
MpI1	*Morella pensylvanica*	Massachussetts, USA	[62]
ORS060501	*Colletia spinosissima*	Argentina	[35]
PtI1	*Purshia tridentata*	Wyoming, USA	[4]
QA3	*Alnus nitida*	Pakistan	[46]
R43	*Casuarina equisetifolia*	Florida, USA	[61]
TA	*Allocasuarina torulosa*	Australia	[8]
Thr	*Casuarina cunninghamiana*	Egypt	[40]

[a] "nod" indicates nodular uncultured strains.

94 °C, annealing for 55 s at 55 °C, and extension for 1 min at 72 °C. A final extension step was performed for 5 min at 72 °C.

Double strand sequencing was performed by the GenoScreen company (Lille, France) for cultured strains and by the Biofidal-DTAMB facility (Villeurbanne, France) for root nodules, using the DyeDeoxy Terminator Cycle Sequencing Kit and an ABI Prism 3730 automated DNA sequencer (Thermo Fisher Scientific, Waltham, MA).

Sequence analysis

Blastococcus aggregatus, *Geodermatophilus obscurus*, *Modestobacter marinus*, *Salinispora arenicola,* and *Stackebrandtia nassauensis* were included as external groups, in accordance with previous studies [108,129]. The *atpD, dnaA, ftsZ, pgk, rpoB* and 16S rRNA outgroup sequences were extracted from complete genomes. Sequences obtained in this study were manually corrected with BioEdit [49]. The sequences for the five housekeeping *loci* of cultured strains were deposited in GenBank under accession numbers JN684910–JN684968 (*atpD*), JN684969–JN68502 (*dnaA*), JN685027–JN685083 (*ftsZ*), JN685084–JN685140 (*pgk*) and JN685141–JN685197 (*rpoB*). 16S rRNA sequences were mostly retrieved from the GenBank database and completed by additional sequences of representative strains of major groups deposited in GenBank under accession numbers JN685198–JN685211, JN698221 and JN698222 (Supplementary information SI 3). The *pgk* sequences obtained for nodular strains were deposited in EMBL under accession numbers LT599837–LT600328 and were pooled with previously published sequences [95–96,104].

After performing a codon-wise alignment including only *Frankia* sequences with the MACSE program [102] and manual curation of the codon sites, several statistics (number of alleles, of polymorphic sites, mean G + C content) were computed for each locus with dnaSP version 5.10.01 [64].

Phylogenetic analyses of Frankia *genus and delineation of taxonomic units*

For MLSA phylogenetic analyses, sequence alignments including outgroup and *Frankia* sequences of the 5 housekeeping *loci* were performed using Muscle 3.8.31 [32] with default parameters, and trimmed with the BMGE program version 1.1 [28] with default parameters and BLOSUM 60 similarity matrix. Alignment concatenation was performed using a home-made Python program. Maximum Likelihood trees were inferred for each of the genes and for the concatenated alignment with PhyML 3.0 [45] with a GTR + G4 model, and the "Best of NNI and SPR" option for topology search. Non-parametric branch support values were computed with the Shimodaira-Hasegawa-like approximated Likelihood Ratio Test statistic (SH-like aLRT) implemented in PhyML [44]. The outgroup branches were used to root the phylogenetic tree. The 16S rRNA phylogenetic analysis was conducted as described above for the Maximum Likelihood MLSA phylogeny, except that the sequences were neither trimmed nor concatenated.

A statistical analysis of topological differences between gene trees and the concatenated tree (MLSA tree) was conducted with the program Prunier version 2.0 [1]. Prunier identifies the clades responsible for significant topological conflicts and those discrepancies can be interpreted as the result of lateral gene transfers (LGT). LGT supports were defined by probability values of the conflicting branches produced by the transferred group misplacement in the gene tree. A topological discrepancy was considered significant if supported by a SH-like *p-value* above 95% (Prunier's conflict threshold).

Patristic pairwise distances between strains were computed from the MLSA phylogeny with the cophenetic.phylo function [ape package,91] and plotted in R software version 3.3.3 [99]. Hierarchical clustering was done with the hclust and cutree functions [stats package,99] to define molecular Operational Taxonomic Units (OTUs). The cutoff threshold for clustering matched the observed barcode gap at 0.03 substitution per nucleotide (subs/nt) on the patristic pairwise distance plot (data not shown).

Average Nucleotide Identity calculation from sequenced Frankia *genomes*

A total of 20 sequenced *Frankia* genomes – representing the totality of available genomes at the time of the study – were retrieved from the JGI IMGer website (https://img.jgi.doe.gov/cgi-bin/mer/main.cgi) [120] and used to compute the Average Nucleotide Identity (ANI) as previously described [41]. A 95% identity threshold has been proposed to consider two organisms to belong to the same species [41]. The correlation between pairwise MLSA distances and pairwise ANI of sequenced strains was assessed by the non-linear least square method [stats package,99]. As normality criterion was not validated (Shapiro–Wilk normality tests; p-values $<10^{-5}$) the correlation between the two variables was tested with the Spearman rank-order method [stats package,99].

Phylogenetic analysis of Alnus*-infective (Cluster 1) strains and OTU delineation*

The *pgk* phylogenetic analysis was conducted as described above for the Maximum Likelihood MLSA phylogeny, except that the sequences were not trimmed or concatenated. Clades were defined on the basis of the strain habitat (host–plant and/or region) and the *in-planta* sporulation phenotype, as previously proposed [95]. The abundance in each clade was scaled (divided by standard deviation) for each host-species and geographic origin prior to plotting against the *pgk* phylogeny. The percent of Spore-positive (Sp+) nodules was also computed for each clade.

OTU identification from the *pgk* topology was conducted as described above for the MLSA phylogeny, except that only Cluster 1 strains were considered and the threshold was slightly different to match the observed barcode gap at 0.04 subs/nt *pgk* distance.

Species diversity was assessed with the iNEXT package in R [27,52]. Sample diversity estimates and accumulation curves were computed for the 615 *Alnus*-infective strains, with Hill's numbers q = 0 (species richness), q = 1 (Shannon diversity, the exponential of Shannon entropy) and q = 2 (Simpson diversity, the inverse of Simpson concentration). The type of strain (cultured *vs.* nodular), the sporulation phenotype of nodules (Sp+ *vs.* Sp−) and the four most sampled *Alnus* species (*Alnus alnobetula, Alnus cordata, Alnus glutinosa* and *Alnus incana*) were investigated.

Results

Loci *characteristics*

The main features of the different *loci* are summarized in Table 2. A minimum of 73 *Frankia* strains under study (Table 1) provided sequences for each of the five housekeeping *loci*. For *atpD, dnaA, ftsZ* and *pgk,* the designed primers resulted in sequences covering more than 38% of the total gene length. In the case of *rpoB* gene, a region considered as phylogenetically discriminant and representing 11% of the gene length [58] was sequenced and analyzed. Mean G + C content values were within the range reported for the genus — 66–74%. The fraction of polymorphic sites in the alleles varied from 28.5% (*pgk*) to 51.3% (*ftsZ*). Mean genetic divergences ranged from 0.047 (*rpoB*) to 0.136 (*pgk*) subs/nt. For the 16S rRNA, the entire gene

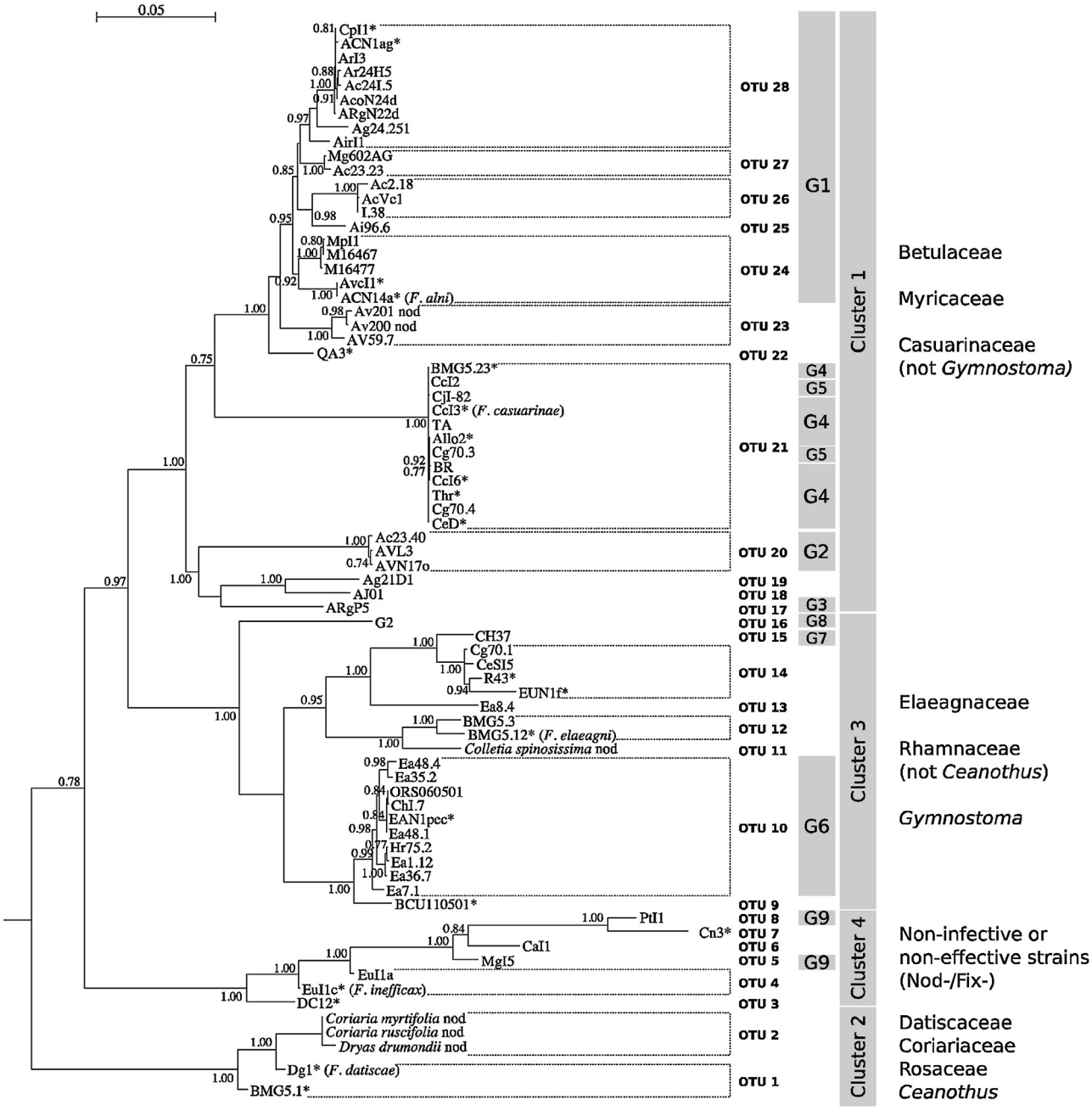

Fig. 1. Multi-Locus Sequence Analysis (MLSA) phylogenetic tree of the *Frankia* genus.
A Maximum Likelihood tree was inferred from the concatenation of five markers (*atpD, dnaA, ftsZ, pgk* and *rpoB*) into a super-alignment. Node reliability was assessed by SH-like aLRT statistic. For clarity, only branch support values higher than 0.70 are indicated and outgroup branches were not displayed. Genomospecies are named as previously [8]. Clusters and host infection groups are indicated as previously [83]. The scale represents the average number of substitutions per nucleotide. A 0.03 substitutions per nucleotide cutoff threshold was applied on the ultrametric MLSA tree to assign OTU (see experimental procedures). (*) Sequenced strains retrieved from GenBank.

Table 2
Sequence information of the five loci used in the Multi-Locus Sequence Analysis (MLSA) analysis compared to the 16S rDNA gene.

Gene	No. Samples	Fragment size (No. filtered sites, bp)	Portion of gene sequenced (%)	G + C Content (%)	No. Alleles	No. Polymorphic Sites (%)	Mean genetic divergence[a]
atpD	75	714 (701)	48	68.7	50	263 (36.8)	0.067
dnaA	74	629 (620)	38	66.5	46	288 (45.8)	0.091
ftsZ	73	717 (713)	43	69.7	45	368 (51.3)	0.096
pgk	73	702 (295)	57	73.7	43	200 (28.5)	0.136
rpoB	74	399 (398)	11	69.7	43	132 (33.1)	0.047
16S rRNA	40	1542 (72)	100	61.0	5	4 (5.5)	0.006

[a] Substitutions per nucleotide.

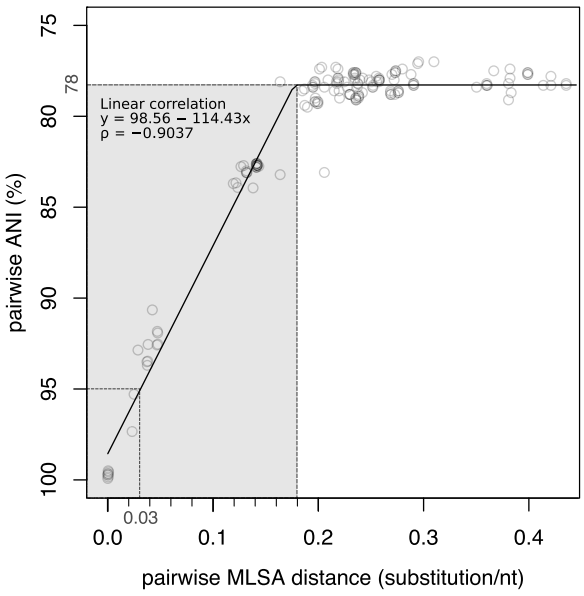

Fig. 2. Correlation between pairwise Multi-Locus Sequence Analysis (MLSA) distances and Average Nucleotide Identity (ANI).
For each couple of the *Frankia* strains sequenced available, pairwise ANI of whole genomes (%) were plotted against pairwise MLSA distances (number of substitution per nucleotide). The correlation between the two variables was assessed by the non-linear least square method implemented in R software. First, a linear response is observed, the two variables being negatively correlated (Spearman rank-order correlation; S = 39645; p-value $<10^{-15}$; $\rho = -0.9037$). Then, a plateau response is observed when the ANI goes below 78%. Note the 0.03 substitutions per nucleotide barcode gap used to define OTUs from the MLSA distances correlates with 95% ANI.

sequence was used, the number of polymorphic sites in the alleles was 5.5% and the mean genetic divergence was 0.006 subs/nt.

MLSA and 16 rRNA phylogenies of Frankia *genus*

The five housekeeping loci were genetically unlinked, thus providing independent markers. A super-alignment based on the concatenation of the five gene alignments was built (3161 bp), and served to reconstruct a maximum-likelihood MLSA phylogeny with Shimodaira–Hasegawa-like approximated Likelihood Ratio Test (SH-like aLRT) branch support (Fig. 1). The resulting phylogenetic tree confirmed the monophyly of *Frankiaceae* (aLRT = 1.0). The previously described Clusters 1–4 [83] were all highly supported monophyletic groups (all aLRT = 1.0), reflecting the host specificity groups. Cluster 1 contains *Alnus*, *Myricaceae* and *Casuarinaceae* infective strains; Cluster 2 groups *Rosaceae*, *Coriariaceae, Datiscaceae* and *Ceanothus* – mostly non-isolated endophytes; Cluster 3 groups *Elaeagnaceae, Rhamnaceae* and *Gymnostoma* infective strains; Cluster 4 comprises the non-nodulating or non-nitrogen-fixing (Nod-/Fix-) strains. Cluster 2 has a basal position in the phylogeny, and seems to be the most ancient group to have emerged (Fig. 1). In contrast, Cluster 4 was the most basal in the 16S topology, while Cluster 2 and Cluster 1 interspersed into groups with weaker branch support (aLRT = 0.72–0.87) and less correlation with host specificity (Supplementary information SI 4). Previously described AFLP-based genomospecies [8] were reliably differentiated as coherent, well-supported clades in the MLSA tree (Fig. 1), except for G4, G5 and G9.

A pairwise patristic distance matrix was derived from the MLSA tree. The MLSA distances were negatively correlated (Spearman rank-order correlation; S = 39645; p-value $<10^{-15}$; $\rho = -0.9037$) with the ANI of sequenced strains (Fig. 2). The hierarchical clustering at the 0.03 subs/nt MLSA distance threshold corresponding to 95% ANI allowed the delineation of 28 OTUs over the 75 strains. Twelve OTUs belonged to Cluster 1, two to Cluster 2, eight to Cluster 3 and six to Cluster 4.

Within Cluster 1, AFLP genomospecies G1 encompassed five well-supported OTUs. OTU 24 corresponded to *Frankia alni* [89] since it included ACN14a together with a closely related strain (AvcI1) and three *Myrica* strains. OTU 28 included CpI1 and ACN1ag and corresponded to another genomospecies based on DDH [34]. ANI results confirmed the grouping of CpI1 and ACN1ag, on the one hand, ACN14a and AvcI1, on the other, as well as the divergence of these two groups (Supplementary information SI 5). The separation of OTUs 26 and 27 from previous OTUs was supported by DDH since only 34 and 53% of DNA homology was found between AcoN24d and AcVc1 and Mg602AG, respectively [34]. OTU 23 included two non-isolated Sp+ *Frankia* strains from *A. alnobetula* and a cultured strain AV59.7, and diverged from other OTUs. OTU 22 only included the sequenced strain QA3 whose ANI with CpI1 (OTU 28) and ACN14a (OTU 24) was below the 95% threshold. Genomospecies G2 and G3 corresponded to OTUs 20 and 17, respectively. Genomospecies G4 and G5 were not separated and constituted the OTU 21, a highly-supported clade with very low genetic divergence corresponding to *Frankia casuarinae* [89] (Fig. 1).

Within Cluster 3, genomospecies G6 corresponded to the well supported and homogeneous OTU 10 (Fig. 1). OTUs 11 to 14 were sister clades of OTU 10 and were well supported (aLRT > 0.95). The separation between OTUs 9, 10, 12 and 14 was also supported (Supplementary information SI 5, ANI < 95%). OTU 12 grouped BMG5.3 and BMG5.12 and corresponded to *Frankia elaeagni* [89]. OTU 14 grouped the sequenced strains R43 and EUN1f (ANI > 95%) and two atypical *Elaeagnus*-infective strains isolated from *Casuarina*. Genomospecies G7 and G8 corresponded to OTU 15 and OTU 16, respectively.

Within Cluster 4, genomospecies G9 was split into two very divergent OTUs 5 and 8 corresponding to MgI5 and PtII, respectively (Fig. 1). While OTUs 5-8 formed a highly-supported group (aLRT = 1.00) comprising physiologically and genetically diverse strains they were not grouped into a single genomospecies as they showed high genetic divergence. OTU 4 grouped EuI1a and EuI1c strains, corresponding to *Frankia inefficax* [88]. OTU 3 corresponded to a single sequenced strain (DC12). ANI results supported the separation between OTUs 3 and 4 (ANI < 95%).

The basal Cluster 2 was split into two OTUs: OTU 1 grouped Dg1 and BMG5.1 strains (ANI > 97%) and corresponded to *Candidatus* F. datiscae [93]. OTU 2 corresponded to three uncultured strains (Fig. 1).

Statistical comparison of the five loci *gene trees with the MLSA phylogeny*

Single-gene topologies (Supplementary information SI 6) showed some differences between each other and with the concatenate MLSA tree (Fig. 1). In order to test whether these differences were statistically significant and indicated the potential results of LGT instead of methodological pitfalls, we used the program Prunier on gene tree topologies, considering the MLSA tree as a reference. Topological differences between a gene tree and the MLSA tree were considered significant for a SH-like aLRT statistic above 0.95 (Supplementary information SI 7). Major clades defined in the MLSA tree (Clusters and genomospecies, Fig. 1) were largely maintained in each single-gene tree (Supplementary information SI 7). Nevertheless, 14 supported topological differences were detected. Nine cases were categorized as “local movers”, meaning the group position was different by only a few branches, at a short evolutionary distance within a given genomospecies. Five cases corresponding to longer-range swaps provide clear evidence of LGT: *atpD*, *dnaA* and *rpoB* trees had respectively 1, 1 and 3 detected incongruences corresponding to such group swaps. In

contrast, the *ftsZ* and *pgk* phylogenetic trees did not display any significant discrepancy at the group level (Supplementary information SI 7). The *pgk*-tree topology was thus congruent with the MLSA phylogeny at the genomospecies level (Fig. 1 and Supplementary information SI 6A). This marker also was the longest and had the highest mean genetic divergence (Table 2) allowing maximal discrimination even between closely related strains. The *pgk* marker was thus chosen for a rapid and reliable strain characterization.

Phenotypic characterization and diversity of Alnus*-infective (Cluster 1)* Frankia

The 615 *Alnus*-infective strain sequences were obtained from 577 nodules sampled worldwide (20 countries) on 20 *Alnus* species or subspecies and 3 Myricaceae species, and 38 isolated reference strains (Supplementary information SI 1). Overall, 47% of strains were Sp+, 37% Sp− and the remaining 16% – mostly isolated reference strains – were not determined. Overall, 57 OTUs were defined from the *pgk* phylogeny at a 0.04 subs/nt *pgk* distance threshold: 25 OTUs were singletons, doubletons or tripletons, 30 had an abundance between 4 and 30, and two OTUs represented more than 20% of the whole set of strains. The observed species richness was estimated to a third of the extrapolated species richness (95% confidence interval: 1/10 to 2/3 of the total species richness). The species richness accumulation curve of all samples did not saturate while the Shannon and Simpson curves did (Fig. 3A), indicating a higher dominance of a few common OTUs. The extrapolated species richness of cultured strains was lower than the observed species richness of nodular strains while no significant differences were observed for the Shannon and Simpson diversity estimates (Fig. 3A). Regarding the effect of *in-planta* sporulation phenotype (Fig. 3B) and host-plant species (Fig. 3C), the species richness confidence intervals overlapped contrary to Shannon and Simpson diversity estimates.

pgk *phylogeny of* Alnus*-infective (Cluster 1)* Frankia

Following the MLSA tree, the *pgk* tree was rooted with the *Frankia* Cluster 2 (Fig. 4). Within Cluster 1, *Casuarina*-infective strains were basal and *Alnus*-infective strains were monophyletic. For reading convenience, outgroup Clusters 2–4 were not displayed and branches were collapsed into 42 clades containing strains with similar features – type of strain, host-plant species, geographic origin, *in-planta* sporulation phenotype – when possible. Branches were mostly well supported by SH-like aLRT. The cultured reference strains were absent from two thirds of the clades and were a small minority when present (MLSA OTUs 17–20 and 23) except in clade 2 that contained half of them (MLSA OTUs 22, 24–28) and grouped a variety of host-plant and geographic origins. All cultured strains belonged to strictly Sp− clades except Av40110 (MLSA OTU 23) and Ag21d1 (MLSA OTU 19) in clades 1a and 8b, respectively. Nodular strains were present in all clades except two (Fig. 4) and thus were far more genetically diverse than cultured strains (Fig. 3A). Uncultured Sp+ strains were distributed over 23 clades and were majority in half of them. The remaining 19 clades had only Sp−. The clustering of nodular strains tended to be more homogeneous with regards to the studied features. For instance, Sp+ *Frankia* from *A. alnobetula* were mainly in clades 1a (European origin) and 1a bis (American and Asian origins) while Sp− were in clade 4a together with cultured strains. *Frankia* from *A. cordata* were mainly in clade 8a–b and sometimes mixed with strains from *A. glutinosa* in clade 5a or with cultured strains in clade 2a–h. *Frankia* from *A. glutinosa* and *A. incana* mostly grouped in clades 5a and 1b, respectively.

Discussion

Frankia has resisted innumerable attempts at isolation and the subsequent taxonomical work resulted in very few species described so far and a sustained lack of clarity concerning the structure and the diversity of the genus. The 16S rRNA is one of the most used phylogenetic marker but the existence of recombination, duplications or lateral gene transfers, as in any other gene, indicates it is sometimes not perfectly depicting the evolution of species [131]. Moreover, among Actinobacteria, 16S signature patterns often provides unresolved or unreliable branching orders [66,108,129,133]. In our study, 16S-rRNA tree positioned Cluster 2 in the middle of Cluster 1 strains (Supplementary information SI 4). This is a major discrepancy with the reference tree we obtained (Fig. 1) and with previous studies. We thus confirm the non-adequacy of 16S rRNA for *Frankia* taxonomy and phylogeny.

Frankia *genus phylogeny*

We used a robust MLSA scheme on a large set of reference strains, including all genome-sequenced strains and type strains available to date. In spite of a few discrepancies, the statistical agreement among single gene trees (Supplementary information SI 6) and between these trees and the reference MLSA phylogeny (Fig. 1) was strong. Indeed, we used informational/housekeeping proteins encoded by single-copy genes which are less likely to be affected by LGT, contrary to metabolic proteins like NifH or GlnII. The concatenation of the five loci gathered a relevant phylogenetic signal and enabled the definition of robust Clusters, genomospecies and OTUs that have the status of species. The MLSA tree is in agreement with most previous taxonomical groupings. Classically, *Frankia* genus appears as monophyletic and divided into four major Clusters strongly linked to infection groups, confirming the importance of the host plant as a major factor driving *Frankia* evolution. Genomospecies defined by genome wide methodologies – DNA–DNA re-association values [2,14,34,109], evolutionary genomic divergence values (EGD) obtained by AFLP analysis [8], and ANIs (this study) – are conserved as strongly supported monophyletic groups, with the exception of G4 and G5. These two genomospecies grouped into a single OTU, suggesting the existence of a unique species.

The placement of Cluster 2 within the phylogeny of the *Frankia* genus is controversial as it depends on the marker used [25,39,83,87,128]. Our results strongly support the basal position of Cluster 2 within *Frankia* phylogeny, in agreement with that proposed based on the whole genome [108]. This group is likely to be the most ancient branching in *Frankia*'s evolutionary history and could have diverged 263–285 Myr bp, long before actinorhizal plants appearance in the fossil records [25] dated between 70 Myr [115] and 100 Myr [127]. The broad host range – encompassing four phylogenetically distant families [57,116] – of Cluster 2 strains could be a remnant of a symbiotic ancestor common to present-day lineages [117]. On the opposite, Clusters 1 and 3 comprise the most recently evolved groups of strains.

The Cluster 1 contains symbionts from three phylogenetically close families – Betulaceae, Casuarinaceae and Myricaceae [56,116] grouped into 11 OTUs – 10 for *Alnus*-infective strains and one for *Casuarina*-infective strains. Members of genomospecies G1 have been isolated from Betulaceae and Myricaceae, two wide ranging families [105], and are considered facultative symbionts or saprobes. Their identification in soils devoid of host plants suggests these strains are well adapted to saprophytic or to rhizospheric biotopes [24,69,114] and may have host-independent dispersion mechanisms. The well-studied G1 was commonly referred to as "*F. alni*" [8,34,83,89]. Actually, as already suggested, we confirmed here that G1 is a complex of species – five different OTUs with a low

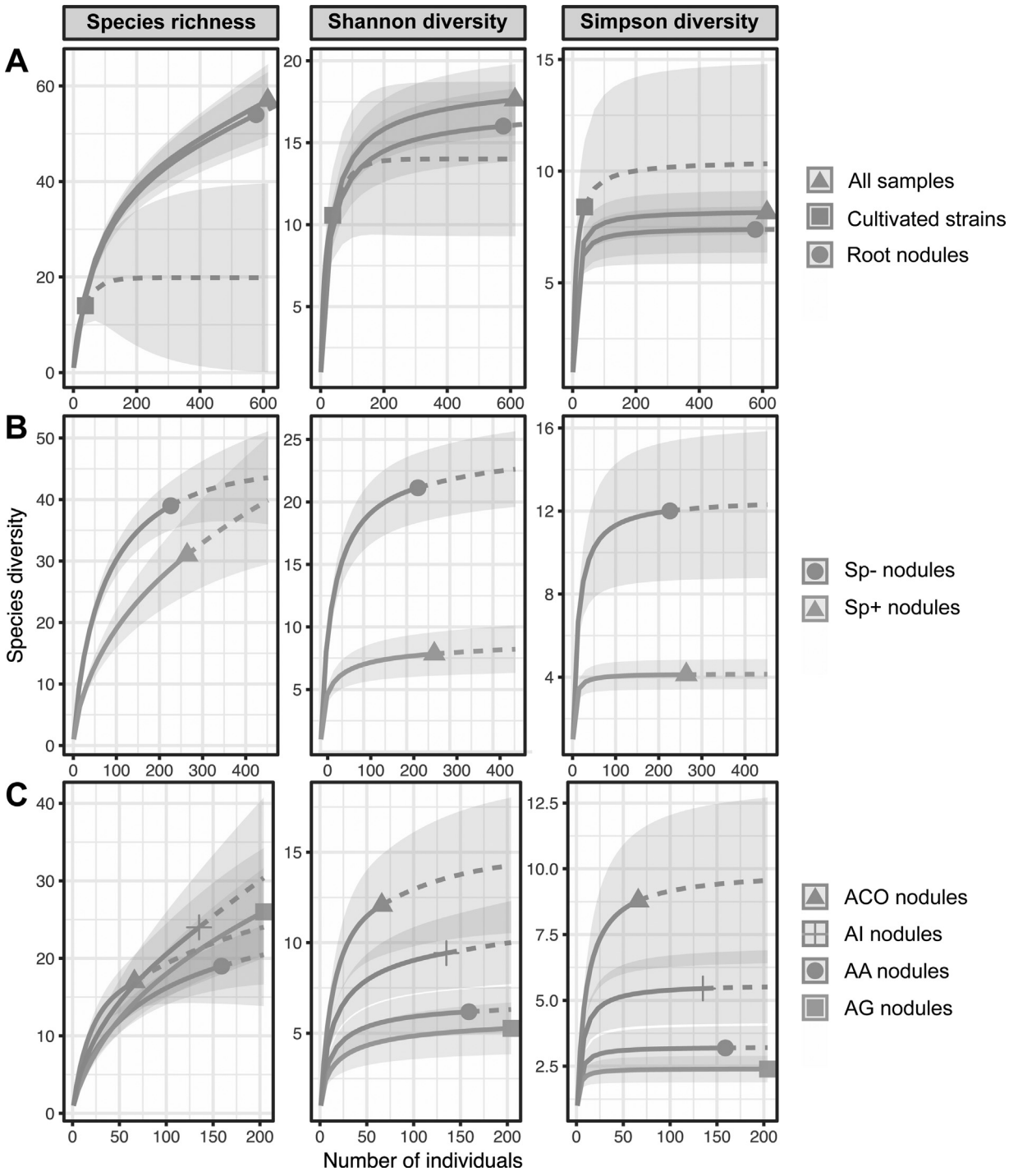

Fig. 3. Richness and diversity of *Alnus*-infective (Cluster 1) *Frankia* based on *pgk* marker.
Accumulation curves for Hill's number q = 0 (species richness), q = 1 (Shannon diversity) and q = 2 (Simpson diversity), with respect to (A) the sample type: all samples together and cultured *vs.* uncultured strains (root nodules), (B) the *in-planta* sporulation phenotype: Spore-positive (Sp+) *vs.* Spore-negative (Sp−) nodules, and (C) the host-plant species: ACO, *Alnus cordata*; AI, *A. incana*; AA, *A. alnobetula*; AG, *A. glutinosa*. Shaded area: 95% confidence interval. Plain lines: interpoled values. Dash lines: extrapoled values.

ANI. Recently, the name *"F. alni"* was attributed to ACN14a, defined as the species type strain [89]. Our results (Fig. 1 and Supplementary information SI 5) indicate that the *F. alni* species should be now restricted to OTU 24 – ACN14a, AvcI1 and three *Myrica*-infective strains – while OTUs 22, 23 and 25–28 should be considered as different species. All the *Casuarina*-infective strains studied here were isolated out of their area of origin (New-Zealand and Australia) essentially from a unique host (*Casuarina equisetifolia*). They belong to a unique species (*F. casuarinae*) [89], corresponding to OTU 21, whose relative position could not be fully resolved. Indeed, *Casuarinaceae* strains are a basal group of Cluster 1 or are interspersed with other Cluster 1 strains, depending on the method used [see for instance Refs. [8,25,87]] or the sampling size as deduced from our two *pgk* trees (Fig. 4 and Supplementary information SI 6A).

The Cluster 3 contains symbionts of Elaeagnaceae, Rhamnaceae and *Gymnostoma* sp. grouped into 8 OTUs. The Elaeagnaceae and Rhamnaceae host plants are also widely distributed; field studies have indicated that some strain subgroups were cosmopolitan

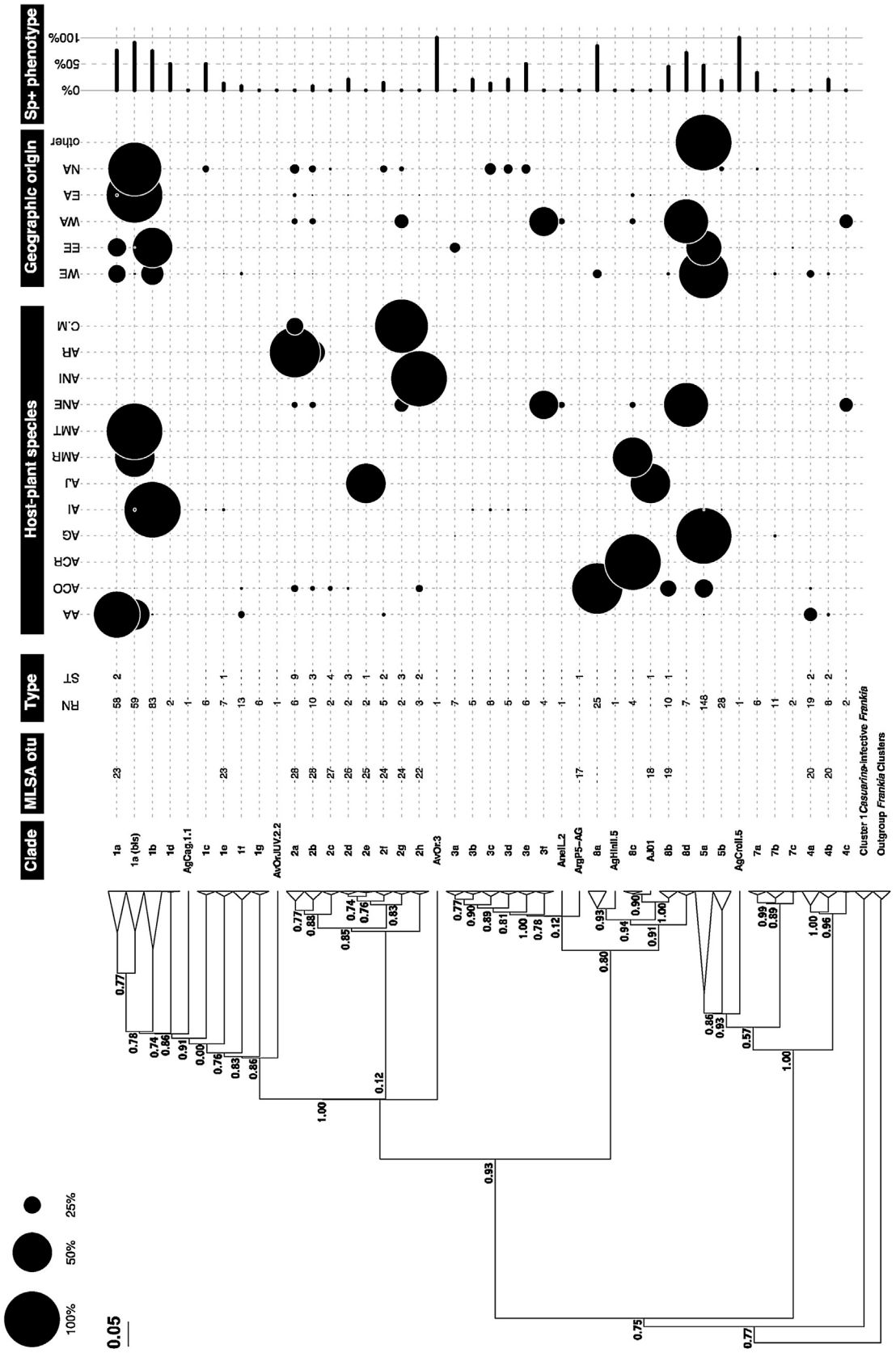

Fig. 4. *pgk* phylogenetic tree of *Alnus*-infective (Cluster 1) *Frankia*.
A Maximum Likelihood tree was inferred for cultured and uncultured strains. Node reliability was assessed by SH-like aLRT statistic. For clarity, outgroup branches were collapsed. Clade numbers are consistent with previous study [95]. Source (number of samples): RN, root nodule; ST, cultured strain. Host-plant species (scaled, per host): AA, *Alnus alnobetula* and subspecies; ACO, *A. cordata*; ACR, *A. cremastogyne*; AG, *A. glutinosa* and subspecies; AI, *A. incana* and subspecies; AJ, *A. jorullensis*; AMR, *A. maritima*; AMT, *A. matsumurae*; ANE, *A. nepalensis*; ANI, *A. nitida*; AR, *A. rubra*; C/M, *Comptonia* or *Myrica* species. Geographic origin (scaled, per origin): WE, Western Europe; EE, Eastern Europe; WA, Western Asia; EA, Eastern Asia; NA, Northern America. Sp+ (%, per clade): number of positive nodules for *in-planta* sporulation phenotype over total number of typed nodules. The scale represents the average number of substitutions per nucleotide.

[22,69,75,126] since introduced *Elaeagnus* are universally and rapidly nodulated by *Frankia* spp. of Cluster 3 [22,42]. Initially proposed for the then-unisolated symbionts of *Elaeagnus, Hippophae* and *Shepherdia* [9], the name *F. elaeagni* was recently proposed for type strain BMG5.12 [89]. In the reference tree, this strain grouped with another *Elaeagnus angustifolia* Tunisian strain (BMG5.3) leading to restrict *F. elaeagni* to OTU 12 (Fig. 1). Genomospecies G6 comprised genomically homogeneous ubiquitous strains with high saprophytic ability. They cross-infect the South-American Rhamnaceae (but not the North-American *Ceanothus*) and Elaeagnaceae [22]. OTU 14 grouped EUN1f strain with three *Elaeagnus*-infective atypical *Casuarina* strains (Cg70.1, CesI5 and R43), confirming the results obtained by MALDI-TOF analysis [48]. Interestingly, these four strains harbor an unusual host specificity. The three strains isolated from *Casuarina* strains are actually unable to nodulate their original host but can infect *Elaeagnaceae,* and EUN1f has the rare capacity for *Elaeagnaceae* strains to nodulate *Alnus* although with low nitrogen-fixation efficiency [15]. Genomic data will help clarifying the genomic processes that have driven the specialization of this group as well as their saprophytic ability or maintenance on alternate actinorhizal hosts. According to ANI, this group would include the G2 sequenced strain [89] (renamed 'G2g' in Supplementary information SI 3) but according to the divergence between G2 and OTU 14 in the MLSA tree, G2g was likely recovered from an isolate different from the genuine G2 [8].

The heterogeneous Cluster 4 grouped strains whose host plants belong to five divergent families. It comprises atypical strains unable to nodulate their original host – strains MgI5 [4] and EuI1c [6] – or strains that have lost their ability to nodulate (Nod−) and to fix nitrogen (Fix−) — strains Cn3 [70] and PtI1 [78]. It has been suggested that non-nodulating strains act as endophytes, coinfective strains or cosymbionts of effective *Frankia* strains without necessarily rendering a benefit to the host [83,101]. Since Cluster 4 strains are not at the root of the *Frankia* radiation, they would likely have emerged from the more ancient Cluster 2 and may have lost symbiotic functions secondarily.

As 26 out of 28 OTUs were monophyletic, we confirmed the validity of the 0.03 subs/nt and 95% ANI thresholds used. The two cases of observed paraphyly – OTUs 1 and 4 – were probably explained by the limited number of available strains and data for these groups.

Alnus-*infective (Cluster 1)* Frankia *phylogeny*

Although *Alnus* is the widest ranging actinorhizal genus [13,21], the diversity of *Alnus*-infective *Frankia* strains has been merely studied at a local or continental scale, encompassing a limited number of host species [51 and references therein,but see Ref. 94]. In the *MLSA* phylogeny of the *Frankia* genus, *Alnus*-infective strains represented 11 out of the 28 OTUs described. To offset the culture bias for *Alnus*-infective strains, our sampling included cultured and uncultured strains from 20 *Alnus* species or subspecies with a worldwide distribution – including areas known as diversity hot-spots (glacial refugia, island) and origin zones [Asia, 21] – as well as three Myricaceae species. The *pgk locus* produced the tree most congruent with the reference MLSA tree and allowed maximal discrimination between closely related strains, even below the species level corresponding to MLSA OTUs. The high amplification success achieved with the *pgk* marker (>95% on nodular strains) is a major improvement compared with another global-scale study — around 50% using *nifH* [94]. Together, these results make *pgk* very reliable for typing new *Frankia* strains and easily applicable to extensive diversity studies. Despite differences in molecular marker and sampled *Alnus* species, the magnitude of the estimated specific richness was similar to that of a previous study [94], with about 50 *Frankia* OTUs on 20 *Alnus* species. The *pgk* phylogeny and richness comparisons showed that strain isolation and cultivation select a limited range of genotypes and favors saprobes adapted to *in-vitro* conditions rather than more host-dependent symbionts [128]. Including numerous uncultured *Frankia* in the sampling revealed a hitherto undescribed diversity. Yet, it may still underestimate the true diversity *in-natura*, as demonstrated by studies on soil communities by qPCR [69,118] or NGS approaches [103]. However, these approaches do not necessarily allow to assign sequences to a specific taxon.

Factors driving Frankia *evolution*

The host-plant has long been considered the major factor driving *Frankia's* evolution [for a review, see Ref. [107]]. Even if *Frankia* clusters are well correlated with host families (Fig. 1) [83], several studies neither found phylogenetic congruence between *Frankia* strains [25,51,95] and *Alnus* species phylogenies [21,25,76] nor phylogenetic structuring of *Frankia* by the host plant species [51,94]. As previously proposed, we confirmed there is no clear evidence of co-speciation between *Alnus* and *Frankia* species. However, our results support a context-dependent co-evolution. For instance, *Frankia* strains associated to the different subspecies of *A. alnobetula* and *A. incana* from three continents, mostly grouped in clades 1a and 1b, respectively (Fig. 4). This supports the host-specificity observed in plant-trapping or cross-inoculation experiments [26 and references therein]. At the same time, we confirmed the possible influence of host biogeographical history on the symbiont phylogenetic structure [51,56,96]. For instance, *A. alnobetula* strains are divided into two groups, corresponding to European (1a) and North-American and East-Asian (1a bis) host-species (Fig. 4). This supports previous hypotheses that North-American *Alnus* species would have originated in Asia [21]. Host taxonomy and biogeography not only influence the symbiont phylogeny but also the symbiont diversity, as demonstrated here for the four most sampled species (Fig. 3). For instance, *A. cordata* thrives over a limited range known as a biodiversity hot-spot (Corsica and South Italy) and is associated with more diverse symbionts than *A. alnobetula* (5 subspecies from 3 continents). The relative influence of the host taxonomy and biogeography on *Frankia* phylogeny and diversity cannot be disentangled based on field studies. Controlled condition experiments may prove to be better but artificial inoculations and/or plant-trapping experiments often revealed higher symbiont diversity and wider host-range than observed in the field. Artificial conditions may force host-permissivity and/or strain specificity and lead to potential associations never achieved *in-natura*.

We recently showed that *in-planta* sporulation phenotype is a relevant trait for *Frankia* phylogenetic and diversity studies [95]. By increasing the sampling effort, we confirmed that (i) the Sp+ trait is not uniformly distributed in the phylogenetic tree, (ii) there are two Sp+ lineages hosted by the three species *A. alnobetula/A. incana* and *A. glutinosa* and (iii) the Sp+ strains have a lower diversity than the Sp−. We also identified a third Sp+ lineage hosted by *A. cordata*. Contrary to our previous study [95], it is worth noting that no pure Sp+ clades were reported in this study while several purely Sp− clades were. We cannot exclude a methodological bias that could underestimate the Sp+ phenotype. *In-planta* sporulation is under the genetic control of the strain [31,71,95,123] but its expression could be modulated by the environment [63,123]. Consequently, a Sp− phenotype could correspond either to a Sp− genotype or to a Sp+ genotype if *in-planta* sporulation was not expressed (false negative) in the given ecological conditions (host-plant and/or abiotic environment). The heterogeneous distribution of the Sp+ trait in the phylogeny raises the question of whether this trait was ancestral or

the result of convergent evolution. This cannot be easily deduced from our results. Even if the occurrence of this trait is higher in *Alnus* species, it has been reported for other actinorhizal plant families whose symbiont belong to Clusters 2 and 3 [106,121]. The phylogenetic position of these non-*Alnus* Sp+ strains is unknown, but like in Cluster 1, we expect them to form new lineages in their respective Cluster.

Conclusion

The MLSA approach used here generated a robust phylogeny of *Frankia* genus with enough discrimination power to clarify the status of unresolved species and complex of species. Together with previous studies and available ANI results, this study based on a high number of reference strains paves the way to a coherent species delineation and nomenclature in genus *Frankia*. This study also uncovered a taxonomical diversity with sound OTUs that correlated with ANI and constitute new species. A hidden diversity was also revealed among *Alnus*-infective (Cluster 1) strains using mostly uncultured symbionts from a wide range of host-species and geographic origins. The *pgk* marker clearly highlighted the effect of cultivation, *in-planta* sporulation phenotype and host-plant species on symbiont richness and diversity. The effects of the type of strain, host-plant species and geographic origins on *Frankia* evolution still have to be elucidated as the phylogenetic clustering cannot be explained by a single factor alone. In the future, the *pgk* marker could permit to accelerate the rate of discovery of new species and the provisional assignation of newly isolated or non-isolated strains to previously described species from all *Frankia* Clusters.

Acknowledgments

This study could never have existed without the immense and strenuous work of the entire community of *Frankia* researchers who have isolated and cultivated strains since 1978. This paper is dedicated to them. Particular thanks are expressed to Louis S. Tisa (University of Durham, New Hampshire) for granting access to IMGer genome data, and to Hiroyuki Tobita (Forestry and Forest Products Research Institute, Tsukuba, Japan), Monique Gardes and Mélanie Roy (University of Toulouse, France), Roger Ruess and Michaela Swanson (University of Fairbanks, Alaska) for the help in nodule samplings. We also thank Ramon Rossello-Mora and one anonymous reviewer for their comments that contributed to improve this paper. This work was supported by the National Center for Scientific Research (CNRS, grant EC2CO), the Mexico-France ECOS program and the research cluster FR41 BioEnviS. A.C. Pozzi was granted a doctoral fellowship by the Ministry of Higher Education and Research (France) and H. Bautista Guerrero had a PhD grant from CONACYT (Mexico). None of the authors have competing interests to declare.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.syapm.2018.03.002.

References

[1] Abby, S.S., Tannier, E., Gouy, M., Daubin, V. (2010) Detecting lateral gene transfers by statistical reconciliation of phylogenetic forests. BMC Bioinf. 11, 324.
[2] Akimov, V.N., Dobritsa, S.V. (1992) Grouping of *Frankia* strains on the basis of DNA relatedness. Syst. Appl. Microbiol. 15, 372–379.
[3] An, C., Riggsby, W., Mullin, B. (1985) Restriction pattern analysis of genomic DNA of *Frankia* isolates. Plant Soil 87, 43–48.
[4] Baker, D. (1987) Relationships among pure-cultured strains of *Frankia* based on host specificity. Physiol. Plant. 70, 245–248.
[5] Baker, D., Pengelly, W., Torrey, J. (1981) Immunochemical analysis of relationships among the isolated Frankiae (Actinomycetales). Int. J. Syst. Bacteriol. 31, 148–151.
[6] Baker, D., Newcomb, W., Torrey, J.G. (1980) Characterization of an ineffective actinorhizal microsymbiont, *Frankia* sp. EuI1 (Actinomycetales). Can. J. Microbiol. 26, 1072–1089.
[7] Baker, D., Torrey, J.G. (1980) Characterization of an effective actinorhizal microsymbiont, *Frankia* sp. AvcI1 (Actinomycetales). Can. J. Microbiol. 26, 1066–1071.
[8] Bautista-Guerrero, H.H., Cruz, H., Nesme, X., Valdés, M., Mendoza, H., Fernandez, M.P. (2011) Genomospecies identification and phylogenomic relevance of AFLP analysis of isolated and non-isolated strains of *Frankia* spp. Syst. Appl. Microbiol. 34, 200–206.
[9] Becking, J.H. (1970) *Frankiaceae* fam. nov. (Actinomycetales) with one new combination and six new species of the genus *Frankia* Brunchorst 1886, 174. Int. J. Syst. Evol. Microbiol. 20 (2), 201–220.
[10] Berry, A., Torrey, J.G. (1979) Isolation and characterization *in vivo* and *in vitro* of an actinomycetous endophyte from *Alnus rubra* Bong. In: Gordon, J.C., Wheeler, C.T., Perry, D.A. (Eds.), Symbiotic nitrogen fixation in the management of temperate forests, Forest Research Laboratory, Oregon State University, Corvallis, Oregon, pp. 69–83.
[11] Benson, D.R., Buchholz, S., Hanna, D. (1984) Identification of *Frankia* strains by two-dimensional polyacrylamide gel electrophoresis. Appl. Environ. Microbiol. 47, 489–494.
[12] Benson, D.R., Silvester, W.B. (1993) Biology of *Frankia* strains, actinomycete symbionts of actinorhizal plants. Microbiol. Rev. 57, 293–319.
[13] Benson, D.R., Dawson, J.O. (2007) Recent advances in the biogeography and genecology of symbiotic *Frankia* and its host plants. Physiol. Plant. 130, 318–330.
[14] Bloom, R.A., Mullin, B.C., Tate, R.L. (1989) DNA restriction patterns and DNA–DNA solution hybridization studies of *Frankia* isolates from *Myrica pensylvanica* (bayberry). Appl. Environ. Microbiol. 55, 2155–2160.
[15] Bosco, M., Fernandez, M.P., 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.
[16] Bosco, M., Lumini, E., Normand, P. (1996) PCR-RFLP direct fingerprinting of uncultured *Frankia* microsymbiont from *Dryas drummondii* nodules. Ann. Microbiol. Enzymol. 46, 115–123.
[17] 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.
[18] Callaham, D., Deltredici, P., Torrey, J.G. (1978) Isolation and cultivation *in vitro* of the actinomycete causing root nodulation in *Comptonia*. Science 199 (4331), 899–902.
[19] Caru, M. (1993) Characterization of native *Frankia* strains isolated from chilean shrubs (Rhamnaceae). Plant Soil 157, 137–145.
[20] Chaia, E. (1998) Isolation of an effective strain of *Frankia* from nodules of *Discaria trinervis* (Rhamnaceae). Plant Soil 205, 99–102.
[21] Chen, Z., Li, J. (2004) Phylogenetics and biogeography of *Alnus* (Betulaceae) inferred from sequences of nuclear ribosomal DNA ITS region. Int. J. Plant Sci. 165, 325–335.
[22] 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.
[23] 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.
[24] 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.
[25] Clawson, M.L., Bourret, A., Benson, D.R. (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, 131–138.
[26] Cotin-Galvan, L., Pozzi, A.C., Schwob, G., Fournier, P., Fernandez, M.P., Herrera-Belaroussi, A. (2016) *In-planta* sporulation capacity enhances infectivity and rhizospheric competitiveness of *Frankia* strains. Microbes Environ. 31 (1), 11–18, http://dx.doi.org/10.1264/jsme2.ME15090.
[27] Chao, A., Gotelli, N.J., Hsieh, T.C., Sander, E.L., Ma, K.H., Colwell, R.K., Ellison, A.M. (2014) Rarefaction and extrapolation with Hill numbers: a framework for sampling and estimation in species diversity studies. Ecol. Monogr. 84 (1), 45–67.
[28] Criscuolo, A., Gribaldo, S. (2010) BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. BMC Evol. Biol. 10, 210.
[29] Diem, H., Dommergues, Y. (1983) The isolation of *Frankia* from nodules of *Casuarina*. Can. J. Bot. 61, 2822–2825.
[30] Diem, H., Gauthier, D., Dommergues, Y. (1982) Isolation of *Frankia* from nodules of *Casuarina equisetifolia*. Can. J. Microbiol. 28, 526–530.
[31] van Dijk, C. (1978) Spore formation and endophyte diversity in root nodules of *Alnus glutinosa* (L.) Vill. New Phytol. 81, 601–615.
[32] Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucl. Acids Res. 32 (5), 1792–1797.

[33] Edwards, U., Rogall, T., Blöcker, H., Emde, M., Böttger, E.C. (1989) Isolation and direct complete nucleotide determination of entire genes. Characterization of a gene coding for 16S ribosomal RNA. Nucl. Acids Res. 17 (19), 7843–7853.
[34] Fernandez, M.P., Meugnier, H., Grimont, P.A.D., Bardin, R. (1989) Deoxyribonucleic acid relatedness among members of the genus *Frankia*. Int. J. Syst. Bacteriol. 39, 424–429.
[35] Gauthier, D., Frioni, L., Diem, H., Dommergues, Y. (1984) The *Colletia spinosissima-Frankia* symbiosis. Oecol. Plant 5, 231–239.
[36] 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.
[37] Gardes, M., Lalonde, M. (1987) Identification and subgrouping of *Frankia* strains using sodium dodecyl sulfate-polyacrylamide gel electrophoresis. Physiol. Plant. 70, 237–244.
[38] Gardes, M., Bousquet, J., Lalonde, M. (1987) Isozyme variation among 40 *Frankia* strains. Appl. Environ. Microbiol. 53, 1596–1603.
[39] Ghodhbane-Gtari, F., Nouioui, I., Chair, M., Boudabous, A., Gtari, M. (2010) 16S-23S rRNA intergenic spacer region variability in the genus *Frankia*. Microb. Ecol. 60, 487–495.
[40] Girgis, M.G.Z., Ishac, Y.Z., El-Hadded, M., Saleh, E.A., Diem, H.G., Dommergues, Y.R., et al. (1990) First report on isolation and culture of effective *Casuarina*-compatible strains of *Frankia* form Egypt. In: El-Lakany, M.H. (Ed.), Advances in Casuarina research and utilization, Desert Development Center, Cairo, Egypt, pp. 156–164.
[41] Goris, J., Konstantinidis, K.T., Klappenbach, J.A., Coenye, T., Vandamme, P., Tiedje, J.M. (2007) DNA–DNA hybridization values and their relationship to whole-genome sequence similarities. Int. J. Syst. Evol. Microbiol. 57 (1), 81–91.
[42] Gtari, M., Brusetti, L., Skander, G., Mora, D., Boudabous, A., Daffonchio, D. (2004) Isolation of *Elaeagnus*-compatible *Frankia* from soils collected in Tunisia. FEMS Microbiol. Lett. 234, 349–355.
[43] Gtari, M., Ghodhbane-Gtari, F., Nouioui, I., Ktari, A., Hezbri, K., Mimouni, W., et al. (2015) Cultivating the uncultured: growing the recalcitrant cluster-2 *Frankia* strains. Sci. Rep. 5, 13112.
[44] Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O. (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. 59 (3), 307–321.
[45] Guindon, S., Gascuel, O. (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol. 52 (5), 696–704.
[46] Hafeez, F. 1983 Nitrogen fixation and nodulation in *Datisca cannabina* L. and *Alnus nitida* Endl. PhD thesis, Quaid-e-Azam University, Islamabad, Pakistan.
[47] Hahn, D., Nickel, A., Dawson, J. (1999) Assessing *Frankia* populations in plants and soil using molecular methods. FEMS Microbiol. Ecol. 29, 215–227.
[48] Hahn, D., Mirza, B., Benagli, C., Vogel, G., Tonoll, M. (2011) Typing of nitrogen-fixing *Frankia* strains by matrix-assisted laser désorption ionization-time-of-flight (MALDI-TOF) mass spectrometry. Syst. App. Microbiol. 34, 63–68.
[49] Hall, T.A. (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp. Ser. 41, 95–98.
[50] Hameed, S., Hafeez, F.Y., Mirza, M.S., Malik, K.A., Akkermans, A.D.L. (1994) Confirmation of an isolate from *Datisca cannabina* as atypical *Frankia* strain using PCR amplified 16 rRNA sequence analysis. Pak. J. Bot. 26, 247–251.
[51] Higgins, L.M., Kennedy, P.G. (2012) Symbiotic *Frankia* bacteria in *Alnus* forests in Mexico and the United States of America: is geographic location a good predictor of assemblage structure? Botany 90, 423–431.
[52] Hsieh, T.C., Ma, K.H., Chao, A. (2016) iNEXT: an R package for rarefaction and extrapolation of species diversity (Hill numbers). Methods Ecol. Evol. 7 (12), 1451–1456.
[53] Jamann, S., Fernandez, M.P., Moiroud, A. (1992) Genetic diversity of Elaeagnaceae-infective *Frankia* strains isolated from various soils. Acta Oecol. 13, 395–405.
[54] 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.
[55] Janssen, P., Coopman, R., Huys, G., Swings, J., Bleeker, M., Vos, P., et al. (1996) Evaluation of the DNA fingerprinting method AFLP as a new tool in bacterial taxonomy. Microbiology 142, 1881–1893.
[56] 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 (9), 1220–1230.
[57] 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.
[58] Kim, B.J., Lee, J.H., Lyu, M.A., Kim, S.J., Bai, G.H., Chae, G.T., et al. (1999) Identification of mycobacterial species by comparative analysis of the RNA polymerase gene (*rpo*B). J. Clin. Microbiol. 37, 1714–1720.
[59] Lalonde, M. (1979) Techniques and observations of the nitrogen fixing *Alnus* root nodule symbiosis. In: Subba Rao, N.S. (Ed.), Recent advances in biological nitrogen fixation, Oxford and IBH Publishing Co., New Delhi, Bombay, Calcutta, pp. 421–434.
[60] Lalonde, M., Calvert, H.E., Pine, S. (1981) Isolation and use of *Frankia* strains in actinorhizae formation. In: Gibson, A.H., Newton, W.E. (Eds.), Current perspectives in nitrogen fixation, Australian Academy of Science, Canberra, pp. 296–299.
[61] Lechevalier, M.P. (1986) Catalog of *Frankia* strains. Actinomycetes 19, 131–162.
[62] Lechevalier, M.P., Lechevalier, H.A. (1979) The taxonomic position of the actinomycetic endophytes. In: Gordon, J.C., Wheeler, C.T., Perry, D.A. (Eds.), Symbiotic nitrogen fixation in the management of temperate forests, Oregon State University, Forest Research Laboratory, Corvallis, OR, pp. 111–121.
[63] Lechevalier, M.P., Lechevalier, H.A. (1990) Systematics, isolation and culture of *Frankia*. In: Schwintzer, C.R., Tjepkema, J.D. (Eds.), The biology of *Frankia* and actinorhizal plants, Academic Press Inc., San Diego, California, pp. 35–60.
[64] Librado, P., Rozas, J. (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25, 1451–1452.
[65] Mansour, S.R., Moussa, L.A.A. (2005) Role of Gamma-radiation on spore germination and infectivity of *Frankia strains* CeI523 and CcI6 isolated from Egyptian. Casuarina Isotope Rad. Res. 37, 1023–1038.
[66] Marchandin, H., Teyssier, C., Siméon de Buochberg, M., Jean-Pierre, H., Carriere, C., Jumas-Bilak, E. (2003) Intra-chromosomal heterogeneity between the four 16S rRNA gene copies in the genus *Veillonella*: implications for phylogeny and taxonomy. Microbiology 149, 1493–1501.
[67] Marechal, J., Clement, B., Nalin, R., Gandon, C., Orso, S., Cvejic, J.H., et al. (2000) A *rec*A gene phylogenetic analysis confirms the close proximity of *Frankia* to *Acidothermus*. Int. J. Syst. Evol. Microbiol. 50, 781–785.
[68] Meudt, H.M., Clarke, A.C. (2007) Almost forgotten or latest practice? AFLP applications, analyses and advances. Trends Plant Sci. 12, 106–117.
[69] Mirza, B.S., Welsh, A., Rieder, J.P., Paschke, M.W., Hahn, D. (2009) Diversity of *Frankiae* in soils from five continents. Syst. Appl. Microbiol. 32, 558–570.
[70] Mirza, M., Hahn, D., Akkermans, A. (1992) Characterization of *Frankia* strains from *Coriaria nepalensis*. Syst. Appl. Microbiol. 15, 289–295.
[71] Monz, C.A., Schwintzer, C.R. (1989) The physiology of spore-negative and spore-positive nodules of *Myrica gale*. Plant Soil 118, 75–87.
[72] Mougel, C., Thioulouse, J., Perriere, G., Nesme, X. (2002) A mathematical method for determining genome divergence and species delineation using AFLP. Int. J. Syst. Evol. Microbiol. 52, 573–586.
[73] Murry, M.M., Fontaine, M.S., Torrey, J.G. (1984) Growth kinetics and nitrogenase induction in *Frankia* sp. HFP ArI3 grown in batch culture. Plant Soil 78, 61–78.
[74] 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.
[75] Nalin, R., Normand, P., Domenach, A.M. (1997) Distribution and N2-fixing activity of *Frankia* strains in relation to soil depth. Physiol. Plant. 99, 732–738.
[76] 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 254, 207–217.
[77] Navarro, E., Jaffre, T., Gauthier, D., Gourbiere, F., Rinaudo, G., Simonet, P., Normand, P. (1999) Distribution of *Gymnostoma* spp. microsymbiotic *Frankia* strains in New Caledonia is related to soil type and to host-plant species. Mol. Ecol. 8, 1781–1788.
[78] 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.
[79] Nichols, R. (2001) Gene trees and species trees are not the same. Trends Ecol. Evol. 16, 358–364.
[80] Normand, P., Benson, D. 2012 Order VI. Frankiales ord. nov. Bergey's manual of systematic bacteriology, 5, pp. , 509–512.
[81] Normand, P., Lalonde, M. (1982) Evaluation of *Frankia* strains isolated from provenances of two *Alnus* species. Can. J. Microbiol. 28 (10), 1133–1142.
[82] Normand, P., Lalonde, M. (1986) The genetics of actinorhizal *Frankia*: a review. Plant Soil 90 (1), 429–453.
[83] 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.
[84] Normand, P., Lapierre, P., Tisa, L.S., Gogarten, J.P., Alloisio, N., Bagnarol, E., et al. (2007) Genome characteristics of facultatively symbiotic *Frankia* sp. strains reflect host range and host plant biogeography. Genome Res. 17, 7–15.
[85] Normand, P., Nguyen, T.V., Battenberg, K., Berry, A.M., Heuvel, B.V., Fernandez, M.P., Pawlowski, K. (2017) Proposal of *Candidatus Frankia* californiensis, the uncultured symbiont in nitrogen-fixing root nodules of a phylogenetically broad group of hosts endemic to western North America. Int. J. Syst. Evol. Microbiol. 67 (10), 3706–3715, http://dx.doi.org/10.1099/ijsem.0.002147.
[86] Nouioui, I., Del Carmen Montero-Calasanz, M., Ghodhbane-Gtari, F., Rohde, M., Tisa, L.S., Klenk, H.P., Gtari, M. (2017) *Frankia discariae* sp. nov.: an infective and effective microsymbiont isolated from the root nodule of *Discaria trinervis*. Arch. Microbiol. 199 (5), 641–647, http://dx.doi.org/10.1007/s00203-017-1337-6.
[87] Nouioui, I., Ghodhbane-Gtari, F., Beauchemin, N.J., Tisa, L.S., Gtari, M. (2011) Phylogeny of members of the *Frankia* genus based on *gyrB, nifH* and *glnII* sequences. Antonie Van Leeuwenhoek 100 (4), 578–587.
[88] Nouioui, I., Ghodhbane-Gtari, F., Del Carmen Montero-Calasanz, M., Rohde, M., Tisa, L.S., Gtari, M., Klenk, H.P. (2017) *Frankia inefficax* sp. nov., an actinobacterial endophyte inducing ineffective, non nitrogen-fixing, root nodules on its actinorhizal host plants. Antonie Van Leeuwenhoek 110 (3), 313–320.
[89] Nouioui, I., Ghodhbane-Gtari, F., Montero-Calasanz, M.D., Goker, M., Meier-Kolthoff, J.P., Schumann, P., Rohde, M., Goodfellow, M., Fernandez, M.P., Normand, P., et al. (2016) Proposal of a type strain for *Frankia alni* (Woronin 1866) Von Tubeuf 1895, emended description of *Frankia alni*, and recognition of *Frankia casuarinae* sp. nov. and *Frankia elaeagni* sp. nov. Int. J. Syst. Evol. Microbiol. 66 (12), 5201–5210.

[90] Nouioui, I., Ghodhbane-Gtari, F., Rohde, M., Klenk, H.P., Gtari, M. (2017) *Frankia coriariae* sp. nov., an infective and effective microsymbiont isolated from *Coriaria japonica*. Int. J. Syst. Evol. Microbiol. 67 (5), 1266–1270, http://dx.doi.org/10.1099/ijsem.0.001797.
[91] Paradis, E., Claude, J., Strimmer, K. (2004) APE: analyses of phylogenetics and evolution in R language. Bioinformatics 20, 289–290.
[92] Persson, T., Benson, D.R., Normand, P., Heuvel, B.V., Pujic, P., Chertkov, O., et al. (2011) Genome sequence of *Candidatus* Frankia datiscae Dg1, the uncultured microsymbiont from nitrogen-fixing root nodules of the dicot *Datisca glomerata*. J. Bacteriol. 193 (24), 7017–7018.
[93] Persson, T., Battenberg, K., Demina, I.V., Vigil-Stenman, T., Vanden Heuvel, B., Pujic, P., et al. (2015) *Candidatus* Frankia datiscae Dg1, the Actinobacterial Microsymbiont of *Datisca glomerata*, expresses the canonical nod genes nodABC in symbiosis with its host plant. PLoS One 10, e0127630.
[94] Pölme, S., Bahram, M., Koljalg, U., Tedersoo, L. (2014) Global biogeography of *Alnus* associated *Frankia* actinobacteria. New Phytol. 204, 979–988.
[95] Pozzi, A.C., Bautista-Guerrero, H.H., Nouioui, I., Cotin-Galvan, L., Pépin, R., Fournier, P., et al. (2015) *In-planta* sporulation phenotype: a major life history trait to understand the evolution of *Alnus*-infective *Frankia* strains. Environ. Microbiol. 17 (9), 3125–3138, http://dx.doi.org/10.1111/1462-2920.12644.
[96] Pozzi, A.C., Roy, M., Nagati, M., Schwob, G., Manzi, S., Moreau, P.A., et al. (2018) Patterns of diversity, endemism and specialization in the root symbiont communities of alder species on the island of Corsica. New Phytol., http://dx.doi.org/10.1111/nph.14996, in press.
[97] Prin, Y., Maggia, L., Picard, B., Diem, H.G., Goullet, P. (1991) Electrophoretic comparison of enzymes from 22 single-spore cultures obtained from *Frankia* strain ORS 140102. FEMS Microbiol. Lett. 77 (2–3), 223–227.
[98] Puppo, A., Dimitrijevic, L., Diem, H., Dommergues, Y. (1985) Homogeneity of superoxide dismutase patterns in *Frankia* strains from Casuarinaceae. FEMS Microbiol. Lett. 30, 43–46.
[99] R Core Team, 2017 R: a language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria https://www.R-project.org/.
[100] Rademaker, J.L.W., Hoste, B., Louws, F.J., Kersters, K., Swings, J., Vauterin, L., et al. (2000) 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.
[101] Ramirez-Saad, H., Janse, J.P., Akkermans, A.D.L. (1998) Root nodules of *Ceanothus caeruleus* contain both the N2-fixing *Frankia* endophyte and a phylogenetically related Nod-/Fix- actinomycete. Can. J. Microbiol. 44, 140–148.
[102] Ranwez, V., Harispe, S., Delsuc, F., Douzery, E.J.P. (2011) MACSE: Multiple Alignment of Coding SEquences accounting for frameshifts and stop codons. PLoS One 6 (9), e22594, http://dx.doi.org/10.1371/journal.pone.0022594.
[103] Rodriguez, D., Guerra, T.M., Forstner, M.R.J., Hahn, D. (2016) Diversity of *Frankia* in soil assessed by Illumina sequencing of *nifH* gene fragments. Syst. Appl. Microbiol. 39 (6), 391–397.
[104] Roy, M., Pozzi, A.C., Gareil, R., Nagati, M., Manzi, S., Nouioui, I., et al. (2017) Alder and the Golden Fleece: high diversity of *Frankia* and ectomycorrhizal fungi revealed from *Alnus glutinosa* subsp. *barbata* roots close to a Tertiary and glacial refugium. Peer J. 5, e3479 https://doi.org/10.7717/peerj.3479.
[105] Schwencke, J., Caru, M. (2001) Advances in actinorhizal symbiosis: host-plant-*Frankia* interactions, biology, and applications in arid land reclamation. A review. Arid Land Res. Manag. 15, 285–327.
[106] Schwintzer, C.R. (1990) Spore-positive and spore-negative nodules. In: Schwintzer, C.R., Tjepkema, J.D. (Eds.), The biology of Frankia and Actinorhizal plants, Academic Press, San Diego, pp. 177–193.
[107] Schwintzer, C.R., Tjepkema, J.D. 1990 The biology of frankia and actinorhizal plants, Academic Press, San Diego.
[108] Sen, A., Daubin, V., Abrouk, D., Gifford, I., Berry, A.M., Normand, P. (2014) Phylogeny of the class Actinobacteria revisited in the light of complete genomes. The orders 'Frankiales' and Micrococcales should be split into coherent entities: proposal of Frankiales ord. nov., Geodermatophilales ord. nov., Acidothermales ord. nov. and Nakamurellales ord. nov. Int. J. Syst. Evol. Microbiol. 64 (11), 3821–3832.
[109] Shi, Y., Ruan, J. (1992) DNA base composition and homology values in the classification of some *Frankia* strains. Acta Microbiol. Sin. 32, 133–136.
[110] 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.
[111] Simonet, P., Normand, P., Moiroud, A., Lalonde, M. (1985) Restriction enzyme digestion patterns of *Frankia* plasmids. Plant Soil 87, 49–60.
[112] 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.
[113] 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.
[114] Smolander, A., Sarsa, M. (1990) *Frankia* strains of soil under *Betula pendula* — behavior in soil and pure culture. Plant Soil 122, 129–136.
[115] Soltis, D.E., Soltis, P.S., Morgan, D.R., Swensen, S.M., Mullin, B.C., Dowd, J.M., Martin, P.G. (1995) Chloroplast gene sequence data suggest a single origin of the predisposition for symbiotic nitrogen fixation in angiosperms. Proc. Natl. Acad. Sci. 92 (7), 2647–2651.
[116] Swensen, S.M. (1996) The evolution of actinorhizal symbioses: evidence for multiple origins of the symbiotic association. Am. J. Bot. 83, 1503–1512.
[117] Swensen, S.M., Benson, D.R. (2008) Evolution of actinorhizal host plants and *Frankia* endosymbionts. In: Pawlowski, K., Newton, W.E. (Eds.), Nitrogen fixing actinorhizal symbioses, Springer, Dordrecht, pp. 73–104.
[118] Tekaya, S.B., Ganesan, A.S., Guerra, T., Dawson, J.O., Forstner, M.R., Hahn, D. (2017) Sybr green-and TaqMan-based quantitative PCR approaches allow assessment of the abundance and relative distribution of *Frankia* clusters in soils. Appl. Environ. Microbiol. 83 (5), e02833–16.
[119] Tisa, L.S., Beauchemin, N., Cantor, M.N., Furnholm, T., Ghodhbane-Gtari, F., Goodwin, L., et al. (2015) Draft genome sequence of *Frankia* sp. strain DC12, an atypical, noninfective, ineffective isolate from *Datisca cannabina*. Genome Announc. 3 (4), e00889–15.
[120] Tisa, L.S., Oshone, R., Sarkar, I., Ktari, A., Sen, A., Gtari, M. (2016) Genomic approaches toward understanding the actinorhizal symbiosis: an update on the status of the *Frankia* genomes. Symbiosis 70 (1–3), 5–16.
[121] Torrey, J.G. (1987) Endophyte sporulation in root nodules of actinorhizal plants. Physiol. Plant. 70, 279–288.
[122] Vinuesa, P., Silva, C., Werner, D., Martinez-Romero, E. (2005) Population genetics and phylogenetic inference in bacterial molecular systematics: the roles of migration and recombination in *Bradyrhizobium* species cohesion and delineation. Mol. Phylogenet. Evol. 34, 29–54.
[123] VandenBosch, K.A., Torrey, J.G. (1985) Development of endophytic *Frankia* sporangia in field- and laboratory- grown nodules of *Comptonia peregrina* and *Myrica gale*. Am. J. Bot. 72, 99–100.
[124] Vos, P., Hogers, R., Bleeker, M., Reijans, M., Van de Lee, T., Hornes, M., et al. (1995) AFLP: a new technique for DNA fingerprinting. Nucl. Acids Res. 23 (21), 4407–4414.
[125] Wayne, L.G., Brenner, D.J., Colwell, R.R., Grimont, P.A.D., Kandler, O., Krichevsky, et al. (1987) Report of the ad hoc committee on reconciliation of approaches to bacterial systematics. Int. J. Syst. Evol. Microbiol. 37 (4), 463–464.
[126] Welsh, A., Mirza, B.S., Rieder, J.P., Paschke, M.W., Hahn, D. (2009) Diversity of Frankiae in root nodules of *Morella pensylvanica* grown in soils from five continents. Syst. Appl. Microbiol. 32, 201–210.
[127] Werner, G.D.A., Cornwell, W.K., Sprent, J.I., Kattge, J., Kiers, T. (2014) A single evolutionary innovation drives the deep evolution of symbiotic N_2-fixation in angiosperms. Nat. Commun. 5, 4087.
[128] Wolters, D.J., Van Dijk, C., Zoetendal, E.G., Akkermans, A.D.L. (1997) Phylogenetic characterization of ineffective *Frankia* in *Alnus glutinosa* (L.) Gaertn. nodules from wetland soil inoculants. Mol. Ecol. 6 (10), 971–981.
[129] Wu, D., Hugenholtz, P., Mavromatis, K., Pukall, R., Dalin, E., Ivanova, N.N., et al. (2009) A phylogeny-driven genomic encyclopaedia of Bacteria and Archaea. Nature 462 (7276), 1056–1060.
[130] Yamanaka, T., Mansour, S.R. (2013) Nodulation of *Alnus japonica* and *Casuarina equisetifolia* in liquid culture after inoculation with *Frankia*. Bull FFPRI 12, 97–103.
[131] Yap, W.H., Zhang, Z., Wang, Y. (1999) Distinct types of rRNA operons exist in the genome of the actinomycete *Thermomonospora chromogena* and evidence for horizontal transfer of an entire rRNA operon. J. Bacteriol. 181, 5201–5209.
[132] 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 Soil 78, 79–90.
[133] Zhi, X.-Y., Li, W.-J., Stackebrandt, E. (2009) An update of the structure and 16S rRNA gene sequence-based definition of higher ranks of the class Actinobacteria, with the proposal of two new suborders and four new families and emended descriptions of the existing higher taxa. Int. J. Syst. Evol. Microbiol. 59, 589–608.