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

Philippe Normand¹, Petar Pujic¹, Danis Abrouk¹, Spandana Vemulapally², Trina Guerra², Camila Carlos-Shanley², Dittmar Hahn²

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.

© The author(s). This is an open access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/). See http://ivyspring.com/terms for full terms and conditions.

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 (N₂), either in pure culture or in nodules, while cluster 4 frankiae for the most part do not fix N₂, except for one strain, and are often unable to fulfill 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* ACN14aᵀ [7], *F. torreyi* Cpl11ᵀ [8], *F. casuarinae* Ccl3ᵀ [7] and *F. canadensis* ARgP5ᵀ [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 1 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 NH₄Cl 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). The 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% (Ccl3) 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, glm, 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, dtaA, phdA, tgsA, ddnB) were also recovered in AgB32 and AgKG’84/4 (Table 1). Conversely, gevp 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.

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    | AgB32     | AgKG’84/4 | AgB32     | AgKG’84/4 |
| G+C content (%) | 72.8 | 72.4 | 72.4 | 72.6 | 73.2 | 73.13 | 71.35 | 70.1 | 71 | 71.3 | 71.9 | 71.42 | 71.8 | 71.93 |
| Genome length (Mb) | 7490754 | 7730285 | 6745225 | 5980833 | 6709355 | 6135377 | 6443882 | 6872691 | 5435268 | 5795263 | 7881711 | 7538133 | 9537992 | 8061519 | 5322373 | 4517564 |
| Number of CDS | 6687 | 7567 | 6977 | 8663 | 8108 | 9428 | 8884 | 8099 | 9262 | 6630 | 6630 | 6630 | 6630 | 6630 | 6630 |
| # of contigs | 27 | 33 | 28 | 30 | 29 | 38 | 26 | 22 | 35 | 37 | 30 | 33 | 29 | 23 | 28 | 15 |

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    | AgB32     | AgKG’84/4 | AgB32     | AgKG’84/4 |
| COG        | Class 1: Cell cycle control, cell division, chromosome partitioning; Class 2: Cell wall/membrane/envelope biogenesis; Class 3: Cell motility; Class 4: Posttranslational modification, protein turnover, chaperones; Class 5: Signal transduction mechanisms; Class 6: Intra cellular trafficking, secretion, and vesicular transport; Class 7: Defense mechanisms; Class 8: Translation, ribosomal structure and biogenesis; Class 9: Transcription; Class 10: Replication, recombination and repair; Class 11: Energy production and conversion; Class 12: Amino acid transport and metabolism; Class 13: Nucleotide transport and metabolism; Class 14: Carbohydrate transport and metabolism; Class 15: Coenzyme transport and metabolism; Class 16: Lipid transport and metabolism; Class 17: Inorganic ion transport and metabolism; Class 18: Secondary metabolites biosynthesis, transport and catabolism; Class 19: General function prediction only; Class 20: Function unknown. |
| COG        | D: 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; J: Inorganic ion transport and metabolism; K: Secondary metabolites biosynthesis, transport and catabolism; L: General function prediction only; M: 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 Cp11 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 Cp11 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

| Strain          | ACN14a | ARgP5 | Cpl1 | QA3 | AgB32 | AgKG’84/4 | Ag5/Mar15 | AgPM24 | Cpl1 | AgB32 | BMG5.13 | BCU119017 | BMG5.12 | G2 | G1 | EUN1 | M1038B | Elt1 | Elt2 | Coa1 | DC12 |
|-----------------|--------|-------|------|-----|-------|-----------|-----------|--------|------|-------|---------|-----------|---------|----|----|-----|--------|------|------|------|------|
| C1              | 1      | 1     | 1    | 1   | 1     | 1         | 1         | 1      | 1    | 1     | 1       | 1         | 1       | 1  | 1  | 1   | 1      | 1    | 1    | 1    | 1    |
| C2              | 1      | 1     | 1    | 1   | 1     | 1         | 1         | 1      | 1    | 1     | 1       | 1         | 1       | 1  | 1  | 1   | 1      | 1    | 1    | 1    | 1    |
| C3              | 1      | 1     | 1    | 1   | 1     | 1         | 1         | 1      | 1    | 1     | 1       | 1         | 1       | 1  | 1  | 1   | 1      | 1    | 1    | 1    | 1    |
| C4              | 1      | 1     | 1    | 1   | 1     | 1         | 1         | 1      | 1    | 1     | 1       | 1         | 1       | 1  | 1  | 1   | 1      | 1    | 1    | 1    | 1    |
| C5              | 1      | 1     | 1    | 1   | 1     | 1         | 1         | 1      | 1    | 1     | 1       | 1         | 1       | 1  | 1  | 1   | 1      | 1    | 1    | 1    | 1    |
| C6              | 1      | 1     | 1    | 1   | 1     | 1         | 1         | 1      | 1    | 1     | 1       | 1         | 1       | 1  | 1  | 1   | 1      | 1    | 1    | 1    | 1    |
| C7              | 1      | 1     | 1    | 1   | 1     | 1         | 1         | 1      | 1    | 1     | 1       | 1         | 1       | 1  | 1  | 1   | 1      | 1    | 1    | 1    | 1    |
| C8              | 1      | 1     | 1    | 1   | 1     | 1         | 1         | 1      | 1    | 1     | 1       | 1         | 1       | 1  | 1  | 1   | 1      | 1    | 1    | 1    | 1    |
| C9              | 1      | 1     | 1    | 1   | 1     | 1         | 1         | 1      | 1    | 1     | 1       | 1         | 1       | 1  | 1  | 1   | 1      | 1    | 1    | 1    | 1    |
| C10             | 1      | 1     | 1    | 1   | 1     | 1         | 1         | 1      | 1    | 1     | 1       | 1         | 1       | 1  | 1  | 1   | 1      | 1    | 1    | 1    | 1    |
| C11             | 1      | 1     | 1    | 1   | 1     | 1         | 1         | 1      | 1    | 1     | 1       | 1         | 1       | 1  | 1  | 1   | 1      | 1    | 1    | 1    | 1    |
| C12             | 1      | 1     | 1    | 1   | 1     | 1         | 1         | 1      | 1    | 1     | 1       | 1         | 1       | 1  | 1  | 1   | 1      | 1    | 1    | 1    | 1    |

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

| Strain        | ACN14a | ARgP5 | Cpl1 | QA3 | AgB32 | AgKG’84/4 | Ag5/Mar15 | AgPM24 | Cpl1 | AgB32 | BMG5.13 | BCU119017 | BMG5.12 | G2 | G1 | EUN1 | M1038B | Elt1 | Elt2 | Coa1 | DC12 |
|---------------|--------|-------|------|-----|-------|-----------|-----------|--------|------|-------|---------|-----------|---------|----|----|-----|--------|------|------|------|------|
| AraC          | 9      | 10    | 16   | 9   | 13    | 6         | 6         | 2      | 5    | 15    | 13       | 17       | 16  | 17 | 17  | 22     | 20   | 21   | 26   | 17   |
| ArsR          | 9      | 6     | 5    | 14   | 13    | 7         | 6         | 2      | 5    | 4     | 11       | 6         | 9   | 8  | 8   | 9      | 16   | 8    | 8    | 8    |
| AsnC          | 9      | 6     | 5    | 14   | 13    | 7         | 6         | 2      | 5    | 4     | 11       | 6         | 9   | 8  | 8   | 9      | 16   | 8    | 8    | 8    |
| CRP           | 9      | 6     | 5    | 14   | 13    | 7         | 6         | 2      | 5    | 4     | 11       | 6         | 9   | 8  | 8   | 9      | 16   | 8    | 8    | 8    |
| DexR          | 9      | 6     | 5    | 14   | 13    | 7         | 6         | 2      | 5    | 4     | 11       | 6         | 9   | 8  | 8   | 9      | 16   | 8    | 8    | 8    |
| DnxR          | 9      | 6     | 5    | 14   | 13    | 7         | 6         | 2      | 5    | 4     | 11       | 6         | 9   | 8  | 8   | 9      | 16   | 8    | 8    | 8    |
| FurC          | 9      | 6     | 5    | 14   | 13    | 7         | 6         | 2      | 5    | 4     | 11       | 6         | 9   | 8  | 8   | 9      | 16   | 8    | 8    | 8    |
| GutR          | 9      | 6     | 5    | 14   | 13    | 7         | 6         | 2      | 5    | 4     | 11       | 6         | 9   | 8  | 8   | 9      | 16   | 8    | 8    | 8    |
| LysR          | 9      | 6     | 5    | 14   | 13    | 7         | 6         | 2      | 5    | 4     | 11       | 6         | 9   | 8  | 8   | 9      | 16   | 8    | 8    | 8    |
| MarR          | 9      | 6     | 5    | 14   | 13    | 7         | 6         | 2      | 5    | 4     | 11       | 6         | 9   | 8  | 8   | 9      | 16   | 8    | 8    | 8    |
| MerR          | 9      | 6     | 5    | 14   | 13    | 7         | 6         | 2      | 5    | 4     | 11       | 6         | 9   | 8  | 8   | 9      | 16   | 8    | 8    | 8    |

1:Class: AraC: arabinose regulator; ArsR: arsenic resistance; AsnC: asparagine synthase regulator; CRP: cyclic AMP receptor protein (catabolite repression); DexR: deoxyribonucleoside synthesis operon regulator; DnxR: diphtheria toxin repressor; FurC: ferric uptake regulator; GutR: glucose transporter; IelR: isocitrate lyase regulator; LysR: quorum-sensing luminescence regulator; MarR: multiple antibiotic resistance regulator; MerR: mercury resistance regulator; TetR: tetracycline repressor; WhiB: regulation of morphological differentiation.
Discussion

The genus *Frankia* has been scantly described for many years because of difficulties to isolate and grow frankia 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 frankia 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 Agd5/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: 203-319.
4. Normand P, Orso S, Cournoyer B, Jeannotin 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[4]. Anton Leeuw. 2019;112: 5-21.
6. Normand P, Fernandez MP. *Frankia* Brunchorst 1886, 174[4]. 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[4] 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, Wieberg 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):326134.
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* phylology, 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, Mieggnier 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, Vemulapalli S, Guerra T, Carlos-Shanley C, et al. Draft genomes of nitrogen-fixing *Frankia* stains Ag45/Mu115 and AgPM24 isolated from root nodules of *Alnus glutinosa*. Journal of Genomics. 2020;7:101505. https://www.jgenomics.com

16. Ben Tekaya S, Ganasan AS, Guerra T, Dawson JO, Forstner MRJ, Hahn D. SybrGreen- and Taq-Man-based quantitative PCR approaches allow assessment of the abundance and relative distribution of *Frankia* clusters in soils. Appl Environ Microbiol. 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. Microbiol. 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 Cc1.2. Arch Microbiol. 1985:143: 137-142.

20. Mirza BS, Welsh A, Hahn D. Saprophytic growth of inoculated *Frankia* sp. in soil microcosms. FEMS Microbiol. 2006;72: 280-289.

21. Samant S, Sha Q, Iyer A, Dhabekar P, Hahn D. Quantification of *Frankia* strains able to sporulate *Frankia* sp. in soil microcosms. FEMS Microbiol. 2007;62: 280-289.

22. Chen S, Zhou Y, Chen Y, Gu J, fastg: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 2018;34: 884-890.

23. Prijbelski 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 and taxonomy in diagnostics for food security: soft-rotting Methanococcus maripaludis extracts of *Frankia* sp. strain BCU110501, a *Frankia* actinobacterium isolated from the root nodule of *Comptonia peregrina*. Announc. 2013;1: e0010313.

25. Gish J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P, Morella pensylvanica. J Bacteriol. 1989;39: 424-429.

26. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. D517-D528.

27. Pritchard L, Glover RH, Humphris S, Elphinstone JG, Toth IK. Genomics trees. Microb Ecol. 2016;71: 1-17.

28. Vallenet D, Calteau A, Cruveiller S, Gachet M, Lajus A, Josso A, et al. Exploring the genomes of *Frankia* populations of the *Elaeogrus* host infection group in nodules of six host plant species after inoculation with soil. Microb Ecol. 2009;58: 384-393.

29. 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*-infected *Frankia* strains. Environ Microbiol. 2015;17: 3125-3138.

30. Welsh AK, Dawson JO, Gottfried CJ, Hahn D. Diversity of *Frankia* in root nodules of geographically isolated Arizona alders in central Arizona (USA). Appl Environ Microbiol. 2009;75: 6913-6918.

31. Welsh A, Mirza BS, Rieder JP, Paschke MW, Hahn D. Diversity of *Frankia* in root nodules of *Morella pensylvanica* grown in soils from five continents. Syst Appl Microbiol. 2009;32: 201-210.

32. Mirza BS, Welsh A, Rasul G, Rieder JP, Paschke MW, Hahn D. Variation in *Frankia* populations of the *Elaeogrus* host infection group in nodules of six host plant species after inoculation with soil. Microb Ecol. 2009;58: 384-393.

33. Ben Tekaya S, Guerra T, Rodriguez D, Dawson JO, Hahn D. Frankia diversity in plant root nodules is independent of abundance or relative diversity of *Frankia* in corresponding rhizosphere soils. Appl. Environ. Microbiol. 2018;84:e02248-17.

34. Normand P, Lapierre P, Tsai LS, Gogarten JP, Alloisio N, Bagaroul E, et al. Genome characteristics of facultatively symbiotic Frankia sp. strains reflect host range and host plant biogeography. Genome Res. 2007:17: 7-15.

35. Pujic P, Alloisio N, Fournier P, Roche D, Sghairer H, Miottello 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.

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, Quetroux T, 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, Garcia-Castro L, Imboud O, et al. Draft genome sequences for three unisolated *Alnus*-infected *Frankia* sp+ strains, AgTrS, Abk and AvVan, the first sequenced *Frankia* strains able to sporulate in planta. J Genomics. 2019;7: 50-55.

39. Tisa LS, Oshone R, Sarkar L, Ktari A, Sen A, Giarni M. Genomic approaches toward understanding the actinorhizal symbiosis: an update on the status of the *Frankia* genomics. Symbiosis. 2016:70: 5-16.

40. Claro-Shanley C, Guerra T, Hahn D. Draft genomes of non-nitrogen-fixing *Frankia* strains. J Genomics. 2021:9: 68-75.

41. Auss AF, van 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: D517-D528.

42. Welsh AK, Dawson JO, Gottfried CJ, Hahn D. Diversity of *Frankia* in root nodules of geographically isolated Arizona alders in central Arizona (USA). Appl Environ Microbiol. 2009;75: 6913-6918.
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.