



## Genomespecies identification and phylogenomic relevance of AFLP analysis of isolated and non-isolated strains of *Frankia* spp.

G.H.H. Bautista<sup>a,c</sup>, H.A. Cruz<sup>b</sup>, X. Nesme<sup>a</sup>, M. Valdés<sup>c</sup>, H.A. Mendoza<sup>b</sup>, M.P. Fernandez<sup>a,\*</sup>

<sup>a</sup> Université de Lyon, Université Lyon 1, CNRS, INRA, Ecologie Microbienne, UMR5557, USC 1193, 43 boulevard du 11 novembre 1918, F-69622 Villeurbanne Cedex, France

<sup>b</sup> Centro de Biotecnología Genómica, IPN, col. Narciso Mendoza CP 88710, Reynosa, Tam, Mexico

<sup>c</sup> Laboratorio de Microbiología Agrícola, Escuela Nacional de Ciencias Biológicas, IPN, México D.F., Mexico

### ARTICLE INFO

#### Article history:

Received 11 June 2010

#### Keywords:

AFLP  
Evolutionary genomic distance  
*Frankia* phylogeny  
Uncultured strains  
Genomespecies

### ABSTRACT

Amplified fragment length polymorphism (AFLP) was tested as an alternative to the DNA–DNA hybridization technique (DDH) to delineate genomespecies and the phylogenetic structure within the genus *Frankia*. Forty *Frankia* strains, including representatives of seven DDH genomespecies, were typed in order to infer current genome mispairing (CGM) and evolutionary genomic distance (EGD). The constructed phylogeny revealed the presence of three main clusters corresponding to the previously identified host-infecting groups. In all instances, strains previously assigned to the same genomespecies 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 range 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 new *Frankia* strains to be assigned to nine genomespecies 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 refractory spore-positive (Sp+, that sporulate *in planta*) strains. Only 1.3% of all AFLP fragments were shown to be generated by the contaminant plant DNA and did not interfere with accurate genomespecies identification of strains. When applied to field nodules, the procedure revealed that *Alnus* Sp+ strains were *bona fide* members of the *Alnus-Myrica* host infecting group. They displayed significant genomic divergence from genomespecies G1 of *Alnus* infecting strains (i.e. *Frankia alni*) and thus may belong to another subspecies or genomespecies.

© 2011 Elsevier GmbH. All rights reserved.

### Introduction

Actinobacteria belonging to the genus *Frankia* are able to establish nitrogen fixing symbiosis with the roots of 25 genera of angiosperms collectively called actinorhizal plants.

*Frankia* spp. is the only known genus of the family Frankiaceae. Isolates display very specific traits that make them easy to distinguish from other actinomycetes [26,27]. On the other hand, there are major morphological, physiological and genetic differences within *Frankia* spp., thus suggesting that there is high diversity within the genus [4,18]. The use of DNA–DNA hybridization (DDH) – still the gold standard for delineating bacterial species [16,48] – has led to the identification of at least 12 genomespecies within *Frankia* spp. [1,2,6,14,21]. However, the *bona fide* species status of *Frankia* genomespecies has yet to be granted because no dis-

criminate phenotypic traits have been found to date. In addition, for more than 50% of actinorhizal plants, *Frankia* specimens are consistently refractory to *in vitro* culture, or host reinfection has not been achieved [4]. The diversity of these uncultured strains at the species level is unknown since DDH cannot be used, suggesting that the total number of *Frankia* species is far more than twelve.

The analysis of 16S rRNA genes, 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, that pools strains infective on *Alnus*, *Comptonia*, *Myrica* and *Casuarina*; cluster 2, that pools the non-isolated symbionts of Rosaceae, Datisceae, Coriariaceae and *Ceanothus* sp.; cluster 3, that pools strains infective on Elaeagnaceae and Rhamnaceae [40]. Cluster 1 is subdivided into clusters 1a and 1b containing strains infective on *Alnus*, *Comptonia* or *Myrica*, and *Casuarina*, respectively. The comparative analysis of partial 16S rRNA gene sequences also showed that culture refractory *Alnus*-infective *Frankia* strains, which are called spore-positive (Sp+) because of their typical ability to synthesize numerous spores *in planta*, represent a different genotype from the Sp– [46]. How-

\* Corresponding author at: 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 address: [maria.fernandez@univ-lyon1.fr](mailto:maria.fernandez@univ-lyon1.fr) (M.P. Fernandez).

ever, 16S rRNA lacks sufficient resolution for accurate and secure genomospecies identification, since *Frankia* genomospecies displaying as little as 2–7% DNA–DNA homology by DDH [14] can share up to 98% 16S rRNA sequence similarity. Comparative analysis based on *glnII* and *nifH* sequences – exhibiting more variability than 16S rRNA – confirmed the major divisions within the genus [8,17]. *nifH* has recently been used to assign uncultured *Frankia* to host infection groups and subgroups [32,51]. However, a study based on a large geographic sampling showed that *Frankia* clustering, based on comparative analysis of the *nifH* sequence, does not consistently describe genomic groups [51], most likely because the *nifH* locus might be subject to lateral transfers [52]. As a general feature, all loci may be subject to lateral transfer, and it is established that a single-gene based phylogeny does not necessarily reflect species phylogeny [16,38,49].

Fingerprinting methods have also been used to gain insight into strain clustering and diversity. These include both multilocus and single locus approaches, such as rep-PCR (repetitive sequence amplification), and PCR-RFLP of intergenic regions 16S–23S rRNA and *nifD-K*, respectively [12,20,24,29,30,36]. These methods were found to be suitable for strain typing, but they generally deliver

a limited number of pattern similarities to allow the inference of robust phylogenesis [31].

This is not the case with AFLP, which is a PCR-based fingerprinting method that randomly targets numerous loci throughout the genome, generating a sufficiently high number of pattern similarities for robust phylogenetic purposes [22,50]. In addition, the method is versatile and can be suitable for targeting genomes. Actually, even though no knowledge of the genome sequence is required for AFLP, complete genome sequences enable *in silico* simulations for optimal estimation of the best parameters in order to achieve maximum genome coverage with the least homoplasmy [44]. In addition, the method can be adapted to produce sufficient fragments for robust analysis, and fluorescent AFLP, using automatic sequencers and fluorescent adaptors, automatically provides accurate estimations of fragment size [28]. AFLP is deemed to be an accurate approach for determining phylogenetic and/or taxonomic relationships and for the assessment of genetic diversity in bacteria [33]. Indeed, AFLP trials with several taxa have shown that evolutionary genomic divergence (EGD) between strains and their common ancestor, deduced from AFLP data, was particularly useful for phylogenomic purposes [23,34,42], and that members of

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

Trivial designation	Registry no.	Original host	Geographic origin	Ref. <sup>a</sup>	Genomospecies	
					Original number <sup>b</sup>	This study
<b><i>Alnus</i> and <i>Myrica</i> strains</b>						
32-44(Air11)	LLR01322	<i>Alnus incana</i> ssp. <i>rugosa</i>	Vermont, USA	[2]	1	G1
ACN1 <sup>AG</sup>	ULQ0102001007	<i>Alnus crispa</i>	Tadoussaq, Canada	[2,14]	1	G1
AcoN24d	ULF01010244	<i>Alnus cordata</i>	Orléans, France	[14]	1	G1
AcVcl	ULF010102171	<i>Alnus cordata</i>	Corse, France	[14]	1	G1
Ar24H3	ULF0131024083	<i>Alnus rubra</i>	Orléans, France	[14]	1	G1
Ar24H5		<i>Alnus rubra</i>	Orléans, France	–	1	G1
ARgN22d	ULQ0132022024	<i>Alnus rugosa</i>	Quebec, Canada	[14]	1	G1
Avcl1	DDB010110	<i>Alnus viridis</i> ssp. <i>crispa</i>	Ontario, Canada	[2,6,14]	1	G1
Cpl1	HFP07010701	<i>Comptonia peregrina</i>	Massachusetts, USA	[2,6,14]	1	G1
MPI1	LLR162001	<i>Myrica pensylvanica</i>	Nantucket, VT, USA	[2]	1	G1
M16467	RBR162021	<i>Myrica pensylvanica</i>	New Jersey, USA	[6]	1	G1
ACN14a	ULQ010201401	<i>Alnus crispa</i>	Tadoussaq, Canada	–	ND	G1
Ag24 <sub>251</sub>	ULF0107024251	<i>Alnus glutinosa</i>	Orléans, France	[14]	1	G1
ArI3	HFP013103	<i>Alnus rubra</i>	Oregon, USA	[14]	1	G1
I38		<i>Alnus incana</i>	La pallud, France	–	ND	G1
M16477		<i>Myrica pensylvanica</i>	New Jersey, USA	–	ND	G1
Ac23 <sub>40</sub>	ULF010102340	<i>Alnus crispa</i>	Orléans, France	[14]	2	G2
AVL3	ULF014102203	<i>Alnus viridis</i>	Lautaret, France	[14]	2	G2
AVN17o	ULF014101715	<i>Alnus viridis</i>	La Toussuire, France	[14]	2	G2
ARgP5 <sup>AG</sup>	ULQ0132105009	<i>Alnus rugosa</i>	Quebec, Canada	[14]	3	G3
<b>Casuarinaceae strains</b>						
BR	ORS020608	<i>Casuarina equisetifolia</i>	Brazil	[14]	9	G4
CcI3	HFP020203	<i>Casuarina cunninghamiana</i>	Florida, USA	[14]	9	G4
CjI-82	ORS021001	<i>Casuarina junghuniana</i>	Thailand	[14]	9	G4
TA	ORS022602	<i>Allocasuarina torulosa</i>	Australia	[14]	9	G4
Cg70 <sub>4</sub>		<i>Casuarina glauca</i>	India	–	ND	G4
Cg70 <sub>5</sub>		<i>Casuarina glauca</i>	India	–	ND	G4
Cg70 <sub>3</sub>		<i>Casuarina glauca</i>	India	–	ND	G5
CcI2	HFP020202	<i>Casuarina cunninghamiana</i>	Florida, USA	[14]	ND	G5
<b>Elaeagnaceae–Rhamnaceae strains</b>						
Ea1 <sub>12</sub>	ULF130100112	<i>Elaeagnus angustifolia</i>	Ecully, France	[14]	4	G6
Ea1 <sub>2</sub>	ULF130100102	<i>Elaeagnus angustifolia</i>	Ecully, France	[14]	4	G6
Ea48 <sub>4</sub>		<i>Elaeagnus angustifolia</i>	France	[21]	4	G6
Ea7 <sub>1</sub>		<i>Elaeagnus angustifolia</i>	Toulon, France	[21]	4	G6
EaCm5 <sub>1</sub>	ULF130100501	<i>Elaeagnus angustifolia</i>	Escarène, France	[14]	4	G6
	ORS060501	<i>Colletia spinosissima</i>	Argentina	[14]	ND	G6
Hr75 <sub>2</sub>		<i>Hippophaë rhamnoides</i>	France	–	ND	G6
CH37	ORS140102	<i>Hippophaë rhamnoides</i>	Nogent/Marne, France	–	ND	G9
GFN14a	GFN140101	<i>Hippophaë rhamnoides</i>	China	[14]	ND	G9
<b>Atypical strains</b>						
Mgl <sub>5</sub>	DDB16110210	<i>Myrica gale</i>	New York, USA		ND	G7
Pti1	DDB170110	<i>Purshia tridentata</i>	Wyoming, USA	[45]	3	G7
G2	ORS020604	<i>Casuarina equisetifolia</i>	Guadeloupe	[2]	2	G8

<sup>a</sup> According to the first genomospecies description.

<sup>b</sup> As given in the reference cited in this table.

the same genomospecies consistently clustered in coherent groups. Since there is a high correlation between DDH values and genomic distances estimated with AFLP data [9,34,43], the International Committee on Taxonomy proposed that molecular traits including AFLP patterns can now also be used to validly distinguish bacterial species [48]. However, AFLP and DDH were claimed to be inapplicable for targeting a specific DNA from a complex source [33], thus, *a priori*, excluding studies of intracellular organisms, such as parasites, obligate symbionts and viruses. This point has, however, not been thoroughly investigated in the case of microsymbionts such as *Frankia* that display very different genome compositions than their hosts, which is a molecular feature that influences AFLP specificity.

In the present study, the AFLP technique was used to analyze taxonomic and 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 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 the suitability of AFLP to be evaluated for genomospecies assignment of both cultured and uncultured strains of *Frankia* spp.

## Materials and methods

### *Frankia* strains and DNA extraction

The *Frankia* strains used in this study are listed in Table 1. They were cultured in BAP medium [35] 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), according to the manufacturer's instructions. Field *Alnus* nodules known to be hosting strains with the Sp+ phenotype ([46] 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 elevations at the La Toussuire site. DNA was extracted from nodules using the classical cetyltrimethylammonium bromide (CTAB; Sigma, St-Louis, MO) method with slight modifications [19].

### *In silico* predictive AFLP

Predictive AFLP was performed *in silico* with complete genome sequences of *Frankia* strains Ccl3, ACN14a and EAN1pec (Accession numbers: CP000249, CT573213 and AAI00000000), with the *In-silico* program (<http://insilico.ehu.es/AFLP/>) to determine the best AFLP parameters for analyzing *Frankia* spp. For this purpose, endonucleases were preferentially focused on that recognized a high G+C content, and that were expected to generate well resolved patterns in the 40–450 bp range 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 the similarities, special care was taken to select AFLP conditions that predicted 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. [50], with some modifications, by using the endonuclease combinations PstI-TaqI and PstI-HhaI, adaptors and PCR primers (Table 2). AFLPs were performed with 55 ng of DNA for pure bacterial cultures, and 100 ng for complex DNA extracted from nodules.

For the PstI-HhaI set, digestion and ligation steps were performed simultaneously by incubating samples at 37 °C for 2 h in

**Table 2**

AFLP oligonucleotides used to construct adaptors and to prime PCRs.

Oligonucleotide	Sequence
PstI-specific adaptors	
F1604-adPst+	CTCGTAGACTGCGTACATGCA
F2757-adPst–	CATCTGACCGATGT
TaqI-specific adaptors	
F1365-adTaq+	GACGATGAGTCTCTGAG
F3878adTaq–	CGCTCAGGACTCAT
HhaI-specific adaptors	
F3879-adHha+	GACGATGAGTCTCTGACC
F3880-adHha–	GTCAGGACTCATC
Core primers	
F2758-corePst	GACTGCGTACATGCGAC
F3798-coreTaq	GATGAGTCTCTGAGGCA
F3881-coreHha	GATGAGTCTCTGACC
Selective primers	
F3804-PstAA-FAM	GACTGCGTACATGCAGAA
F3805-PstAC-HEX	GACTGCGTACATGCAGAC

11 µL (final volume) of mixture containing the target DNA, PstI (5 U), HhaI (5 U), T4 DNA ligase (1 U), 1× T4 DNA ligase buffer (Boehringer-Mannheim, 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 the PstI-TaqI set, digestion and ligation were performed in two steps because of differences in the optimal temperatures for enzyme activities. TaqI digestion was performed first at 65 °C for 2 h, and then at 37 °C for 2 h with PstI. The adaptor ligation was performed with 55 ng of digested DNAs, as described above.

### Non-selective and selective PCR amplification

Adapted DNAs (4 µL) were subjected to non-selective 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 0.25 µM primer F3798-coreTaq, for the PstI-TaqI set. Non-selective PCR reactions were carried out in a PE-9600 thermocycler (Perkin-Elmer Applied Biosystems, Foster City, Calif.) with an initial denaturation at 94 °C for 5 min, and 20 denaturation cycles at 94 °C for 20 s, annealing at 56 °C for 30 s, and elongation at 72 °C for 2 min. The quality of non-selective PCRs was checked by visual inspection of 1% agarose gel electrophoresis. A good non-selective PCR typically displayed many fragments (smeared) between 200 bp and 600 bp.

Non-selective PCR products were diluted 1/50 before being 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, Carlsbad, USA). 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 reduced by 1 °C per cycle) for 30 s, and elongation at 72 °C for 2 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 20 elongation cycles at 72 °C for 2 min.

Total selective PCR products (10 µL) were purified on a Sephadex G-50 column (Amersham Biosciences, Orsay, France) before separation of AFLP fragments by electrophoresis with a capillary sequencer (MegaBACE 1000; Amersham Pharmacia Biotech Europe, Orsay, France). The device automatically calculated the length of fluorescent fragments by comparison with the MegaBACE ET-900-R size standard (Amersham Pharmacia Biotech Europe,

Orsay, France). 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 Thresholdfilter 1.3 (Yann Legros, Amersham), as previously described [42]. A threshold fluorescence value of 100 arbitrary units was used to eliminate background noise before subsequent scoring and bioinformatics treatments.

The accuracy and reliability of the AFLP method was tested by comparing patterns independently in duplicate or triplicate for 60% and 20% of the tested strains, respectively.

#### 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 current genome mispairing (CGM), expressed as the number of mismatches per site using the Jaccard index, and evolutionary genomic divergence (EGD), expressed as the number of nucleotide substitutions per site using the Dice index, with 12 as the number of nucleotides involved in restriction and selective PCR steps, as described by Mougél et al. [34]. LecPCR, Align2, and DistAFLP programs are accessible at <http://pbil.univ-lyon1.fr/ADE-4/microb/>. Dendrograms were calculated with the Neighbor/UPGMA module of the PHYLIP package. Bootstrap values were calculated using the bootstrap option (1000 replicates) in DistAFLP, and the Neighbor/UPGMA and the CONSENSE modules of the PHYLIP package were used to draw dendrograms with NJ-Plot (<http://pbil.univ-lyon1.fr/software/njplot.html>).

## Results and discussion

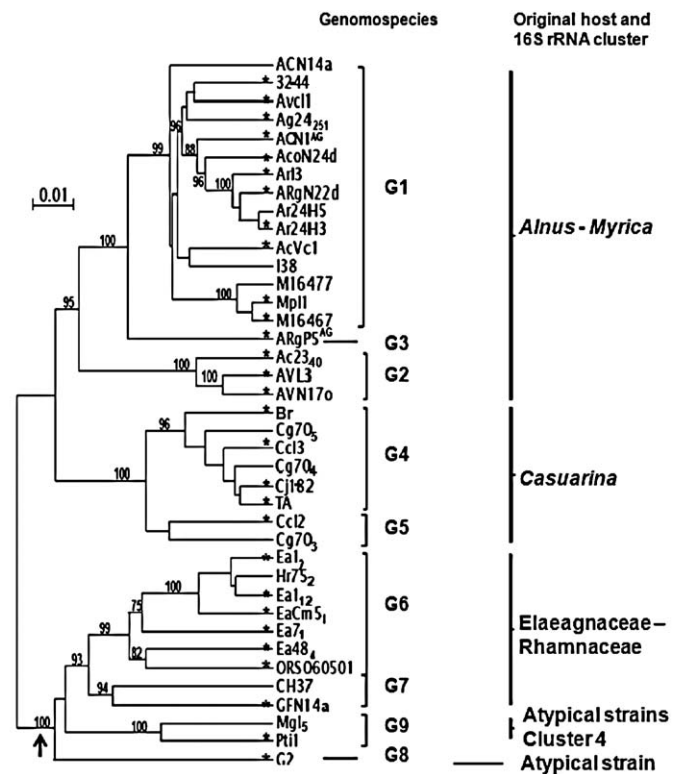
#### AFLP conditions and reliability

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  $60.7 \pm 6.5$ ,  $58.6 \pm 6.9$ ,  $61.3 \pm 8.4$  and  $61.4 \pm 8.7$  for PstI + AA/TaqI + 0, PstI + AC/TaqI + 0, PstI + AA/Hha + 0 and PstI + AC/Hha + 0, respectively, for a pooled total of  $242.0 \pm 21.7$ . This number varied considerably according to 16S rRNA clusters and/or host infectivity groups with  $249 \pm 10.1$ ,  $206 \pm 10.6$  and  $259 \pm 17.7$  for clusters 1a, 1b and 3, respectively, that corresponded to host infectivity groups *Alnus-Myrica*, *Casuarina* and *Elaeagnaceae-Rhamnaceae*, respectively.

Duplicate or triplicate samples produced more than 95% and 93% shared fragments, respectively (data not shown). The AFLP method thus allowed the pattern similarities between strains to be determined reliably and, in turn, the confident calculation of genomic distances between different strains. However, the method was found to be unreliable for generating absolutely identical patterns for a given strain (data not shown). This phenomenon has already been observed in other taxa [9,11,42]. The latter authors suggested that this unreliability could be related to epigenetic factors causing differential accessibility of genome regions to endonucleases [41]. Therefore, because of this known limitation of AFLP technology, *Frankia* isolates could not be definitively assigned to given strains or clone categories.

#### Phylogenomic analysis of cultured *Frankia* spp.

Percentages of polymorphic fragments (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 the pairwise genomic distances expressed as CGM and EGD (data not



**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 from elsewhere were available. Arrow indicates significant gathering of atypical and standard *Elaeagnaceae-Rhamnaceae* infective strains at the deep branching level.

shown) to be calculated. Phylogenomic dendrograms drawn with EGD data calculated with each enzyme-primer combination were very similar (data not shown). Thus, EGD data were pooled in order to infer a more robust phylogeny with UPGMA (Fig. 1), which was almost identical to 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 were always grouped in significant AFLP clusters (Fig. 1). This was the case for 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 analysis were always grouped in different AFLP clusters. This clearly demonstrated that AFLP was a suitable method to clearly delineate *Frankia* genomospecies previously determined by DDH.

In the present analysis, other isolates that were not analyzed previously by DDH grouped together with DDH tested strains. Indeed, it was confidently assumed that they readily belonged to the same genomospecies as their analyzed neighbors. This allowed us to infer that the sequenced strain ACN14a belonged to genomospecies G1.

#### AFLP threshold values for *Frankia* species definition

The possibility of defining a threshold of genomic distances to serve as a cut-off value for genomospecies delineation was questioned. Therefore, AFLP based estimates of genome mispairing (CGM) were plotted against DDH re-association values, and a strong significant linear relationship ( $r^2 = 0.87$ ) was obtained. However, as displayed in Fig. 2, the 60% DDH threshold, which we previously

**Table 3**  
Current genome mispairing (CGM) limits for *Frankia* genomospecies delineation.

Host-infectivity group	<i>Alnus-Myrica</i>			<i>Casuarina</i>		Elaeagnaceae		
Genomospecies <sup>a</sup>	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

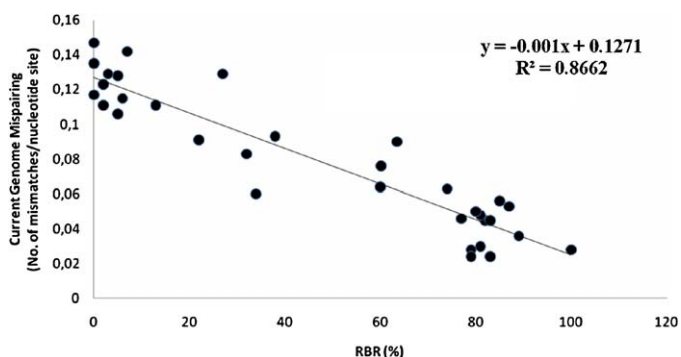
<sup>a</sup> This table only summarizes genomospecies for which DDH values are available.

found to be suitable for delineating *Frankia* genomospecies based on  $\Delta T_m$  considerations [1,2,14,21], did not correspond to a single CGM cut-off value. More precisely, the lowest value of genomospecies pairwise CGM was 0.080 mismatches per site between G1 and G3, while the largest CGM value found within a single genomospecies was in G6 (i.e. 0.098 mismatches per site).

Distinctive cut-offs could be determined according to major phylogeny clades (Table 3). Actually, it appeared that the value was between 0.071 and 0.080 mismatches per site in *Alnus-Myrica* and *Casuarina*-infecting genomospecies (16S rRNA cluster 1), while it was over 0.098 and below 0.105 mismatches per site in Elaeagnaceae–Rhamnaceae infecting genomospecies (e.g. 16S rRNA clusters 3 and 4). This cut-off discrepancy was also revealed in phylogenetic dendrograms inferred from EGD values (Fig. 1). Cut-off differences for genomospecies determination were also found with the *nifH* molecular marker [51].

The inconsistency of the genomic distance cut-off for the *Frankia* genus could be related to the dramatic differences in genome size between strains belonging to different host infecting groups. Actually, ACN14a (G1, *Alnus-Myrica* infecting group), Ccl3 (G4, *Casuarina* infecting group) and EaN1pec (G6, Elaeagnaceae–Rhamnaceae infecting group) have genome sizes of 7.5, 5.4 and 9.03 Mb, respectively [39]. It has been estimated that large genomes generate more numerous and specific AFLP fragments than small ones [13,15,31], while strains with reduced genomes that must keep essential and thus conserved genes would generate less strain specific AFLP fragments. Therefore, it is likely that genomospecies with large genomes, such as members of the Elaeagnaceae–Rhamnaceae infecting group, must generate more strain specific fragments and thus display much larger genomic distances between their members than groups with smaller genomes, such as members of the *Casuarina* infecting group.

By considering the different genomic distance cut-off values, it was assumed that AFLP data could be used to assign novel strains to their respective genomospecies. *Casuarina* infecting strains Ccl2 and Cg70<sub>3</sub> that displayed a CGM of 0.068 mismatches per site – thus below 0.071, the largest infra-specific CGM recorded in 16S rRNA cluster 1 – were hence thought to belong to the same genomospecies G5. In the Elaeagnaceae–Rhamnaceae infecting group, GFN14a and CH37 were distantly related (CGM = 0.094 mismatches per site), but still in the range of infra-specific distances of



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

the group. It is thus likely they are members of the same genomospecies G9.

In addition, with a CGM value of 0.071 mismatches per site below all cut-off values, atypical strains Mgl<sub>5</sub> and Pti1, which were grouped in 16S rRNA cluster 4 by Huguet et al. [19], could confidently be assigned to the same genomospecies G7 in the present AFLP study.

#### Deep branching of *Frankia* phylogeny

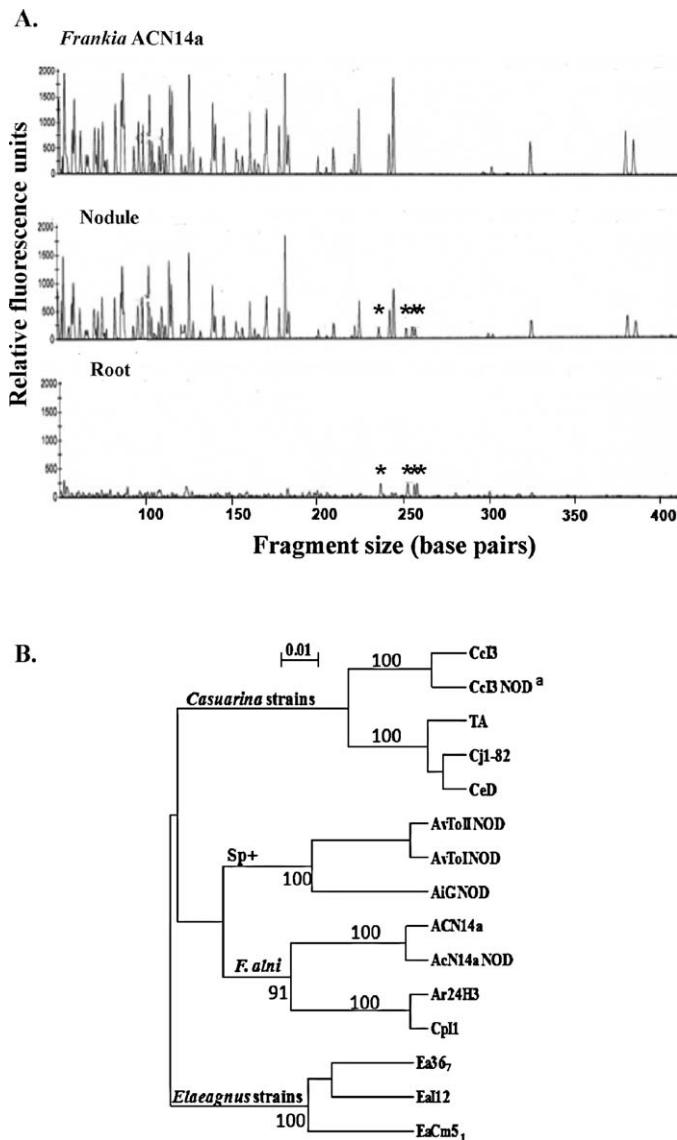
Remarkably, contrary to what has been found with several other taxa [7,25,42], bootstrap values strongly supported the deep branching of the *Frankia* phylogenetic tree obtained with AFLP data. In all instances, the phylogeny significantly grouped, in a standard fashion, infective *Frankia* strains according to their actual host infectivity groups *Alnus-Myrica*, *Casuarina*, and Elaeagnaceae–Rhamnaceae. The present results showed that this was also the case for atypical strains G2 and Mgl<sub>5</sub>, which are unable to re-infect their original host, *Casuarina* and *Myrica*, respectively, but which were found to be able to infect Elaeagnaceae. These readily Elaeagnaceae infective but still atypical strains coherently grouped in the same major AFLP cluster made up of standard Elaeagnaceae–Rhamnaceae infecting strains (arrow, Fig. 1).

Superimposed on host infectivity groups, the AFLP tree grouped *Frankia* strains according to previously defined 16S clusters 1a, 1b, 3 and 4 (Fig. 1). Here, cluster 4 (i.e. genomospecies G9) clearly appeared to be a sister clade of cluster 3. This feature was not supported by, or contradicted, the 16S rRNA phylogeny results because this marker is not resolvable enough to support significant bootstrap values at this level [19]. However, this feature is in line with the assignment of Mgl<sub>5</sub> to the Elaeagnaceae and Rhamnaceae-infective group, also obtained by using the symbiotic *nifH* gene [51]. It is thus worth noting the position of the atypical strain Pti1 (G9) together with Mgl<sub>5</sub>. It is known that Pti1 isolated from *Purshia tridentata* is unable to re-infect any host [37], most likely because it is a defective strain lacking symbiotic *nif* genes [37]. The actual position of Pti1 in G9 close to Mgl<sub>5</sub> showed that Pti1 has a genomic background similar to that of Mgl<sub>5</sub> and probably most true Elaeagnaceae and Rhamnaceae-infective strains but, for obvious reasons, this could not be confirmed by using *nifH*.

However, the actual phylogenetic position of G9/cluster 4 members within or close to the true Elaeagnaceae and Rhamnaceae-infective group is still questionable because this position may result from artifacts in tree reconstruction at the deep branching level caused by the well known long branch attraction phenomenon [5]. Actually, as explained above, larger genomes produce more fragments that in turn generate longer branches instead of smaller ones. It thus cannot be excluded that deep branching 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 clarify *Frankia* phylogeny definitively at the deepest levels.

#### Genomospecies assignment of uncultivated *Frankia* present in nodules

As numerous *Frankia* members are only available in nodules, since they have until now been refractory to *in vitro* culture,



**Fig. 3.** (A) AFLP–electrophoretic profiles yielded with PstI/AA/HhaI + 0 for the 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 AFLP analysis of cultured and uncultured *Frankia* strains. Numbers indicate significant bootstrap values for 1000 resamplings. Definition of scale bar required (e.g. percentage nucleotide substitutions).

they cannot be assigned to genomospecies by DDH because contaminant plant DNA would hamper precise analysis. Therefore, it was thought in this study that AFLP could be used instead, since this methodology targets precise DNA sequences (i.e. restriction sites flanking selective nucleotides), thus allowing plant- and bacterium-derived AFLP fragments to be discriminative and consequently overcoming the presence of plant DNA contaminants in the nodules.

In a first step, AFLP analyses were therefore performed 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, which did not significantly

interfere in the resulting phylogeny, could be explained by the very different G + C content between plant and *Frankia* genomes (i.e. about 44% and 70% for *Alnus glutinosa* and *Frankia*, respectively). AFLPs performed with DNAs from nodules and their pure culture inciting strains actually yielded AFLP patterns with 0.020 and 0.032 mismatches per site between ACN14aNOD and ACN14a, and Cc13NOD and Cc13, respectively. They were thus positioned very close together in the phylogenomic dendrograms (Fig. 3B). This small difference is in the range of unreliability reported above between AFLP duplicates obtained with the same strain, thus showing that the AFLP method is accurate for assessing the phylogenomic position of *Frankia* present in nodules.

In a second step, the AFLP method was used to define the phylogenomic position of culture refractory *Frankia*. In this experiment, field nodules were analyzed from *Alnus viridis* and *A. incana* hosting endophytic *Frankia*, which displayed a typical Sp+ phenotype known to repeatedly fail in *in vitro* culture (Moiroud, personal communication). AFLP patterns of the three Sp+ endophytes were found to be closely related, thus most likely belonging to the same genomospecies (greatest CGM = 0.046 mismatches per site). These endophytes appeared to be closely related but they probably differed from genomospecies G1. The CGM values were generally in a range that indicated an out position (0.085–0.120 mismatches per site), except with ACN14a and the AiGNOD endophyte (0.066 mismatches 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* [10,42,43], AFLP was found to be a suitable alternative to DDH for the assignment of strains to known or new genomospecies within the whole *Frankia* genus. It enabled us to assign new isolates to previously defined genomospecies. Moreover, the method also allowed genomospecies determination of culture refractory isolates. This indicated that a complete phylogeny of the genus could now be drawn up. The present data also further confirmed, 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 [3,47]. This co-evolutionary shaping was also confirmed for culture refractory Sp+ which infects *Alnus* plants, and was found to be 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*, *Datisceae* and *Ceanothus*.

## Acknowledgements

We thank D. Chapulliot (UMR555-Ecologie Microbienne, Lyon, France) for 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 the French Région Rhône-Alpes (Cluster 9 program).

## References

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

- [2] An, C.S., Riggsby, W.S., Mullin, B.C. (1985) Relationships of *Frankia* isolates based on deoxyribonucleic acid homology studies. *Int. J. Syst. Bacteriol.* 35, 140–146.
- [3] 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.
- [4] Benson, D.R., Silvester, W.B. (1993) Biology of *Frankia* strains, actinomycete symbionts of actinorhizal plants. *Microbiol. Rev.* 57, 293–319.
- [5] Bergsten, J. (2005) A review of long-branch attraction. *Cladistics* 21, 163–193.
- [6] 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.
- [7] 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.
- [8] 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.
- [9] Clerc, A., Manceau, C., Nesme, X. (1998) 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.
- [10] 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, 1657–1666.
- [11] 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.*
- [12] Faten, G., Nouiou, I., Chair, M., Boudabous, A., Gtari, M. (2010) 16S-23S rRNA Intergenic spacer region variability in the genus *Frankia*. *Microb. Ecol.* 60, 487–495.
- [13] 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.
- [14] 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.
- [15] 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.
- [16] Gevers, D., Cohan, F.M., Lawrence, J.G., Spratt, B.G., Coenye, T., Feil, E.J., Stackebrandt, E., Van de Peer, Y., Vandamme, P., Thompson, F.L., Swings, J. (2005) Opinion: re-evaluating prokaryotic species. *Nat. Rev. Microbiol.* 3, 733–739.
- [17] 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.
- [18] Hahn, D. (2007) Polyphasic taxonomy of the genus *Frankia*. In: Pawlowski, K., Newton, W.E. (Eds.), *Nitrogen-fixing Actinorhizal Symbiosis*, pp. 25–47.
- [19] Huguet, V., Batzli, J.M., 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. *Appl. Environ. Microbiol.* 67, 2116–2122.
- [20] Huguet, V., Mergeay, M., Cervantes, E., Fernandez, M.P. (2004) 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.
- [21] Jamann, S., Fernandez, M.P., Moiroud, A. (1992) Genetic diversity of Elaeagnaceae-infective *Frankia* strains isolated from various soils. *Acta Oecolog.* 13, 395–405.
- [22] 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.
- [23] 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.
- [24] 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.
- [25] Koopman, W.J. (2005) Phylogenetic signal in AFLP data sets. *Syst. Biol.* 54, 197–217.
- [26] Lechevalier, M.P. (1986) Catalog of *Frankia* strains. *Actinomycetes* 19, 131–162.
- [27] Lechevalier, M.P. (1984) The taxonomy of the genus *Frankia*. *Plant Soil* 78, 1–6.
- [28] 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.
- [29] 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.
- [30] 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.
- [31] Meudt, H.M., Clarke, A.C. (2007) Almost forgotten or latest practice? AFLP applications, analyses and advances. *Trends Plant Sci.* 12, 106–117.
- [32] 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.
- [33] Mortimer, P., Arnold, C. (2001) FAFLP: last word in microbial genotyping? *J. Med. Microbiol.* 50, 393–395.
- [34] Mougél, C., Thioulouse, J., Perrière, G., Nesme, X. (2002) A mathematical method for determining genome divergence and species delineation using AFLP. *Int. J. Syst. Evol. Microbiol.* 52, 573–586.
- [35] Murry, M.A., Fontaine, M.S., Torrey, J.G. (1984) Growth kinetics and nitrogenase induction in *Frankia* sp. HFPAr13 grown in batch culture. *Plant Soil* 78, 61–78.
- [36] 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.
- [37] 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.
- [38] Nichols, R. (2001) Gene trees and species trees are not the same. *Trends Ecol. Evol.* 16, 358–364.
- [39] Normand, P., Lapiere, P., Tisa, L., Gogarten, J.P., Alloisio, N., Bagnarol, E., Bassi, C.A., Berry, A.M., Bickhard, D.M., Choise, N., Couloux, A., Cournoyer, B., Cruveiller, S., Daubin, V., Demange, N., Francino, M.P., Goltzman, E., Huang, Y., Kopp, O., Labarre, L., Lapidus, A., Lavire, C., Marechal, J., Martinez, M., Mastronunzio, J.E., Mullin, B.C., Niemann, J., Pujic, P., Rawnsley, T., Rouy, Z., Schenowitz, C., Sellstedt, A., Tavares, F., Tomkins, J.P., Vallenet, D., Valverde, C., Wall, L., Wang, Y., Medigue, C., Benson, D.R. (2007) Genome characteristics of facultatively symbiotic *Frankia* sp. strains reflect host range and host plant biogeography. *Genome Res.* 17, 7–15.
- [40] 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.
- [41] Portier, P., Costechareyre, D., Muller, D., Chapulliot, D., Oger, C., Perrière, 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.
- [42] Portier, P., Fischer-Le Saux, M., Mougél, C., Lerondelle, C., Chapulliot, D., Thioulouse, J., Nesme, X. (2006) Identification of genomic species in *Agrobacterium biovar 1* by AFLP genomic markers. *Appl. Environ. Microbiol.* 72, 7123–7131.
- [43] Rademaker, J.L.W., Hoste, B., Louws, F.J., Kersters, K., Swings, J., Vauterin, L., Vauterin, P., De Bruijn, F.J. (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.
- [44] Rombauts, S., Van de Peer, Y., Rouzé, P. (2003) AFLPinSilico simulating AFLP fingerprints. *Bioinformatics* 19, 776–777.
- [45] Shi, Y., Ruan, J. (1992) DNA base composition and homology values in the classification of some *Frankia* strains. *Acta Microbiol. Sin.* 32, 133–136.
- [46] 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.
- [47] Simonet, P., Navarro, E., Rouvier, C., Reddell, P., Zimpfer, J., Dommergues, Y., Bardin, R., Combarro, P., Hamelin, J., Domenach, A.M., Gourbiere, F., Prin, Y., Dawson, J.O., Normand, P. (1999) Co-evolution between *Frankia* populations and host plants in the family Casuarinaceae and consequent patterns of global dispersal. *Environ. Microbiol.* 1, 525–533.
- [48] Stackebrandt, E., Frederiksen, W., Garrity, G.M., Grimont, P.A., Kampfer, P., Maiden, M.C., Nesme, X., Rossello-Mora, R., Swings, J., Truper, H.G., Vauterin, L., Ward, A.C., Whitman, W.B. (2002) Report of the ad hoc committee for the re-evaluation of the species definition in bacteriology. *Int. J. Syst. Evol. Microbiol.* 52, 1043–1047.
- [49] 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.
- [50] Vos, P., Hogers, R., Bleeker, M., Reijans, M., Van de Lee, T., Hornes, M., Frijters, A., Pot, J., Peleman, J., Kuiper, M., Zabeau, M. (1995) A new technique for DNA fingerprinting. *Nucleic Acids Res.* 23, 4407–4414.
- [51] 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.
- [52] 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.