IVYSPRING INTERNATIONAL PUBLISHER

Journal of Genomics

2022; 10: 61-68. doi: 10.7150/jgen.75779

Research Paper

Draft Genomes of Symbiotic *Frankia* Strains AgB32 and AgKG'84/4 from Root Nodules of *Alnus Glutinosa* growing under Contrasted Environmental Conditions

Philippe Normand[1]✉, Petar Pujic[1], Danis Abrouk[1], Spandana Vemulapally[2], Trina Guerra[2], Camila Carlos-Shanley[2], Dittmar Hahn[2]

1. Université Claude-Bernard Lyon 1, Université de Lyon, UMR 5557 CNRS Ecologie Microbienne, Villeurbanne, Cedex 69622, France.
2. Texas State University, Department of Biology, 601 University Drive, San Marcos, TX 78666, USA.

✉ Corresponding author: Philippe Normand, Université Claude-Bernard Lyon 1, Université de Lyon, UMR 5557 CNRS Ecologie Microbienne, Villeurbanne, Cedex 69622, France. E-mail: philippe.normand@univ-lyon1.fr.

Received: 2022.06.03; Accepted: 2022.07.08; Published: 2022.08.08

Abstract

The genomes of two nitrogen-fixing *Frankia* strains, AgB32 and AgKG'84/4, were isolated from spore-containing (spore+) and spore-free (spore-) root nodules of *Alnus glutinosa*, but they did not sporulate upon reinfection. The two strains are described as representatives of two novel candidate species. Phylogenomic and ANI analyses indicate that each strain represents a novel species within cluster 1, with genome sizes of 6.3 and 6.7 Mb smaller than or similar to those of other cultivated *Alnus*-infective cluster 1 strains. Genes essential for nitrogen-fixation, clusters of orthologous genes, secondary metabolite clusters and transcriptional regulators analyzed by comparative genomic analyses were typical of those from *Alnus*-infective cluster 1 cultivated strains in both genomes. Compared to other cultivated *Alnus*-infective strains with large genomes, those of AgB32 and AgKG'84/4 had lost 380 or 409 genes, among which one *hup* cluster, one *shc* gene and the *gvp* cluster, which indicates genome erosion is taking place in these two strains.

Key words: *Frankia*, Actinorhizal symbiosis, genome, nitrogen-fixing frankiae, biosynthetic gene clusters

Introduction

Bacteria classified in the genus *Frankia* constitute a heterologous group of filamentous soil bacteria that can trigger the development of symbiotic root nodules on a range of host plants belonging to 25 genera of perennial, dicotyledonous, angiosperms [1-3]. Isolates have been classified into four distinct clusters, among which three comprise strains that fix atmospheric nitrogen (N_2), either in pure culture or in nodules, while cluster 4 frankiae for the most part do not fix N_2, except for one strain, and are often unable to fulfil Koch's postulates [4, 5]. Cluster 1 comprises strains infective on *Alnus* and *Casuarina*, with currently four described species and two candidate species [6]. The species have type strains deposited in culture collections as *Frankia alni* ACN14a^T [7], *F. torreyi* CpI1^T [8], *F. casuarinae* CcI3^T [7] and *F. canadensis* ARgP5^T [9]. Candidate species represent uncultured *Frankia* populations in root nodules of host plants, i.e. Candidatus *F. nodulisporulans* AgTrS, AgUmASt1 and AgUmASH1 [10] and Candidatus *F. alpina* AiOr, and AvVan [10] that have resisted all attempts at culture.

Several published works on genus *Frankia* using sub-cluster, OTU, group and genomospecies assignments did provide grounds permitting to affirm that cluster 1 is probably much more diverse than the four species and two candidate species described so far [4, 11-14]. This statement is supported by recent genome analyses of strains Ag45/Mut14 and AgPM24 as representatives of a yet undescribed species [15], and by comparative sequence analyses of amplicons of an actinobacteria-specific insertion in the 23S rRNA genes of frankiae that identified several strains clustering together but that are distinct from type strains of cluster 1 [16]. Strains AgB32 and AgKG'84/4

are two such strains, isolated from root nodules of *Alnus glutinosa* growing under contrasted environmental conditions at two locations in Germany about 350 km apart. Strain AgB32 was isolated from spore[+] root nodules of *Alnus glutinosa* of a forest ecotype that was interspersed with oak (*Quercus robur*) in an established riverside forest on a wet, but well aerated sandy loam in Bad Bentheim, Germany (52.320319, 7.159997) [17]. Strain AgKG'84/4 was isolated from spore[-] root nodules of *A. glutinosa* of the pioneer ecotype growing in a pure stand at a lake shore marsh in water-logged soil rich in organic material in Krems II-Goels, Germany (53.989103, 10.360772) [17]. Both strains had previously been identified as members of cluster 1, representing a subcluster designated as subgroup I [18] or cluster 1d [16]. In order to assess the viability of the previous amplicon-based analysis and to potentially amend and refine the species diversity of cluster 1 frankiae, we used whole genome sequence analyses trying to affirm the potential of strains AgB32 and AgKG'84/4 for the description of new species.

Materials and Methods

Sample preparation

Frankia strains AgB32 and AgKG'84/4 were grown from stocks preserved in 20% vol/vol glycerol at -80 °C since 2003 in Defined Propionate Medium (DPM) containing propionate and NH_4Cl as C and N source, respectively (19), at 30 °C for two weeks. Cells were harvested by centrifugation (15,000 × g, 5 min) and homogenized through sonication (10 s at 20% output in a S-450 sonifier, Branson Ultrasonics, Danbury, CT) [20]. DNA was extracted from cell pellets after an additional centrifugation step using the SurePrep™ Soil DNA Isolation Kit (Fisher Scientific, Houston, TX) [21], and concentrations measured with a Qubit® 2.0 Fluorometer (Life Technologies, Carlsbad, USA). Library preparation and sequencing using the Illumina tagmentation protocol and the NextSeq Illumina platform (2 × 150 bp) using standard protocols were done at the Microbial Genomics Sequencing Center, Pittsburgh, PA, USA.

Genome assembly

Default settings of fastp were used to filter and trim sequence reads [22], with reads with average %GC<54 removed using bbduk (https://jgi.doe.gov/data-and-tools/bbtools/bb-tools-user-guide/). SPAdes 3.13.0 was used to assemble genomes [23] and QUAST to check their quality [24]. Genome completeness was estimated using the lineage workflow (lineage_set) CheckM v1.0.18 with default values [25].

Comparative genomic analysis

Assembled genomes of strains AgB32 and AgKG'84/4 as well as *Frankia* genomes of type strains of all described species and other selected genomes were selected for Average Nucleotide Identity (ANI) comparisons [26] using the pyani platform with the b (Blast) setting ([27]; https://pyani.readthedocs.io). Genomes were further analyzed on the Mage platform [28] to compute clusters of orthologous genes (COGs) [29], to identify secondary metabolite clusters through antiSMASH [30] and to identify genes specific to or lost in the new genomes. A MASH distance matrix [31] was used to construct a phylogenetic tree using a rapid neighbour joining algorithm [32] on the Mage platform.

Results

Characteristics of the two *Frankia* genomes

The genomes of the two strains AgB32 and AgKG'84/4 were considered complete given their CheckM scores of 99.59% and 98.05%, respectively. The N50 were 55 309 and 112 139, respectively and the total length were 6 667 069 and 6 426 475. They were considered pure with contamination indices of 1.09 and 2.37, respectively. Genomes of AgB32 and AgKG'84/4 harbored 214 and 1,305 contigs with the largest contig being 223 506 nt and 54 816 nt, respectively. Their GC contents of 72.23 and 71.88% for AgB32 and AgKG'84/4, respectively (Table 1).

Phylogenetic analysis of *Frankia* spp

A phylogenetic tree generated from the MASH matrix with *Frankia* genomes of type strains revealed that the closest strains to AgB32 and AgKG'84/4 were members of cluster 1 (Figure 1). Average nucleotide identity (ANI) between strains AgB32 and AgKG'84/4 was 89%, indicating that they belong to two separate genospecies (Figure 2). ANI values at or below 80% were obtained for both strains in comparison with *Frankia* genomes of type strains of all described species (Figure 2). The ANI values with other cluster 1 genomes ranged from 79% (CcI3) to 81% (ACN14a), while 76-77% values were obtained with cluster 2 genomes, and 77-78% with cluster 3 and 4 genomes (Figure 2).

Analysis of functional genes in *Frankia* spp. isolates

All genes identified as playing a role in the symbiosis were found to be present in the genomes of AgB32 and AgKG'84/4, i.e. *nif, hup, suf, shc, cel, glx, bcsA* (Table 1). Furthermore, all genes that are more abundant in symbiotic lineages (clusters 1, 2 and 3)

than in non-symbiotic lineages (cluster 4) (*sodF, geoA, argF, accA, rhbE, dctA, phdA, tgsA, ddnB*) were also recovered in AgB32 and AgKG'84/4 (Table 1). Conversely, *gvp* that codes for gas vesicle proteins, one of the two *shc* genes and one of the two *hup* clusters that are found in cluster 1 strains were not found in the two genomes while the symbiotic cluster was maintained [33].

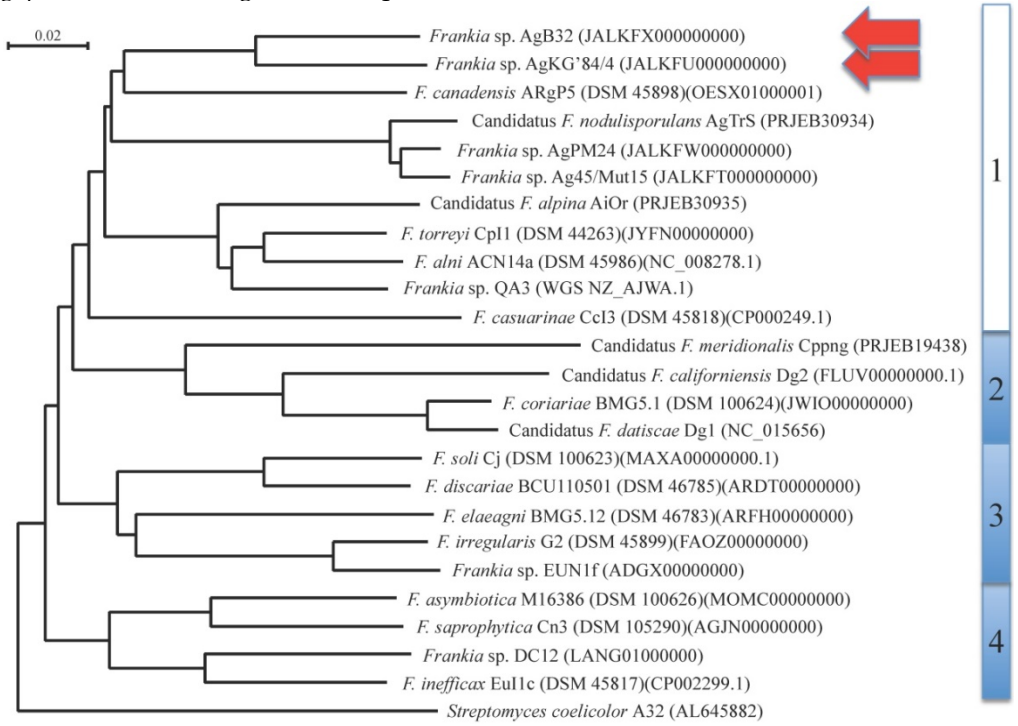

Figure 1. Phylogenetic tree based on comparative sequence analyses of complete genomes of *Frankia* species and candidate species, using *Streptomyces coelicolor* (AL645882) as outgroup. *Frankia* clusters 1 to 4 are indicated on the right. Scale units are substitutions per site. The two genomes described in the present study have red arrows.

	1-*F. alni*-ACN14a	1-*F. canadensis*-ARgP5	1-*F. torreyi*-CpI1	1-*Frankia*-QA3	1-*F. alpina*-AiOr	1-*Frankia*-AgKG'84/4	1-*Frankia*-AgB32	1-*Frankia*-Ag45/Mut15	1-*Frankia*-AgPM24	1-*F. nodulisporulans*-AgTrs	1c-*F. casuarinae*-CcI3	2-*F. californiensis*-Dg2	2-*F. coriariae*-BMG5-1	2-*F. datiscae*-Dg1	2-*F. meridionalis*-Cppng	3-*F. discariae*-BCU110501	3-*F. elaeagni*-BMG5-12	3-*F. irregularis*-G2	3-*F. soli*-Cj	3-*Frankia*-EAN1pec	3-*Frankia*-EUN1f	4-*F. asymbiotica*-B16386	4-*F. inefficax*-EuI1c	4-*F. saprophytica*-CN3	4-*Frankia*-DC12
1-*F. alni*-ACN14a	1.00	0.81	0.91	0.89	0.88	0.81	0.81	0.79	0.79	0.80	0.81	0.77	0.77	0.77	0.76	0.77	0.77	0.77	0.77	0.77	0.77	0.77	0.76	0.77	0.77
1-*F. canadensis*-ARgP5	0.81	1.00	0.81	0.81	0.81	0.81	0.81	0.80	0.80	0.80	0.79	0.76	0.76	0.77	0.76	0.78	0.77	0.77	0.77	0.78	0.77	0.77	0.77	0.77	0.77
1-*F. torreyi*-CpI1	0.91	0.81	1.00	0.90	0.88	0.81	0.81	0.79	0.79	0.80	0.81	0.77	0.77	0.77	0.76	0.78	0.77	0.77	0.78	0.77	0.77	0.77	0.76	0.77	0.77
1-*Frankia*-QA3	0.89	0.82	0.90	1.00	0.89	0.81	0.81	0.80	0.80	0.80	0.81	0.77	0.77	0.77	0.76	0.78	0.78	0.77	0.78	0.78	0.77	0.77	0.77	0.77	0.77
1-*F. alpina*-AiOr	0.88	0.81	0.89	0.90	1.00	0.81	0.81	0.80	0.79	0.80	0.80	0.77	0.77	0.77	0.76	0.78	0.77	0.77	0.77	0.78	0.77	0.77	0.77	0.77	0.77
1-*Frankia*-AgKG'84/4	0.81	0.81	0.81	0.81	0.81	1.00	0.89	0.80	0.80	0.80	0.79	0.76	0.76	0.76	0.75	0.77	0.77	0.77	0.77	0.77	0.77	0.77	0.77	0.77	0.77
1-*Frankia*-AgB32	0.81	0.81	0.81	0.81	0.81	0.89	1.00	0.80	0.80	0.80	0.79	0.76	0.76	0.77	0.76	0.78	0.77	0.77	0.78	0.77	0.77	0.77	0.77	0.77	0.77
1-*Frankia*-Ag45/Mut15	0.80	0.80	0.79	0.80	0.79	0.80	0.80	1.00	0.97	0.97	0.78	0.76	0.76	0.76	0.75	0.77	0.77	0.77	0.77	0.77	0.77	0.77	0.77	0.76	0.76	0.77
1-*Frankia*-AgPM24	0.80	0.80	0.79	0.80	0.79	0.80	0.80	0.97	1.00	0.97	0.78	0.76	0.76	0.76	0.75	0.77	0.77	0.76	0.77	0.77	0.77	0.76	0.76	0.76	0.76
1-*F. nodulisporulans*-AgTrs	0.80	0.80	0.80	0.80	0.80	0.80	0.80	0.97	0.97	1.00	0.79	0.76	0.76	0.77	0.76	0.77	0.77	0.77	0.77	0.77	0.77	0.77	0.77	0.77	0.77
1c-*F. casuarinae*-CcI3	0.81	0.80	0.81	0.81	0.81	0.79	0.79	0.78	0.78	0.79	1.00	0.77	0.77	0.77	0.76	0.78	0.78	0.78	0.78	0.78	0.77	0.77	0.77	0.77	0.77
2-*F. californiensis*-Dg2	0.77	0.76	0.77	0.77	0.76	0.76	0.76	0.76	0.76	0.76	0.77	1.00	0.86	0.86	0.82	0.77	0.76	0.76	0.77	0.77	0.76	0.76	0.76	0.76	0.76
2-*F. coriariae*-BMG5.1	0.77	0.77	0.77	0.77	0.77	0.76	0.77	0.76	0.76	0.77	0.77	0.86	1.00	0.96	0.80	0.77	0.77	0.76	0.77	0.77	0.76	0.76	0.76	0.76	0.76
2-*F. datiscae*-Dg1	0.77	0.77	0.77	0.77	0.77	0.77	0.78	0.77	0.77	0.77	0.77	0.87	0.96	1.00	0.81	0.77	0.78	0.76	0.77	0.78	0.77	0.76	0.76	0.76	0.76
2-*F. meridionalis*-Cppng	0.76	0.76	0.76	0.76	0.76	0.76	0.76	0.75	0.76	0.76	0.76	0.82	0.81	0.81	1.00	0.76	0.76	0.76	0.76	0.76	0.76	0.76	0.76	0.76	0.76
3-*F. discariae*-BCU110501	0.77	0.77	0.78	0.78	0.77	0.77	0.77	0.77	0.77	0.77	0.77	0.77	0.76	0.77	0.76	1.00	0.81	0.80	0.90	0.90	0.81	0.77	0.77	0.77	0.77
3-*F. elaeagni*-BMG5.12	0.77	0.77	0.77	0.77	0.77	0.77	0.77	0.77	0.77	0.77	0.77	0.76	0.77	0.77	0.76	0.81	1.00	0.82	0.82	0.82	0.82	0.77	0.77	0.77	0.77
3-*F. irregularis*-G2	0.77	0.77	0.77	0.77	0.77	0.77	0.77	0.76	0.76	0.77	0.77	0.76	0.76	0.76	0.75	0.80	0.81	1.00	0.80	0.80	0.94	0.77	0.77	0.77	0.77
3-*F. soli*-Cj	0.78	0.78	0.78	0.78	0.77	0.77	0.77	0.77	0.77	0.77	0.77	0.77	0.77	0.77	0.76	0.90	0.82	0.81	1.00	0.96	0.81	0.77	0.77	0.77	0.77
3-*Frankia*-EAN1pec	0.78	0.78	0.78	0.78	0.77	0.77	0.77	0.77	0.77	0.77	0.78	0.77	0.77	0.77	0.76	0.90	0.82	0.81	0.96	1.00	0.81	0.77	0.77	0.77	0.77
3-*Frankia*-EUN1f	0.77	0.77	0.77	0.77	0.77	0.77	0.77	0.77	0.77	0.77	0.77	0.76	0.76	0.76	0.76	0.81	0.81	0.94	0.80	0.80	1.00	0.77	0.77	0.77	0.77
4-*F. asymbiotica*-B16386	0.77	0.77	0.77	0.77	0.76	0.77	0.77	0.76	0.76	0.76	0.77	0.76	0.76	0.76	0.76	0.77	0.77	0.77	0.77	0.77	0.77	1.00	0.81	0.87	0.80
4-*F. inefficax*-EuI1c	0.76	0.77	0.76	0.76	0.76	0.77	0.76	0.76	0.76	0.76	0.76	0.76	0.76	0.76	0.75	0.77	0.77	0.77	0.77	0.76	0.77	0.81	1.00	0.81	0.88
4-*F. saprophytica*-CN3	0.76	0.77	0.76	0.76	0.76	0.76	0.76	0.76	0.76	0.76	0.76	0.76	0.76	0.76	0.75	0.77	0.77	0.77	0.77	0.77	0.77	0.87	0.81	1.00	0.80
4-*Frankia*-DC12	0.77	0.77	0.77	0.77	0.77	0.77	0.77	0.76	0.76	0.77	0.77	0.76	0.76	0.76	0.76	0.77	0.77	0.77	0.77	0.77	0.77	0.81	0.88	0.81	1.00

Figure 2. Heatmap matrix of Average Nucleotide Identity (ANI) comparisons (in percent) for the *Frankia* genomes of type strains of described species using the pyani platform with the b (Blast) setting [27]; https://pyani.readthedocs.io). The two genomes described in the present study are highlighted in grey. Those ANI values above the 95% threshold are highlighted in red. ANI values of clusters are boxed.

Table 1. Basic genome characteristics (G+C%, genome length, number of CDS, number of secondary metabolite clusters, presence of selected genes, # of contigs and references) of *Frankia* strains AgB32 and AgKG'84/4 compared to those of type strains of *Frankia* species in clusters 1 to 4

	Cluster 1									Cluster 2	Cluster 3					Cluster 4			
Strain	ACN14a^T	ARgP5^T	CpI1^T	QA3	AgB32	AgKG'84/4	Ag45/Mut 15	AgPM24	CcI3^T	BMG5.1^T	BCU110501^T	BMG5.12^T	G2^T	Cj^T	EUN1f	M16386^T	Eu1c^T	Cn3^T	DC12
	alni	*canadensis*	*torreyi*						*casuarinae*	*coriariae*	*discariae*	*elaeagni*	*irregularis*	*soli*		*asymbiotica*	*inefficax*	*saprophytica*	
G+C content (mol%)	72.8	72.4	72.4	72.6	72.22	72.13	71.37	71.35	70.1	71	72.3	71.7	70.9	71.1	70.82	71.93	72.3	71.8	71.93
Genome length (nt)	7497934	7730285	7624758	7590853	6709935	6513357	6443382	6672691	5433628	5795263	7891711	7589313	9537992	8061539	9322173	9435764	8815781	9978592	6884336
# CDS	6714	7500	7201	7307	6364	6122	6088	6370	5593	6487	7567	6977	8663	8108	9428	8884	8099	9262	6630
secondary metabolite clusters*	27	33	28	33	42	30	29	38	26	22	36	35	37	30	33	29	23	28	15
*nifH***	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0
shc	2	2	2	2	1	1	2	2	1	2	2	2	2	2	2	2	2	2	2
hupL	2	2	2	2	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1
sufD	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
celA1	2	2	2	0	2	2	2	2	0	1	1	0	1	1	1	0	0	0	1
glxA	1	1	1	1	1	1	1	1	0	1	1	0	1	1	1	0	0	0	1
bcsA	1	1	1	0	1	1	1	1	0	1	1	0	0	1	1	1	1	1	1
gvpJ	1	1	1	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1
sodF	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0	0	0
geoA	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0	0
argG	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0
accA	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0
can	2	2	2	2	2	2	2	2	2	2	2	2	2	1	2	0	1	0	0
rhbE	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0
lac	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	0	1	1	1
phdA	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0
dctA	1	1	1	1	1	1	1	1	1	0	1	0	1	1	1	0	0	0	0
tgsA	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	0	1	0
ddnB	1	1	1	1	1	0	1	1	0	0	1	1	1	1	1	1	0	0	0
mopB	1	1	2	2	2	2	2	2	1	1	2	2	2	1	2	0	0	0	0
qorB	1	2	1	1	1	0	0	0	0	0	1	1	1	1	1	0	0	0	0
glbN	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	0	1	0	1
# contigs	1	568	153	120	274	342	157	230	1	116	207	139	83	289	396	174	1	2	1
Reference	(48)	(9)	(52)	(53)	This study	This study	(15)	(15)	(48)	(54)	(55)	(56)	(57)	(58)	(48)	(59)	(39)	(48)	(60)

Table 2. COG characteristics of *Frankia* strains AgB32 and AgKG'84/4 compared to those of type strains of *Frankia* species in clusters 1 to 4

	Cluster 1									Cluster 2	Cluster 3					Cluster 4			
Strain	ACN14a^T	ARgP5^T	CpI1^T	QA3	AgB32	AgKG'84/4	Ag45/Mut15	AgPM24	CcI3^T	BMG5.1^T	BCU110501^T	BMG5.12^T	G2^T	Cj^T	EUN1f	M16386^T	Eu1c^T	Cn3^T	DC12
COG[1]	*alni*	*canadensis*	*torreyi*						*casuarinae*	*coriariae*	*discariae*	*elaeagni*	*irregularis*	*soli*		*asymbiotica*	*inefficax*	*saprophytica*	
D	56	66	75	64	64	51	56	61	57	80	65	63	80	66	78	63	62	64	67
M	241	189	253	236	216	212	225	241	207	203	292	259	297	248	311	299	258	266	255
N	19	15	26	22	17	13	12	17	12	30	20	16	28	21	29	16	20	11	17
O	181	134	181	190	143	150	133	140	147	149	200	165	176	200	195	177	173	200	149
T	325	226	320	326	318	281	291	290	232	253	405	336	436	400	418	415	405	494	282
U	42	38	50	38	45	54	45	50	48	50	54	53	66	56	64	53	52	52	50
V	94	74	86	102	83	76	77	81	60	78	107	84	117	126	110	130	113	153	113
J	212	226	212	257	222	219	209	212	202	243	207	197	203	219	226	243	241	247	232
K	565	402	594	646	537	507	509	525	369	409	739	577	778	688	755	785	809	945	520
L	270	254	351	356	332	266	308	319	433	289	613	398	398	518	468	380	286	409	399
C	435	323	455	472	348	355	346	347	256	362	492	394	527	451	530	555	507	589	332
E	523	386	482	534	440	463	452	451	335	396	577	461	630	516	623	670	661	704	447
F	111	82	104	108	91	100	96	94	94	92	107	94	103	101	97	129	116	114	107
G	326	274	321	342	307	298	289	297	233	249	418	326	372	360	428	450	426	488	302
H	192	149	186	187	181	186	170	184	174	173	187	177	192	181	182	188	186	208	163
I	432	258	400	460	299	331	296	303	191	297	513	412	643	405	619	586	624	619	313
P	311	243	323	332	274	293	307	313	210	293	381	298	408	343	387	402	394	427	278
Q	376	226	368	371	323	331	304	339	197	320	488	369	565	417	550	531	534	569	256
R	1009	704	1005	1059	863	869	814	836	619	682	1216	969	1323	1064	1280	1343	1332	1508	865
S	301	226	315	286	281	268	258	278	223	243	323	297	336	328	338	341	334	375	284

[1]class: **D**: Cell cycle control, cell division, chromosome partitioning; **M**: Cell wall/membrane/envelope biogenesis; **N**: Cell motility; **O**: Posttranslational modification, protein turnover, chaperones; **T**: Signal transduction mechanisms; **U**: Intracellular trafficking, secretion, and vesicular transport; **V**: Defense mechanisms; **J**: Translation, ribosomal structure and biogenesis; **K**: Transcription; **L**: Replication, recombination and repair; **C**: Energy production and conversion; **E**: Amino acid transport and metabolism; **F**: Nucleotide transport and metabolism; **G**: Carbohydrate transport and metabolism; **H**: Coenzyme transport and metabolism; **I**: Lipid transport and metabolism; **P**: Inorganic ion transport and metabolism; **Q**: Secondary metabolites biosynthesis, transport and catabolism; **R**: General function prediction only; **S**: Function unknown.

The COG computation showed values for AgB32 and AgKG'84/4 characteristic of other *Alnus*-infective cluster 1 strains with a low number of categories "N" (Cell motility), and "P" (Inorganic ion transport and metabolism) (Table 2). These results are similar for the antiSMASH computation that showed AgB32 and AgKG'84/4 to have values characteristic of other *Alnus*-infective cluster 1 strains with a high number of T1PKS and NRPS (Table 3). T1PKS and NRPS typically code for antibiotics and a high number of such clusters is evocative of a good capacity for keeping other soil microbes at bay. The numbers of

transcriptional regulators were on the whole comparable to other strains with a low number of ArsR, and LuxR regulators (Table 4).

A search for genes present in *F. alni* ACN14a, *Frankia* sp. QA3, *F. torreyi* CpI1 and *F. canadensis* ARgP5 but absent in AgB32 and AgKG'84/4 yielded 380 or 409 hits, respectively among which an alkane sulfonate, a acetyl/propionyl CoA carboxylase locus, an uptake hydrogenase locus, a dicarboxylate transporter, a Hup locus, the GVP locus, several transporters (Table S1). Conversely, there were 565 genes present in both AgB32 and AgKG'84/4 but absent in *F. alni* ACN14a, *Frankia* sp. QA3, *F. torreyi* CpI1 and *F. canadensis* ARgP5, of which about half (277) were of unknown function.

Table 3. Number of secondary metabolites clusters (antiSMASH) of *Frankia* strains AgB32 and AgKG'84/4 compared to those of cultivated type strains of *Frankia* species in clusters 1 to 4

	Cluster 1									Cluster 2	Cluster 3					Cluster 4			
Strain	ACN14a^T	ARgP5^T	CpI1^T	QA3	AgB32	AgKG'84/4	Ag45/Mut15	AgPM24	CcI3^T	BMG5.1^T	BCU110501^T	BMG5.12^T	G2^T	Cj^T	EUN1f	M16386^T	EuI1c^T	Cn3^T	DC12
	alni	*canadensis*	*torreyi*						*casuarinae*	*coriariae*	*discariae*	*elaeagni*	*irregularis*	*soli*		*asymbiotica*	*inefficax*	*saprophytica*	
t1PKS[1]	6	9	8	8	10	13	9	11	1	6	16	13	6	9	9	6	5	2	1
t2PKS	1	3	1	3	2	1	1	1	2	2	1	1	2	1	2	1	2	1	1
t3PKS	1	1	1	1	1	1	1	1	0	2	0	1	1	1	1	3	1	1	2
otherKS	4	4	3	3	5	4	3	5	4	1	4	3	6	4	6	2	1	2	1
t1pks-NRPS	1	0	1	0	1	0	1	2	1	0	1	0	0	0	1	0	0	0	0
NRPS	3	6	2	2	6	8	6	6	0	1	1	2	9	5	5	4	2	7	1
terpene	5	3	5	5	3	6	4	4	4	3	4	4	3	5	4	5	4	4	3
lanthipeptide	1	1	1	3	2	1	0	3	6	2	4	3	2	1	3	1	2	1	2
bacteriocin	2	1	2	2	2	3	1	2	1	1	2	2	2	2	0	3	1	2	0
siderophore	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	0	0	0	0
lassopeptide	1	1	1	2	1	1	1	1	0	1	0	0	1	0	0	1	2	1	2
betalactone	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	1	2	0	1
thiopeptide	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0
butyrolactone	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
phosphonate	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
arylpolyene	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
nucleoside	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
ladderane	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
oligosaccharide	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
resorcinol	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
LAP	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0
other	0	2	2	3	2	3	1	0	4	1	1	5	1	1	1	2	1	5	1
Total/strain	27	33	28	33	38	43	29	38	26	22	36	35	37	30	33	29	23	28	15

[1]tnPKS is type "n" Polyketide Synthase;
NRPS is Non Ribosomal Peptide Synthase, LAP is Linear Azole/azoline-containing Peptide.

Table 4. Number of transcriptional regulators of *Frankia* strains AgB32 and AgKG'84/4 compared to those of type strains of *Frankia* species in clusters 1 to 4

	Cluster 1									Cluster 2	Cluster 3					Cluster 4			
Strain	ACN14a^T	ARgP5^T	CpI1^T	QA3	AgB32	AgKG'84/4	Ag45/Mut15	AgPM24	CcI3^T	BMG5.1^T	BCU110501^T	BMG5.12^T	G2^T	Cj^T	EUN1f	M16386^T	EuI1c^T	Cn3^T	DC12
Class[1]	*alni*	*canadensis*	*torreyi*						*casuarinae*	*coriariae*	*discariae*	*elaeagni*	*irregularis*	*soli*		*asymbiotica*	*inefficax*	*saprophytica*	
AraC	9	9	10	16	9	13	6	6	2	5	15	13	17	16	17	20	22	21	6
ArsR	9	6	5	1	14	13	7	6	6	5	4	4	11	6	9	9	16	8	8
AsnC	3	2	2	4	3	4	4	3	3	2	3	3	3	3	4	5	5	5	3
CRP	4	2	1	1	4	3	4	4	2	3	3	3	5	2	5	3	5	2	3
DeoR	4	1	0	0	4	4	1	2	0	0	2	1	0	2	0	2	2	2	1
DtxR	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
FurC	2	3	3	4	3	3	3	3	2	2	5	4	5	5	4	5	4	4	5
GntR	25	19	10	20	12	15	7	5	6	8	21	12	19	19	24	27	35	30	11
IclR	3	6	4	9	7	8	4	3	2	1	4	7	12	6	12	6	13	11	7
LuxR	10	19	19	36	40	36	10	14	20	15	22	9	18	15	40	18	64	29	14
LysR	18	16	12	22	18	18	11	10	5	5	14	10	20	13	17	24	20	22	13
MarR	21	19	13	33	23	20	16	15	15	23	18	20	31	25	30	32	33	35	19
MerR	8	17	9	22	19	18	10	10	12	4	13	12	15	13	17	16	19	18	7

[1]class: **AraC**: arabinose regulator; **ArsR**: arsenic resistance; **AsnC**: asparagine synthase regulator; **CRP**: cyclic AMP receptor protein (catabolite repression); **DeoR**: deoxyribonucleoside synthesis operon regulator; **DtxR**: diphtheria toxin repressor; **FurC**: ferric uptake regulator; **GntR**: gluconate regulator; **IclR**: isocitrate lyase regulator; **LuxR**: quorum-sensing luminescence regulator; **LysR**: lysine regulator; **MarR**: Multiple antibiotic resistance regulator; **MerR**: mercury resistance regulator; **TetR**: Tetracycline repressor; **WhiB**: regulation of morphological differentiation.

Discussion

The genus *Frankia* has been scantily described for many years because of difficulties to isolate and grow frankiae in pure culture, a major prerequisite for the description of strains [34, 35]. Some populations to this day have even resisted isolation attempts so far [36]. Differentiation of isolates has also been hampered by the availability of few distinguishing features between populations [14]. Starting in 2007, new developments in whole genome sequencing techniques have overcome these difficulties and resulted in the determination of genome sequences of three *Frankia* isolates [37], and ultimately even of uncultured *Frankia* populations in root nodules [38]. Comparative analyses of whole genome sequences between *Frankia* populations have resulted in the description of twelve species and five candidate species for uncultured populations so far [6]. These numbers were based on the availability of 37 genomes [39], a number that is increasing regularly [15, 40]. Comparative sequences analyses of whole genomes and metrics such as ANI [26] or dDDH [41] are now used as foundation for the description of microbial genera, species and subspecies.

Members of the genus *Frankia* have been assigned into four clusters, numbered 1 to 4, within the genus [4]. These assignments have proven quite solid over the years, with cluster 1 in particular found to remain coherent with all *Alnus*-infective symbiotic strains. Cluster 1c with *Casuarina*-infective strains remains at the root of this cluster with several distinguishing features such as the lack of vesicles in nodules, a host-derived hemoglobin protection against oxygen and a distinct host range [6]. *Alnus*-infective symbiotic strains have been described initially on the basis of DNA/DNA homology as quite close to one another [14] but the full extent of diversity has slowly emerged with studies targeting new cultured strains and uncultured frankiae from specific environments [38, 42-47].

Genomes of *Alnus*-infective symbiotic strains have initially been found to be quite large at 7.5 Mb with several ancient duplicated genes such as the *shc* gene coding for the synthesis of hopanoid lipids [48], the *hup* genes coding for hydrogen uptake for the recycling of hydrogen derived from nitrogenase [33], the *cel* coding for cellulases [49], the *can* coding for the carbonic anhydrase necessary for feeding short chain fatty acids (SCFA) into the tricarboxylic acid (TCA) cycle or the *kor* genes coding for 2-oxoglutarate ferredoxin oxidoreductase that connects the TCA cycle with nitrogenase (with the nitrogen-fixation process) [50, 51]. Some of these duplications have been found to be lost in lineages with smaller genome size as is the case for Ag45/Mut15 and AgPM24 [15]. It appears the genomes of strains AgB32 and AgKG'84/4 are also undergoing a parallel process of genome erosion. This process is similar with some of the genes lost in common such as *hup* but also other genes such as *shc* only lost in AgB32 and AgKG'84/4.

AgB32 and AgKG'84/4 are two distinct lineages with an ANI of 89%, well below the threshold of 95 set by Goris [26] to delineate species but markedly above the 80% average between other *Alnus*-infective cluster 1 species. This would indicate the two strains should constitute two distinct species yet sharing many features due to a recent common ancestry.

Supplementary Material

Supplementary table.
https://www.jgenomics.com/v10p0061s1.txt

Acknowledgements

The authors are indebted to the Graduate College (Doctoral Research Support Fellowship to S. Vemulapally), and the Department of Biology at Texas State University for financial support.

Competing Interests

The authors have declared that no competing interest exists.

References

1. Schwintzer CR, Tjepkema JD. The biology of *Frankia* and actinorhizal plants, Academic Press, San Diego, 1990.
2. Huss-Danell K. Actinorhizal symbioses and their N-2 fixation. New Phytol. 1997;136: 375-405.
3. Benson DR, Silvester WB. Biology of *Frankia* strains, actinomycete symbionts of actinorhizal plants. Microbiol Rev. 1993;57: 293-319.
4. Normand P, Orso S, Cournoyer B, Jeannin P, Chapelon C, Dawson J, et al. Molecular phylogeny of the genus *Frankia* and related genera and emendation of the family Frankiaceae. Int J Syst Bacteriol. 1996;46: 1-9.
5. Gtari M, Nouioui I, Sarkar I, Ghodhbane-Gtari F, Tisa LS, Sen A, et al. An update on the taxonomy of the genus *Frankia* Brunchorst, 1886, 174^{AL}. Anton Leeuw. 2019;112: 5-21.
6. Normand P, Fernandez MP. *Frankia* Brunchorst 1886, 174^{AL}. In: Whitman WB, Rainey FA, Kämpfer P, Trujillo ME, DeVos P, Hedlund B, Dedysh S. (Eds.) Bergey's Manual of Systematics of Archaea and Bacteria, 2019. 10.1002/9781118960608.gbm00042.pub2.
7. Nouioui I, Ghodhbane-Gtari F, Montero-Calasanz MD, Goker M, Meier-Kolthoff JP, Schumann P, et al. 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. 2016;66: 5201-5210.
8. Nouioui I, Ghodhbane-Gtari F, Jando M, Tisa LS, Klenk HP, Gtari M. *Frankia torreyi* sp. nov., the first actinobacterium of the genus *Frankia* Brunchorst 1886, 174^{AL} isolated in axenic culture. Anton Leeuw. 2019;112: 57-65.
9. Normand P, Nouioui I, Pujic P, Fournier P, Dubost A, Schwob G, et al. *Frankia canadensis* sp nov., isolated from root nodules of *Alnus incana* subspecies *rugosa*. Int J Syst Evol Microbiol. 2018;68: 3001-3011.
10. Herrera-Belaroussi A, Normand P, Pawlowski K, Fernandez MP, Wibberg D, Kalinowski J, et al. Candidatus *Frankia nodulisporulans* sp. nov., an *Alnus glutinosa*-infective *Frankia* species unable to grow in pure culture and able to sporulate in-planta. Syst Appl Microbiol. 2020; 43(6):126134.
11. Hahn D. Polyphasic taxonomy of the genus *Frankia*. In: Pawlowski K, Newton WE (eds), Nitrogen-fixing actinorhizal symbioses. Springer Verlag, Berlin, Germany, 2007; pp. 25-48.

12. Hahn D, Mirza B, Benagli C, Vogel G, Tonolla M. Typing of nitrogen-fixing *Frankia* strains by matrix-assisted laser desorption ionization-time-of-flight (MALDI-TOF) mass spectrometry. Syst Appl Microbiol. 2011;34: 63-68.
13. Pozzi AC, Bautista-Guerrero HH, Abby SS, Herrera-Belaroussi A, Abrouk D, Normand P, et al. Robust *Frankia* phylogeny, species delineation and intraspeciesdiversity based on Multi-Locus Sequence Analysis (MLSA) and Single-Locus Strain Typing (SLST) adapted to a large sample size. Syst Appl Microbiol. 2018;41: 311-323.
14. Fernandez MP, Meugnier H, Grimont PAD, Bardin R. Deoxyribonucleic acid relatedness among members of the genus *Frankia*. Int J Syst Bacteriol. 1989;39: 424-429.
15. Normand P, Pujic P, Abrouk D, Vemulapally S, Guerra T, Carlos-Shanley C, et al. Draft genomes of nitrogen-fixing *Frankia* strains Ag45/Mut15 and AgPM24 isolated from root nodules of *Alnus glutinosa*. Journal of Genomics. 2022; 10.7150/jgen.74788.
16. Ben Tekaya S, Ganesan AS, Guerra T, Dawson JO, Forstner MRJ, Hahn D. SybrGreen- and TaqMan-based quantitative PCR approaches allow assessment of the abundance and relative distribution of *Frankia* clusters in soils. Appl Environ Microb. 2017; 83(5): e02833-16.
17. Hahn D, Dorsch M, Stackebrandt E, Akkermans ADL. Synthetic oligonucleotide probes for identification of *Frankia* strains. Plant Soil. 1989;118: 211-219.
18. Samant S, Huo T, Dawson JO, Hahn D. Abundance and relative distribution of *Frankia* host infection groups under actinorhizal *Alnus glutinosa* and non-actinorhizal *Betula nigra* trees. Microb Ecol. 2016;71: 473-481.
19. Meesters TM, van Genesen ST, Akkermans ADL. Growth, acetylene reduction activity and localization of nitrogenase in relation to vesicle formation in *Frankia* strains Cc1.17 and Cp1.2. Arch Microbiol. 1985;143: 137-142.
20. Mirza BS, Welsh A, Hahn D. Saprophytic growth of inoculated *Frankia* sp. in soil microcosms. FEMS Microbiol Ecol. 2007;62: 280-289.
21. Samant S, Sha Q, Iyer A, Dhabekar P, Hahn D. Quantification of Frankia in soils using SYBR Green based qPCR. Syst Appl Microbiol. 2012;35: 191-197.
22. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 2018;34: i884-i890.
23. Prjibelski A, Antipov D, Meleshko D, Lapidus A, Korobeynikov A. Using SPAdes De Novo Assembler. Curr Protoc Bioinformatics. 2020;70: e102.
24. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. Bioinformatics. 2013;29: 1072-1075.
25. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 2015;25: 1043-1055.
26. Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P, Tiedje JM. DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. Int J Syst Evol Microbiol. 2007;57: 81-91.
27. Pritchard L, Glover RH, Humphris S, Elphinstone JG, Toth IK. Genomics and taxonomy in diagnostics for food security: soft-rotting enterobacterial plant pathogens. Anal Methods-Uk. 2016;8: 12-24.
28. Vallenet D, Calteau A, Cruveiller S, Gachet M, Lajus A, Josso A, et al. MicroScope in 2017: an expanding and evolving integrated resource for community expertise of microbial genomes. Nucleic Acids Res. 2017;45: D517-D528.
29. Tatusov RL, Natale DA, Garkavtsev IV, Tatusova TA, Shankavaram UT, Rao BS, et al. The COG database: new developments in phylogenetic classification of proteins from complete genomes. Nucleic Acids Res. 2001;29: 22-28.
30. Medema MH, Blin K, Cimermancic P, de Jager V, Zakrzewski P, Fischbach MA, et al. antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. Nucleic Acids Res. 2011;39: W339-346.
31. Ondov BD, Treangen TJ, Melsted P, Mallonee AB, Bergman NH, Koren S, et al. Mash: fast genome and metagenome distance estimation using MinHash. Genome Biol. 2016;17: 132.
32. Simonsen M, Mailund T, Pedersen CNS, Rapid Neighbour-Joining, in: K.A. Crandall, J. Lagergren (Eds.) WABI 2008: Algorithms in Bioinformatics, Springer Verlag, Heidelberg, 2008, pp. 113-122.
33. Richau KH, Kudahettige RL, Pujic P, Kudahettige NP, Sellstedt A. Structural and gene expression analyses of uptake hydrogenases and other proteins involved in nitrogenase protection in *Frankia*. J Biosciences. 2013;38: 703-712.
34. Lechevalier MP. Taxonomy of the genus *Frankia* (Actinomycetales). Int J Syst Bacteriol. 1994;44: 1-8.
35. Lechevalier MP, Lechevalier HA, Systematics, isolation, and culture of *Frankia*. In: Schwintzer CR, Tjepkema JD (eds), The biology of *Frankia* and actinorhizal plants. Academic Press, New York, 1990, pp. 35-60.
36. Cotin-Galvan L, Pozzi AC, Schwob G, Fournier P, Fernandez MP, Herrera-Belaroussi A. In-planta sporulation capacity enhances infectivity and rhizospheric competitiveness of *Frankia* strains. Microbes Environ. 2016;31: 11-18.
37. Normand P, Queiroux C, Tisa LS, Benson DR, Rouy Z, Cruveiller S, et al. Exploring the genomes of *Frankia*. Physiol Plantarum. 2007;130: 331-343.
38. Bethencourt L, Vautrin F, Taib N, Dubost A, Castro-Garcia L, Imbaud O, et al. Draft genome sequences for three unisolated *Alnus*-infective *Frankia* Sp+ strains, AgTrS, AiOr and AvVan, the first sequenced *Frankia* strains able to sporulate *in-planta*. J Genomics. 2019;7: 50-55.
39. Tisa LS, Oshone R, Sarkar I, Ktari A, Sen A, Gtari M. Genomic approaches toward understanding the actinorhizal symbiosis: an update on the status of the *Frankia* genomes. Symbiosis. 2016;70: 5-16.
40. Carlos-Shanley C, Guerra T, Hahn D. Draft genomes of non-nitrogen-fixing *Frankia* strains. J Genomics. 2021;9: 68-75.
41. Auch AF, von Jan M, Klenk HP, Goker M. Digital DNA-DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison. Stand Genomic Sci. 2010;2: 117-134.
42. Welsh AK, Dawson JO, Gottfried GJ, Hahn D. Diversity of *Frankia* in root nodules of geographically isolated Arizona alders in central Arizona (USA). Appl Environ Microbiol. 2009;75: 6913-6918.
43. Welsh A, Mirza BS, Rieder JP, Paschke MW, Hahn D. Diversity of frankiae in root nodules of *Morella pensylvanica* grown in soils from five continents. Syst Appl Microbiol. 2009;32: 201-210.
44. Mirza BS, Welsh A, Rasul G, Rieder JP, Paschke MW, Hahn D. Variation in *Frankia* populations of the *Elaeagnus* host infection group in nodules of six host plant species after inoculation with soil. Microb Ecol. 2009;58: 384-393.
45. Pozzi AC, Bautista-Guerrero HH, Nouioui I, Cotin-Galvan L, Pepin R, Fournier P, et al. In-planta sporulation phenotype: a major life history trait to understand the evolution of *Alnus*-infective *Frankia* strains. Environ Microbiol. 2015;17: 3125-3138.
46. Pozzi ACM, Herrera-Belaroussi A, Schwob G, Bautista-Guerrero HH, Bethencourt L, Fournier P, et al. Proposal of 'Candidatus *Frankia alpina*', the uncultured symbiont of *Alnus alnobetula* and *A. incana* that forms spore-containing nitrogen-fixing root nodules. Int J Syst Evol Microbiol. 2020;70: 5453-5459.
47. Ben Tekaya S, Guerra T, Rodriguez D, Dawson JO, Hahn D. *Frankia* diversity in host-plant root nodules is independent of abundance or relative diversity of *Frankia* in corresponding rhizosphere soils. Appl. Environ. Microbiol. 2018;84:e02248-17.
48. Normand P, Lapierre P, Tisa LS, Gogarten JP, Alloisio N, Bagnarol E, et al. Genome characteristics of facultatively symbiotic *Frankia* sp strains reflect host range and host plant biogeography. Genome Res. 2007;17: 7-15.
49. Pujic P, Alloisio N, Fournier P, Roche D, Sghaier H, Miotello G, et al. Omics of the early molecular dialogue between *Frankia alni* and *Alnus glutinosa* and the cellulase synton. Environ Microbiol. 2019;21: 3328-3345.
50. Dodsworth JA, Cady NC, Leigh JA. 2-oxoglutarate and the PII homologues NifI(1) and NifI(2) regulate nitrogenase activity in cell extracts of *Methanococcus maripaludis*. Mol Microbiol. 2005;56: 1527-1538.
51. Huergo LF, Dixon R. The emergence of 2-Oxoglutarate as a master regulator metabolite. Microbiol Mol Biol Rev. 2015;79: 419-435.
52. Oshone R, Hurst SGt, Abebe-Akele F, Simpson S, Morris K, Thomas WK, et al. Permanent draft genome sequences for two variants of *Frankia* sp. strain CpI1, the first *Frankia* strain isolated from root nodules of *Comptonia peregrina*. Genome Announc. 2016; 4(1): e01588-15.
53. Sen A, Beauchemin N, Bruce D, Chain P, Chen A, Walston Davenport K, et al. Draft genome sequence of *Frankia* sp. strain QA3, a nitrogen-fixing actinobacterium isolated from the root nodule of *Alnus nitida*. Genome Announc. 2013;1: e0010313.
54. Gtari M, Ghodhbane-Gtari F, Nouioui I, Ktari A, Hezbri K, Mimouni W, et al. Cultivating the uncultured: growing the recalcitrant cluster-2 *Frankia* strains. Sci Rep-Uk. 2015;5: 13112; doi: 13110.11038/srep13112.
55. Wall LG, Beauchemin N, Cantor MN, Chaia E, Chen A, Detter JC, et al. Draft genome sequence of *Frankia* sp. strain BCU110501, a nitrogen-fixing actinobacterium isolated from nodules of *Discaria trinevis*. Genome Announc. 2013; 1(4): e00503-13.
56. Nouioui I, Beauchemin N, Cantor MN, Chen A, Detter JC, Furnholm T, et al. Draft genome sequence of *Frankia* sp. strain BMG5.12, a nitrogen-fixing actinobacterium isolated from Tunisian soils. Genome Announc. 2013; 1(4):e00468-13.
57. Nouioui I, Gtari M, Goker M, Ghodhbane-Gtari F, Tisa LS, Fernandez MP, et al. Draft genome sequence of *Frankia* strain G2, a nitrogen-fixing actinobacterium isolated from *Casuarina equisetifolia* and able to nodulate actinorhizal plants of the order Rhamnales. Genome Announc. 2016; 4(3):e00437-16.

58. Gtari M, Ghodhbane-Gtari F, Nouioui I. *Frankia soli* sp. nov., an actinobacterium isolated from soil beneath *Ceanothus jepsonii*. Int J Syst Evol Microbiol. 2020;70: 1203-1209.
59. Nouioui I, Gueddou A, Ghodhbane-Gtari F, Rhode M, Gtari M, Klenk HP. *Frankia asymbiotica* sp nov., a non-infective actinobacterium isolated from *Morella californica* root nodule. Int J Syst Evol Microbiol. 2017;67: 4897-4901.
60. Tisa LS, Beauchemin N, Cantor MN, Furnholm T, Ghodhbane-Gtari F, Goodwin L, et al. Draft genome sequence of *Frankia* sp. strain DC12, an atypical, noninfective, ineffective isolate from *Datisca cannabina*. Genome Announc. 2015; 3(4): e00889-15.