



HAL
open science

Candidatus Frankia nodulisporulans sp. nov., an *Alnus glutinosa*-infective Frankia species unable to grow in pure culture and able to sporulate in-planta

Aude Herrera-Belaroussi, Philippe Normand, Katharina Pawlowski, Maria Fernandez, Daniel Wibberg, Jörn Kalinowski, Andreas Brachmann, Fede Berckx, Natuschka Lee, Jochen Blom, et al.

► To cite this version:

Aude Herrera-Belaroussi, Philippe Normand, Katharina Pawlowski, Maria Fernandez, Daniel Wibberg, 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. Systematic and Applied Microbiology, 2020, 43 (6), pp.126134. 10.1016/j.syapm.2020.126134 . hal-03150467

HAL Id: hal-03150467

<https://hal.science/hal-03150467>

Submitted on 17 Oct 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

REVISED VERSION

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27

***Candidatus Frankia nodulisporulans* sp. nov., an *Alnus glutinosa*-infective *Frankia* species unable to grow in pure culture and able to sporulate *in-planta* ~~growing in France and Sweden~~**

Aude Herrera-Belaroussi^{1*}, Philippe Normand¹, Katharina Pawlowski², Maria P. Fernandez¹, Daniel Wibberg³, Jörn Kalinowski³, Andreas Brachmann⁴, Fede Berckx², Natuschka Lee⁵, Jochen Blom⁶, Adrien C. Pozzi¹, Pascale Fournier¹, Lorine Bethencourt¹, Audrey Dubost¹, Danis Abrouk¹ and Anita Sellstedt^{7*}

¹Université de Lyon, Université Lyon 1, CNRS, UMR5557, Ecologie Microbienne, INRA, UMR 1418, 43 bd du 11 novembre 1918, F-69622 Villeurbanne, France

²Department of Ecology, Environment and Plant Sciences, Stockholm University, 10691 Stockholm, Sweden

³The Center for Biotechnology CeBiTec, Bielefeld University, Universitätsstraße 25, 33615 Bielefeld, Germany

⁴Biocenter of the LMU Munich, Genetics Section, Grosshaderner Str. 2-4, D-82152 Planegg-Martinsried, Germany

⁵Department of Ecology and Environmental Science, Umeå University, 90187 Umeå, Sweden

⁶Bioinformatics & Systems Biology, Justus-Liebig-University, 35392 Giessen, Hesse, Germany

⁷Department of Plant Physiology, UPSC, Umeå University, 90187 Umeå, Sweden

*Corresponding authors:

Aude Herrera-Belaroussi, Tel: +33 472 448 200 - E-mail: aude.herrera-belaroussi@univ-lyon1.fr

Anita Sellstedt, Tel: +46 90 786 77 56 - E-mail: anita.sellstedt@umu.se

28 **Key-words**

29 *In-planta* sporulation, Sp+ *Frankia* strains, *Frankia-Alnus* symbiosis, hydrogenase,
30 comparative genomics

31

32 **Abbreviations:** ANI, average nucleotide identity; ML, maximum-likelihood; ~~NJ, Neighbour-~~
33 ~~Joining~~

34

35 The GenBank/EMBL/DDBJ accession numbers for the 16S rRNA gene and draft genome
36 sequences reported in this study for AgTrS, AgUmASt1 and AgUmASH1 are
37 MT023539/LR778176/LR778180 and
38 NZ_CADCWS000000000.1/CADDZU010000001/CADDZW010000001, respectively.

39

40

41 **ABSTRACT**

42 We describe a new *Frankia* species, for three non-isolated strains obtained from *Alnus*
43 *glutinosa* in France and Sweden, respectively. These strains can nodulate several *Alnus*
44 species (*A. glutinosa*, *A. incana*, *A. alnobetula*), they form hyphae, vesicles and sporangia
45 in the root nodule cortex but have resisted all attempts at isolation in pure culture. Their
46 genomes have been sequenced, they are significantly smaller than those of other *Alnus*-
47 infective species (5 Mb instead of 7.5 Mb) and are very closely related to one another (ANI
48 of 100%). The name *Candidatus Frankia nodulisporulans* is proposed.

49

50

51 Introduction

52

53 The *Frankia* genus encompasses morphologically homogeneous soil actinobacteria
54 able to fix nitrogen and to establish an intracellular symbiosis with actinorhizal plants, leading
55 to the formation of perennial root nodules wherein bacteria are hosted and nitrogen is fixed.
56 Actinorhizal plants represent ca. 220 species belonging to eight angiosperm families that
57 have evolved to establish endophytic symbiotic interactions with *Frankia* [6, 36, 50], allowing
58 them to grow under adverse environmental conditions (low nitrogen, low fertility, high acid
59 and metals content, *etc.*) and to act as pioneers initiating plant succession in nitrogen-poor
60 biotopes (*e.g.* glacial moraines, burned forests or mine spoils) [6]. Able to grow
61 endophytically and saprotrophically, *Frankia* strains have been reported in diverse areas,
62 including extreme environments and in soils devoid of compatible host plants [21]. However,
63 their diversity and phylogeny have remained very difficult to estimate because numerous
64 strains could not be cultured to date.

65 *Frankia* strains with the ability to sporulate *in-planta*, called Sp+ (for spore-positive
66 phenotype) have so far never been cultured despite numerous attempts [41, 42]. They can
67 be distinguished from Sp- strains, most of which have been described as sporulating *in-vitro*,
68 by their ability to produce numerous and massive multilocular sporangia inside host plant
69 cells [41, 47, 49]. Sp+ strains have been reported from nodules of nine of the 23 actinorhizal
70 plant genera, with a particularly high frequency for the *Alnus* genus – also the most studied
71 genus –, especially for the species *A. glutinosa*, *A. incana* and *A. alnobetula* (formerly *A.*
72 *viridis*) [35, 41, 44]. Over the past twelve years, an increasing number of studies has focused
73 on describing Sp+ *Alnus*-infective strains in terms of diversity, ecology and metabolism [5,
74 12, 15, 35, 36, 42]. The most recent molecular studies revealed that Sp+ strains form distinct
75 phylogenetic lineages, strongly correlated with the host plant species, suggesting that Sp+
76 strains had emerged several times independently over the course of *Frankia* diversification.

77 Endophytes belonging to the different lineages had different ecology and traits. While Sp+
78 strains associated with the species *A. alnobetula* growing in alpine habitats showed a very
79 narrow host specificity, mostly limited to their original host species, strains associated to *A.*
80 *glutinosa* were more permissive [42].

81 *Frankia* strains are grouped into four major Clusters, three of which containing
82 symbiotic strains, linked to infection groups, one of which, Cluster 1, comprises most *Alnus*-
83 infective strains [25]. However, 16S rRNA-based phylogenies were shown to lack resolution
84 within the *Frankia* genus, and single gene phylogenies – known to be often not congruent
85 with whole genome phylogenies due to recombination events and Lateral Gene Transfers
86 (LGT; [35]) led to different results with regard to the question of which cluster was the earliest
87 one to diverge [21]. Recently, a Multi-Locus Sequence Analysis (MLSA) using *atpD*, *dnaA*,
88 *ftsZ*, *pgk*, and *rpoB* loci was performed on a large set of cultured and non-cultivated strains
89 [35], providing a robust resolution of the phylogeny of *Frankia* Clusters 1, 3 and 4, in
90 agreement with the 16S rRNA phylogeny and core genome phylogenies [11, 25, 34, 35].
91 Cluster 1 contains strains infective on *Alnus*, actinorhizal Myricaceae, and actinorhizal
92 members of the Casuarinaceae except for *Gymnostoma* sp. This cluster is subdivided into
93 two **main** subclusters [25]: 1a for *Alnus*- and Myricaceae-infective strain and 1c for
94 Casuarinaceae-infective strains. Cluster 2 encompasses strains infective on actinorhizal
95 Cucurbitales, Rosaceae and the Rhamnaceae genus *Ceanothus*; Cluster 3 strains nodulate
96 Elaeagnaceae, Rhamnaceae except for *Ceanothus*, and the Casuarinaceae outlier genus
97 *Gymnostoma*. Cluster 4 contains atypical non-nodulating or non-nitrogen-fixing strains. With
98 the increasing amount of sequenced bacterial genomes, **15** *Frankia* species have been
99 confirmed or named to date; *Frankia alni* [30], *Frankia torreyi* [28], *Frankia canadensis* [24],
100 *Frankia casuarinae* [30], *Frankia elaeagni* [30], *Frankia coriariae* [32], *Frankia discariae* [26],
101 *Frankia irregularis* [31], *Frankia inefficax* [27], *Frankia asymbiotica* [33], *Frankia*
102 *saprophytica* [29] and *Frankia soli* [10] as well as non-cultivated *Candidatus Frankia*

103 *daticae* [34], *Candidatus Frankia californiensis* [23] and *Candidatus Frankia meridionalis*
104 [20].

105 The MLSA studies highlighted several Sp+ lineages among *Alnus*-infective strains
106 with a genetic distance between each other and with the different Sp- lineages that
107 suggested these lineages constituted new species [35]. Recently, Sp+ genomes were
108 obtained directly from *Frankia* spores isolated from nodules of different *Alnus* species and
109 confirmed *Alnus*-infective Sp+ strains are likely to represent new species within the *Frankia*
110 genus [1].

111 We hereby aimed to describe a new species, *Candidatus Frankia nodulisporulans* sp.
112 nov., encompassing Sp+ *Alnus glutinosa* symbionts. The present study is based on the
113 genome sequences of three non-cultivated *Alnus glutinosa*-infective strains: AgTrS from
114 nodules of *A. glutinosa* growing in France, whose genome has recently been published [1]
115 and two newly sequenced strains, AgUmASt1 and AgUmASH1, sequenced from *A.*
116 *glutinosa* nodules induced by the so-called “local source” of *Frankia* in Sweden, one of the
117 earliest Sp+ inocula identified, characterized by the lack of uptake hydrogenase activity in
118 nodules [13, 44].

119

120

121 **Materials and Methods**

122

123 ***Frankia strain origin***

124 Three non-cultivated *Alnus*-infective Sp+ strains were selected: AgTrS, AgUmASt1
125 and AgUmASH1. These strains originated from two distinct alder stands. AgTrS came from
126 *Alnus glutinosa* nodules collected at Le Tremblay (**La motte Servolex**, Savoie, France),
127 previously described [36], and AgUmASt1 and AgUmASH1 were from nodules of *A.*
128 *glutinosa* induced by the “local source” of *Frankia* [13, 44] on plants grown in the greenhouse

129 (the “local source” designated an inoculum made from crushed nodules of *A. incana*
130 collected in Sweden and propagated for many years on *A. incana* as well as *A. glutinosa*
131 plants grown under greenhouse conditions [13, 44]). AgUmAS1 came from a nodule
132 harvested in 2019 and AgUmASH1 from a nodule harvested in 2014. All inocula were
133 maintained on *A. glutinosa* seedlings under controlled greenhouse conditions (with a 8-h
134 and 16-h nycthemeral regime with temperatures of 21 and 25°C, respectively), in the
135 UMR5557 Microbial Ecology of Lyon (France) for AgTrS and in greenhouses of the
136 Department of Plant Physiology at UPSC (Umeå University, Sweden) for AgUmAS1 and
137 AgUmASH1.

138

139 ***In-planta morphology***

140 *In-planta* sporulation phenotype was determined (from at least one lobe) by
141 microscopic observations of hand or semi-thin nodule sections stained with Lactophenol
142 Blue (Réactifs RAL, Martillac, France), as previously described [35]. The Sp+ phenotype
143 was established when more than one sporangium was observed within 50 infected plant
144 cortical cells.

145

146 ***DNA extraction from nodules and genome sequencing***

147 For the genome of AgTrS, DNA was extracted from a spores suspension isolated from
148 crushed nodules and sequenced as previously described by Bethencourt et al. [1].

149 For both AgUmAS1 and AgUmASH1, DNA was extracted from whole nodules of *Alnus*
150 *glutinosa* as described by Nguyen et al. [20]. Genomic DNA was sheared on a Covaris M220
151 grinder with Covaris MicroCaps 50 µl (75 W Peak Incident Power, Duty Factor 10%, 200
152 Cycles/Burst, Treatment Time 42 sec, 20°C) to approx. 570 bp. Library preparation was
153 performed with 140 ng sheared gDNA using the sparQ DNA Library Prep Kit (QuantaBio)
154 according to the manufacturer's instructions. Libraries were quality controlled with a DNA

155 High Sensitivity DNA Kit on Bioanalyzer (Agilent) and quantified on Qubit 2.0 Fluorometer
156 (ThermoFisher Scientific with ds HS Assay Kit). Genome sequencing was performed in the
157 Genomics Service Unit (LMU Biocenter, Munich) on Illumina MiSeq with v3 chemistry (2x
158 250 bp paired-end sequencing). Genome assembly and annotation were performed as
159 recently described with a few modifications [20]. In brief, raw reads were assembled using
160 the gsAssembler 2.8 (Roche) with default settings.

161 Completeness, contamination, and strain heterogeneity for the three genomes were
162 estimated with BUSCO (v3.0.2; [46]), using the bacteria-specific single-copy marker genes
163 database (odb9). For genome annotation, Prokka [43] and GenDB [17] were used. Draft
164 genome sequences were deposited at the EMBL/GenBank/DDBJ databases with accession
165 numbers NZ_CADCWS000000000.1, CADDZU010000001-CADDZU010000211 and
166 CADDZW010000001-CADDZW010000231 for AgTrS, UmAst1 and UmASH1, respectively.

167

168 ***Average nucleotide identity (ANI) and different molecular phylogenies***

169 Average Nucleotide Identity (ANI) comparisons were performed in order to accurately
170 determine the species borders within the *Frankia* genus, using the usually applied ANI
171 threshold range of 95% for species delineation [9]. The analysis was performed for 16
172 genomes of *Frankia* Cluster 1 strains including the three Sp+ strains described in this study,
173 using the comparative genomics program EDGAR 2.0 [2, 3].

174 Further analysis of the 16 *Frankia* genomes used for ANI analysis, and a set of strains
175 representing all Cluster 2, Cluster 3 and Cluster 4 *Frankia* species proposed so far, was
176 performed using the comparative genomics program EDGAR 2.0. These analyses
177 comprised the identification of orthologous genes, classification of genes as core genes or
178 singletons as well as the creation of a phylogenetic tree **based on the core genome using**
179 **the FastTree software (<http://www.microbesonline.org/fasttree/>) to generate approximately-**
180 **maximum-likelihood phylogenetic trees [37, 38].**

181 For 16S rRNA based phylogeny, sequences were aligned by using CLUSTAL X [14]
182 as implemented on SEAVIEW [8]. Phylogenetic trees were constructed by the Maximum
183 Likelihood method running on SEAVIEW [8]. A bootstrap confidence analysis was performed
184 with 1,000 replicates to determine the reliability of the distance tree topologies obtained [7],
185 only those results above 95% are indicated.

186 MLSA phylogeny was performed based on a Maximum Likelihood tree inferred from
187 the concatenation of five markers (*atpD*, *dnaA*, *ftsZ*, *pgk*, *rpoB*) as previously described [36],
188 but only *Alnus*-infective strains were included in order to obtain optimal resolution for Cluster
189 1a.

190

191

192 **Results**

193

194 All three strains AgTrS, AgUmASt1 and AgUmASH1 expressed the Sp⁺ phenotype in
195 nodules collected from the field and when they were maintained on *A. glutinosa* seedlings
196 under controlled greenhouse conditions. In addition to hyphae and nitrogen-fixing vesicles,
197 numerous spores contained in sporangia were indeed observed within *Alnus* nodule cells,
198 sometimes filling the whole volume of the nodule cells (Figure 1).

199

200 ***AgTrS*, *AgUmASt1* and *AgUmASH1* genome description**

201 The final draft assembly for AgTrS consisted of 281 contigs (≥ 5 kb), with a total
202 genome size of 4,882,652 bp and an overall G+C content of 71.6% (Table 1). For AgUmASt1
203 and AgUmASH1, the final draft assembly consisted of 211 (N50: 25 kb) and 231 (N50: 24
204 kb) scaffolds (≥ 3 kb) containing 4,315,411 bp and 4,261,328 bp, respectively. These two
205 draft genomes had an overall G+C content of 71.3 and 71.2 %, respectively. Genome
206 completeness was estimated at 94.1% for AgTrS, 93.9 % for AgUmASt1 and 90.5 %

207 AgUmASH1, using BUSCO software that assesses completeness of genome assemblies
208 based on the presence of a specific number of near-universal orthologs from related
209 organisms in OrthoDB (148 single core genes for *Frankia* genomes) [45]. The assembled
210 genomes of AgTrS, AgUmASt1 and AgUmASH1 strains resulted in 4,388, 3,709 and 3,681
211 predicted protein-encoding genes, respectively. Interestingly, genome exploration using
212 AntiSMASH [16] revealed the presence of 16, 12 and 10 gene clusters coding for secondary
213 metabolites in AgTrS, AgUmASt1 and AgUmASH1 genomes, respectively (Table 1).

214

215 ***Taxonomy and phylogenetic affiliation***

216 AgUmASt1, AgUmASH1 and AgTrS showed median Average Nucleotide Identity (ANI)
217 values of 100% with one another and values equal to or below 90.7 % with other *Alnus*-
218 infective *Frankia* species, the closest one being *F. torreyi* Cpl1^T. ANI values with *F.*
219 *casuarinae* Ccl3^T, the closest non *Alnus*-infective *Frankia* species, were about 77.3% (Table
220 2).

221 The 16S rRNA tree (Figure 2A) permitted to recover the known clusters, except
222 Clusters 3 and 4 that were part of Cluster 1, though with a poor bootstrap value. The only
223 clusters with strong bootstrap values were Cluster 2 and Cluster 4. Within Cluster 1, the
224 three Sp+ strains AgUmASt1, AgUmASH1 and AgTrS were closest to *F. torreyi* with strong
225 bootstrap support.

226 The core genome tree (Figure 2B) shows a different structure. The three Sp+ strains
227 AgUmASt1, AgUmASH1 and AgTrS form their own cluster within Cluster 1. All clusters are
228 recovered with Cluster 2 in the earliest divergent position.

229 The MLSA tree shows that the three Sp+ strains AgUmASt1, AgUmASH1 and AgTrS
230 cluster close together, forming a coherent group, closest to AVN17-AVL3-Ac23.40, gathered
231 with the other *Alnus*-infective strains (Figure 3). The three Sp+ strains together with AVN17-
232 AVL3-Ac23.40 represent the earliest divergent group of alder-infective strains.

233

234 **Genes encoding uptake hydrogenase**

235 In Cluster 1 strains, [NiFe] uptake hydrogenase is encoded by two syntons [22]. *HupS*
236 and *HupL* encode the small and large subunit of [NiFe] uptake hydrogenase, respectively.
237 The three genomes of Sp+ strains described in this study lack synton #1, while containing
238 an incomplete version of synton #2.

239

240

241 **Discussion**

242 In the last few years, *Frankia* systematics has made significant progress thanks to a
243 community effort to sequence the genomes of cultivated strains. Now, the technology is
244 available to sequence genomes from non-cultivated species [1, 19, 20, 34], which allows
245 the exploration of previously inaccessible lineages. Based on these genomes, one objective
246 of the present study was to name another non-cultivated *Frankia* species. Inability to grow
247 in pure culture is seen in several microbial lineages such as *Mycobacterium leprae* where it
248 has been explained by massive genome decay [4]. In the case of its close relative
249 *Mycobacterium haemophilum*, this defect was explained by its loss of siderophore genes
250 [48]. Genome-based studies of *Tropheryma whipplei* (that causes Whipple's disease) have
251 permitted to circumvent such blockage and to grow the microbe in pure culture [39]. In the
252 case of *Frankia* Cluster 2, an empirical approach with a Biolog analysis of nodular tissues
253 permitted to choose an alkaline medium and obtain a pure culture of *Frankia coriariae* [11].
254 The genomes of Sp+ *Frankia* were obtained [1] in part to address the question of culturability
255 and to understand what had happened in evolutionary terms.

256 Intriguingly, Sp+ strains of Cluster 1a show the strongest genome reduction observed
257 so far in the entire genus, with Sp+ genomes reduced by more than one third from 7.5 Mb
258 to less than 5 Mb. The *Frankia* genus is characterized by an unusually large divergence of

259 genome sizes, a fact that has been linked to their capability to survive in diverse (large
260 genomes) or narrow (small genomes) ecological niches in the soil [22]. Lineages that so far
261 resisted efforts of cultivation are assumed to have lower saprotrophic capabilities than
262 culturable lineages. It is likely that nodules became the only habitat for *Frankia* in
263 inhospitable biotopes such as cold regolith-like mountain locales [36] or acidic bogs with low
264 microbial population levels [50]. *Frankia* genome reduction so far had been seen in *Frankia*
265 Cluster 2 strains [11, 19, 20, 34] and in Cluster 1c [22]. The substantial genomic purge of
266 Sp+ strains of Cluster 1, combined with the fact that they could never be cultured despite
267 numerous attempts, suggest they are dependent on the host plant for a large part of their
268 life cycle and support the obligate symbiont scenario previously discussed [13, 35]. Under
269 the hypothesis of a stronger dependence on their host, Sp+ strains could have a narrower
270 host-range compared with Sp- strains [35, 41, 47]. For the three strains AgTrS, AgUmAST1
271 and AgUmASH1, previous experiments based on cross-inoculation experiments (*e.g.* using
272 crushed nodules to inoculate seedlings grown under greenhouse conditions) have been
273 reported. The results showed that AgTrS was compatible with the three *Alnus* species *A.*
274 *glutinosa*, *A. incana* and *A. alnobetula* [5] and AgUmAST1 / AgUmASH1 inocula formed N₂-
275 fixing nodules on nine species of *Alnus* and also on *Myrica gale* [13]. However, a recent
276 study based on *Frankia* strain trapping experiments from soil collected under Sp+ and Sp-
277 alder stands (including soil collected from the AgTrS alder stand) revealed clear
278 incompatibilities of *Alnus*-infective Sp+ strains for different *Alnus* species [42]. This study
279 supported a narrower host-range of Sp+ strains compared to Sp- strains.

280 In evolutionary terms, the existence of several distinct lineages of Sp+ *Frankia* strains
281 [1, 35, 36] indicates that the emergence of *in-planta* sporulating strains occurred
282 independently. One hypothesis to explain such evolutionary events in several lineages is
283 that when a bacterium ceases to grow outside its host, it can no longer exchange genes to
284 compensate deletion events and then starts to accumulate deleterious mutations in non-

285 essential genes, a process called Müller's ratchet [18]. Coupled with reductive genome
286 evolution, such lineages often have an accelerated mutation rate and a lower G+C% than
287 their saprotrophic neighbours. In addition, the genomes of the three AgTrS, AgUmAS1 and
288 AgUmASH1 strains were found to be closely related (median ANI values of 100% with one
289 another) despite their distant geographical origin (Southern France *versus* Northern
290 Sweden), suggesting the common ancestor appeared relatively recently and was then
291 distributed across Europe, maybe following the latest glaciation and the subsequent
292 recolonization from refugia [45].

293 During genome reduction in Sp+ strains, several genes could have been lost, with
294 among them the symbiotic up-regulated *hup* genes (~~but not during genome reduction in~~
295 ~~Cluster 1e~~). The presence of an uptake hydrogenase has been shown to be beneficial for
296 both rhizobial and *Frankia* symbioses, allowing high biomass production and nitrogenase
297 activities [40, 44]. In the present study, the three genomes of Sp+ strains were found to lack
298 synton #1, while containing an incomplete version of synton #2. This observation is coherent
299 with the lack of uptake hydrogenase activity that was consistently observed for nodules
300 induced by the "local source" - corresponding to AgUmAS1 and AgUmASH1 strains [13,
301 44]. Among other genes lost in AgTrS, AgUmASH1 and AgUmAS1, we also found the
302 cellulase operon (FRAAL4955-4960), some lectin genes (FRAAL2972), geosmin synthesis
303 (FRAAL6507), gas vesicle proteins possibly used for flotation on the soil watertable
304 (FRAAL3021-3029), an aspect that is being investigated in detail.

305 Finally, several other *Frankia* strains seem to belong to the same lineage as AgTrS,
306 AgUmAS1 and AgUmASH1 [35], and distinct potential Sp+ species could exist in Cluster
307 1a of the *Frankia* genus [1, 35]. Our effort to sequence additional Sp+ *Frankia* genomes has
308 to be continued and thorough analyses based on comparative genomic approaches with
309 *Frankia* Sp- genomes already available have to be performed to deeply explore the evolution
310 of their life history traits. Such analyses would for example give us clues to their ability to

311 sporulate *in-planta*, to their reduced saprotrophic capability/host dependence, and clarify
312 their hypothetical status of obligatory symbionts.

313

314 In summary, both genomic and phylogenetic data (ANI, 16S, MLSA) grouped together
315 AgTrS, AgUmAS1 and AgUmASH1 strains in the same species, proposed as *Candidatus*
316 *Frankia nodulisporulans*. The species does not grow in the tested growth media, induces
317 the formation of nodules when inoculated on *Alnus glutinosa*, *A. incana* or *A. alnobetula*
318 under greenhouse conditions and sporulates in infected cells in cortical tissues. This species
319 can be found in root nodules of alder in lowland regions of France or in shores of rivers and
320 sea of northern Sweden.

321 The description of *Candidatus* *Frankia nodulisporulans* is given in Table 3.

322

323

324 **Acknowledgements**

325 The bioinformatics support of the BMBF-funded project “Bielefeld-Gießen Center for
326 MicrobialBioinformatics” (BiGi) and the BMBF grant FKZ 031A533 within the German
327 Network for Bioinformatics Infrastructure (de.NBI) are gratefully acknowledged. A.C. Pozzi
328 was granted a doctoral fellowship by the Ministry of Higher Education and Research (MESR,
329 France). We also acknowledge support from TC4F (Trees and Crops for the Future) and
330 initial assistance from Han Nguyen.

331

332 **Tables**

333

334 Table 1. Genomic characteristics of AgTrS, AgUmAST1 and AgUmASH1 *Frankia* strains.

335

336 Table 2. Median ANI values between genomes of strains from *Frankia* Cluster 1 (including
337 all available sequenced genomes in Cluster 1). All currently known *Frankia* Cluster 1 species
338 are represented.

339

340 Table 3. Description of *Candidatus Frankia nodulisporulans* sp. nov.

341

342

343 **Figure legends**

344

345 Figure 1. Semi-thin section of *Alnus glutinosa* nodule stained with lactophenol blue. Some
346 colonized cells mostly contain hyphae (H) and diazo-vesicles (V) while others are full of
347 spores gathered in sporangia (S). The scale bar represents 10 µm. Similar observations
348 were done from nodules containing *Frankia* strain AgTrS, AgUmAH1 or AgUmAst1.

349

350 Figure 2. (A) Phylogenetic tree based on 16S rDNA sequences inferred using CLUSTAL X
351 [14] as implemented on SEAVIEW [8]. Phylogenetic trees were constructed by the Maximum
352 Likelihood [8] method. A bootstrap confidence analysis was performed with 1,000 replicates
353 to determine the reliability of the distance tree topology obtained [7], only those results above
354 95% are indicated. (B) Phylogenetic tree for 20 genomes, built out of a core of 531 genes
355 per genome, 10,620 in total, by EDGAR 2.0 [2, 3] using the FastTree software
356 (<http://www.microbesonline.org/fasttree/>) to generate approximately-maximum-likelihood
357 phylogenetic trees [37, 38]. The values at the branches show local support values computed

358 by FastTree using the Shimodaira-Hasegawa test. The core has 199469 amino acid
359 residues per genome, 3,989,380 in total. Genomes used were *Candidatus Frankia datisca*
360 Dg1^T (NC_015656.1), *Frankia coriariae* BMG5.1^T (NZ_JWIO00000000.1), *Candidatus*
361 *Frankia californiensis* Dg2^T (FLUV00000000.1), *Candidatus Frankia meridionalis* Cppng1^T
362 (CADDZT010000001-CADDZT010000101), *Frankia asymbiotica* M16386^T
363 (NZ_MOMC00000000.1), *Frankia saprophytica* CN3^T (NZ_AGJN00000000.2), *Frankia*
364 *inefficax* Eu1c^T (NC_014666.1), *Frankia irregularis* DSM45899^T (NZ_FAOZ00000000.1),
365 *Frankia elaeagni* BMG5.12^T (NZ_ARFH00000000.1), *Frankia discariae* BCU110501^T
366 (NZ_ARDT00000000.1), *Frankia soli* NRRL B-16219^T (MAXA00000000.1), *Frankia*
367 *casuarinae* Ccl3^T (NC_007777.1), *Frankia canadensis* ARgP5^T (GCF_900241035.1),
368 *Candidatus Frankia nodulisporulans* AgTrS^T (NZ_CADCWS00000000.1), *Candidatus*
369 *Frankia nodulisporulans* AgUmAst1 (CADDZU010000001-CADDZU010000211),
370 *Candidatus Frankia nodulisporulans* AgUmASH1 (CADDZW010000001-
371 CADDZW010000231), *Frankia alni* ACN14a^T (NC_008278.1), *Frankia torreyi* Cpl1^T
372 (NZ_JYFN00000000.1) and *Frankia sp.* QA3^T (AJWA00000000). The genome of
373 *Cryptosporangium arzum* DSM44712^T (JFBT01000000) was used as the outgroup. The size
374 bar denotes 0.5 changes. ~~Bootstrap values are 100 for every branch; they were calculated~~
375 ~~in R (<http://www.R-project.org/>) [48] using the packages APE [35] and phangorn [41]~~
376 For both A and B panels, numbers in shaded boxes on the right correspond to the previously
377 described clusters in *Frankia* genus [25]. The arrows point at the three strains of *Candidatus*
378 *Frankia nodulisporulans*.

379

380 Figure 3. *Alnus*-infective *Frankia* multilocus sequence analysis (MLSA) phylogeny. A
381 Maximum Likelihood tree was inferred from the concatenation of 5 markers (*atpD*, *dnaA*,
382 *ftsZ*, *pgk*, *rpoB*) into a super-alignment as previously described [35]. Only *Alnus*-infective
383 strains were included in order to obtain optimal resolution for Cluster 1a. The bar denotes a

384 0.01 substitution/site distance. Numbers reported at nodes are the aLRT SH-like branch
385 support. The arrow points at the three strains of *Candidatus Frankia nodulisporulans*.

386

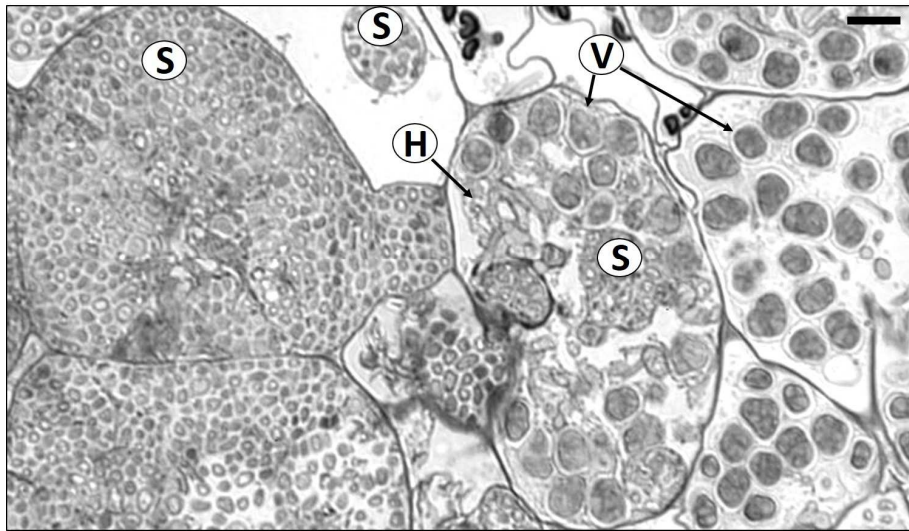
387 References

- 388 [1] Bethencourt, L, Vautrin, F, Taib, N, Dubost, A, Castro-Garcia, L, Imbaud, O et al. (2019) Draft
389 genome sequences for three unisolated *Alnus*-infective *Frankia* Sp+ strains, AgTrS, AiOr and
390 AvVan, the first sequenced *Frankia* strains able to sporulate in-plant. J. Genomics 7, 50-55.
- 391 [2] Blom, J, Albaum, SP, Doppmeier, D, Puhler, A, Vorholter, FJ, Zakrzewski, M et al. (2009)
392 EDGAR: a software framework for the comparative analysis of prokaryotic genomes. BMC
393 Bioinformatics 10, 154.
- 394 [3] Blom, J, Kreis, J, Spänig, S, Juhre, T, Bertelli, C, Ernst, C et al. (2016) EDGAR 2.0: an enhanced
395 software platform for comparative gene content analyses. Nucleic Acids Res. 44(W1), W22–
396 W28.
- 397 [4] Cole, ST, Eiglmeier, K, Parkhill, J, James, KD, Thomson, NR, Wheeler, PR et al. (2001) Massive
398 gene decay in the leprosy bacillus. Nature 409, 1007-1011.
- 399 [5] Cotin-Galvan, L, Pozzi, AC, Schwob, G, Fournier, P, Fernandez, MP, Herrera-Belaroussi, A
400 (2016) In-plant sporulation capacity enhances infectivity and rhizospheric competitiveness
401 of *Frankia* strains. Microbes Environ. 31, 11–18.
- 402 [6] Dawson, JO (2008) Ecology of actinorrhizal plants. In Newton KPWE (ed.), Nitrogen-Fixing
403 Actinorrhizal Symbioses-Nitrogen Fixation Research: Origins and Progress. Springer, New
404 York. pp. 199-234.
- 405 [7] Felsenstein, J (1985) Confidence limits on phylogenies: an approach using the bootstrap.
406 Evolution 39, 783-791.
- 407 [8] Galtier, N, Gouy, M, Gautier, C (1996) SEAVIEW and PHYLO_WIN: two graphic tools for
408 sequence alignment and molecular phylogeny. Comput Appl Biosci 12, 543-548.
- 409 [9] Goris, J, Konstantinidis, KT, Klappenbach, JA, Coenye, T, Vandamme, P, Tiedje, JM (2007)
410 DNA-DNA hybridization values and their relationship to whole-genome sequence
411 similarities. Int J Syst Evol Microbiol 57, 81-91.
- 412 [10] Gtari, M, Ghodhbane-Gtari, F, Nouioui, I (2019) *Frankia soli* sp. nov., an actinobacterium
413 isolated from soil beneath *Ceanothus jepsonii*. Int J Syst Evol Microbiol.
- 414 [11] Gtari, M, Ghodhbane-Gtari, F, Nouioui, I, Ktari, A, Hezbri, K, Mimouni, W et al. (2015)
415 Cultivating the uncultured: growing the recalcitrant cluster-2 *Frankia* strains. Sci. Rep. 5,
416 13112.
- 417 [12] Hay De-Bettignies, AE, Boubakri, H, Buonomo, A, Rey, M, Meiffren, G, Cotin-Galvan, L et al.
418 (2017) Control of endophytic *Frankia* sporulation by *Alnus* nodule metabolites. Mol Plant
419 Microbe Interact. 30, 205-214.
- 420 [13] Huss-Danell, K (1991) Influence of host (*Alnus* and *Myrica*) genotype on infectivity, N₂ fixation,
421 spore formation and hydrogenase activity in *Frankia*. New Phytol 119, 121-127.
- 422 [14] Larkin, MA, Blackshields, G, Brown, NP, Chenna, R, McGettigan, PA, McWilliam, H et al.
423 (2007) Clustal W and Clustal X version 2.0. Bioinformatics 23, 2947-2948.
- 424 [15] Markham, JH (2008) Variability of nitrogen fixing *Frankia* on *Alnus* species. Botany 86, 501–
425 510.
- 426 [16] Medema, MH, Blin, K, Cimermancic, P, de Jager, V, Zakrzewski, P, Fischbach, MA et al. (2011)
427 antiSMASH: rapid identification, annotation and analysis of secondary metabolite
428 biosynthesis gene clusters in bacterial and fungal genome sequences. Nucleic Acids Res 39,
429 W339–W346.

- 430 [17] Meyer, F, Goesmann, A, McHardy, AC, Bartels, D, Bekel, T, Clausen, J et al. (2003) GenDB –
431 an open source genome annotation system for prokaryote genomes. Nucl. Acids Res. 31,
432 2187–2195.
- 433 [18] Muller, HJ (1964) The relation of recombination to mutational advance. Mutat Res. 106, 2–9.
- 434 [19] Nguyen, TV, Wibberg, D, Battenberg, K, Blom, J, Vanden Heuvel, B, Berry, AM et al. (2016)
435 An assemblage of *Frankia* Cluster II strains from California contains the canonical *nod* genes
436 and also the sulfotransferase gene *nodH*. BMC Genomics 17, 796.
- 437 [20] Nguyen, TV, Wibberg, D, Vigil-Stenman, T, Berckx, F, Battenberg, K, Demchenko, KN et al.
438 (2019) *Frankia*-enriched metagenomes from the earliest diverging symbiotic *Frankia* cluster:
439 They come in teams. Genome Biol Evol. 11, 2273-2291.
- 440 [21] Normand, P, Benson, DR (2012) Genus I *Frankia* Brunchorst 1886, 174AL. In Goodfellow M,
441 Kämpfer P, Busse H-J, Trujillo ME, Ludwig W, Suzuki KI, Parte A (ed.), Bergey's Manual
442 of Systematic Bacteriology, The Actinobacteria., vol. 5. Bergey's Manual Trust, Springer,
443 New York. pp. 512-520.
- 444 [22] Normand, P, Lapierre, P, Tisa, LS, Gogarten, JP, Alloisio, N, Bagnarol, E et al. (2007) Genome
445 characteristics of facultatively symbiotic *Frankia* sp. strains reflect host range and host plant
446 biogeography. Genome Res. 17, 7-15.
- 447 [23] Normand, P, Nguyen, TV, Battenberg, K, Berry, AM, Vanden Heuvel, B, Fernandez, MP et al.
448 (2017) Proposal of *Candidatus Frankia californiensis*, the uncultured symbiont in nitrogen-
449 fixing root nodules of a phylogenetically broad group of hosts endemic to western North
450 America. Int. J. Syst. Evol. Microbiol. 67, 3706-3715.
- 451 [24] Normand, P, Nouioui, I, Pujic, P, Fournier, P, Dubost, A, Schwob, G et al. (2018) *Frankia*
452 *canadensis* sp. nov., isolated from root nodules of *Alnus incana* subspecies *rugosa*. Int. J.
453 Syst. Evol. Microbiol. 10.1099/ijsem.0.002939.
- 454 [25] Normand, P, Orso, S, Cournoyer, B, Jeannin, P, Chapelon, C, Dawson, J et al. (1996) Molecular
455 phylogeny of the genus *Frankia* and related genera and emendation of the family
456 *Frankiaceae*. Int J Syst Bacteriol 46, 1-9.
- 457 [26] Nouioui, I, Del Carmen Montero-Calasanz, M, Ghodhbane-Gtari, F, Rohde, M, Tisa, LS, Klenk,
458 HP et al. (2017) *Frankia discariae* sp. nov.: an infective and effective microsymbiont isolated
459 from the root nodule of *Discaria trinervis*. Arch. Microbiol. 199, 641-647.
- 460 [27] Nouioui, I, Ghodhbane-Gtari, F, Del Carmen Montero-Calasanz, M, Rohde, M, Tisa, LS, Gtari,
461 M et al. (2017) *Frankia inefficax* sp. nov., an actinobacterial endophyte inducing ineffective,
462 non nitrogen-fixing, root nodules on its actinorhizal host plants. Antonie Van Leeuwenhoek
463 110, 313-320.
- 464 [28] Nouioui, I, Ghodhbane-Gtari, F, Jando, M, Tisa, LS, Klenk, HP, Gtari, M (2018) *Frankia torreyi*
465 sp. nov., the first actinobacterium of the genus *Frankia* Brunchorst 1886, 174AL isolated in
466 axenic culture. Antonie van Leeuwenhoek //doi.org/10.1007/s10482-018-1131-8.
- 467 [29] Nouioui, I, Ghodhbane-Gtari, F, Klenk, H-P, Gtari, M (2018) *Frankia saprophytica* sp. nov., an
468 atypical, non-infective (Nod–) and non-nitrogen fixing (Fix–) actinobacterium isolated from
469 *Coriaria nepalensis* root nodules. Int. J. Syst. Evol. Microbiol. 68, 1090-1095.
- 470 [30] Nouioui, I, Ghodhbane-Gtari, F, Montero-Calasanz, MD, Goker, M, Meier-Kolthoff, JP,
471 Schumann, P et al. (2016) Proposal of a type strain for *Frankia alni* (Woronin 1866) Von
472 Tubeuf 1895, emended description of *Frankia alni*, and recognition of *Frankia casuarinae*
473 sp. nov. and *Frankia elaeagni* sp. nov. Int. J. Syst. Evol. Microbiol. 66, 5201-5210.
- 474 [31] Nouioui, I, Ghodhbane-Gtari, F, Rhode, M, Sangal, V, Klenk, HP, Gtari, M (2018) *Frankia*
475 *irregularis* sp. nov., an actinobacterium unable to nodulate its original host, *Casuarina*
476 *equisetifolia*, but effectively nodulates members of the actinorhizal *Rhamnales*. Int J Syst Evol
477 Microbiol 10.1099/ijsem.0.002914.
- 478 [32] Nouioui, I, Ghodhbane-Gtari, F, Rohde, M, Klenk, HP, Gtari, M (2017) *Frankia coriariae* sp.
479 nov., an infective and effective microsymbiont isolated from *Coriaria japonica*. Int. J. Syst.
480 Evol. Microbiol. 67, 1266-1270.

- 481 [33] Nouioui, I, Gueddou, A, Ghodhbane-Gtari, F, Rhode, M, Gtari, M, Klenk, HP (2017) *Frankia*
482 *asymbiotica* sp. nov., a non-infective actinobacterium isolated from *Morella californica* root
483 nodule. Int. J. Syst. Evol. Microbiol. doi: 10.1099.
- 484 [34] Persson, T, Battenberg, K, Demina, IV, Vigil-Stenman, T, Vanden Heuvel, B, Pujic, P et al.
485 (2015) *Candidatus Frankia datiscaae* Dg1, the actinobacterial microsymbiont of *Datisca*
486 *glomerata*, expresses the canonical *nod* genes *nodABC* in symbiosis with its host plant. PLoS
487 One 10, e0127630.
- 488 [35] Pozzi, AC, Bautista-Guerrero, HH, Abby, SS, Herrera-Belaroussi, A, Abrouk, D, Normand, P et
489 al. (2018) Robust *Frankia* phylogeny, species delineation and intraspecies diversity based on
490 Multi-Locus Sequence Analysis (MLSA) and Single-Locus Strain Typing (SLST) adapted to
491 a large sample size. Syst Appl Microbiol. 41, 311-323.
- 492 [36] Pozzi, AC, Bautista-Guerrero, HH, Nouioui, I, Cotin-Galvan, L, Pepin, R, Fournier, P et al.
493 (2015) In-plant sporulation phenotype: a major life history trait to understand the evolution
494 of *Alnus*-infective *Frankia* strains. Environ Microbiol. 17, 3125-3138.
- 495 [37] Price, MN, Dehal, PS, Arkin, AP (2010) FastTree 2--approximately maximum-likelihood trees
496 for large alignments. PLoS ONE 5, e9490.
- 497 [38] Price, MN, Dehal, PS, Arkin, AP (2009) FastTree: computing large minimum-evolution trees
498 with profiles instead of a distance matrix. Molec Biol Evol 26, 1641-1650.
- 499 [39] Renesto, P, Crapoulet, N, Ogata, H, La Scola, B, Vestris, G, Claverie, J et al. (2003) Genome-
500 based design of a cell-free culture medium for *Tropheryma whippelii*. Lancet 362, 447-449.
- 501 [40] Schubert, K, Evans, H (1976) Hydrogen evolution : a major factor affecting the efficiency of
502 nitrogen fixation in nodulated symbionts. Proc nat Acad Sci 73(4), 1207-1211.
- 503 [41] Schwintzer, CR (1990) Spore-positive and spore-negative nodules. In C.R. S, J.D. T (ed.), The
504 Biology of *Frankia* and Actinorhizal Plants. Academic Press, San Diego. pp. 177-193.
- 505 [42] Schwob, G, Roy, M, Pozzi, AC, Herrera-Belaroussi, A, Fernandez, MP (2018) In planta
506 sporulation of *Frankia* spp. as a determinant of alder-symbiont interactions. Appl Environ
507 Microbiol 84, e01737-01718.
- 508 [43] Seemann, T (2014) Prokka: rapid prokaryotic genome annotation. Bioinformatics 30, 2068-
509 2069.
- 510 [44] Sellstedt, A, Ahlqvist, A, Huss-Danell, K (1986) Nitrogen fixation and biomass production in
511 symbioses between *Alnus incana* and *Frankia* strains with different nitrogen metabolism.
512 Physiol Plant 66, 99-107.
- 513 [45] Seppey, M, Manni, M, Zdobnov, EM (2019) BUSCO: Assessing Genome Assembly and
514 Annotation Completeness. In M. K (ed.), Gene Prediction. Methods in Molecular Biology,
515 vol. 1962. , Humana, New York, NY. pp.
- 516 [46] Simão, FA, Waterhouse, RM, Ioannidis, P, Kriventseva, EV, Zdobnov, EM (2015) BUSCO:
517 assessing genome assembly and annotation completeness with single-copy orthologs.
518 Bioinformatics 31, 3210–3212.
- 519 [47] Torrey, J (1987) Endophyte sporulation in root nodules of actinorhizal plants. Physiol Plant 70,
520 279-288.
- 521 [48] Tufariello, JM, Kerantzas, CA, Vilchèze, C, Calder, RB, Nordberg, EK, Fischer, JA et al. (2015)
522 The complete genome sequence of the emerging pathogen *Mycobacterium haemophilum*
523 explains its unique culture requirements. mBio 6, e01313-01315.
- 524 [49] van Dijk, C (1978) Spore formation and endophyte diversity in root nodules of *Alnus glutinosa*
525 (L.) Vill. New Phytol 81, 601-615.
- 526 [50] Wolters, DJ, Akkermans, ADL, Van Dijk, C (1997) Ineffective *Frankia* strains in wet stands of
527 *Alnus glutinosa* L. Gaertn. in The Netherlands. Soil Biol. Biochem. 29, 1707-1712.
- 528

2 **Figures**

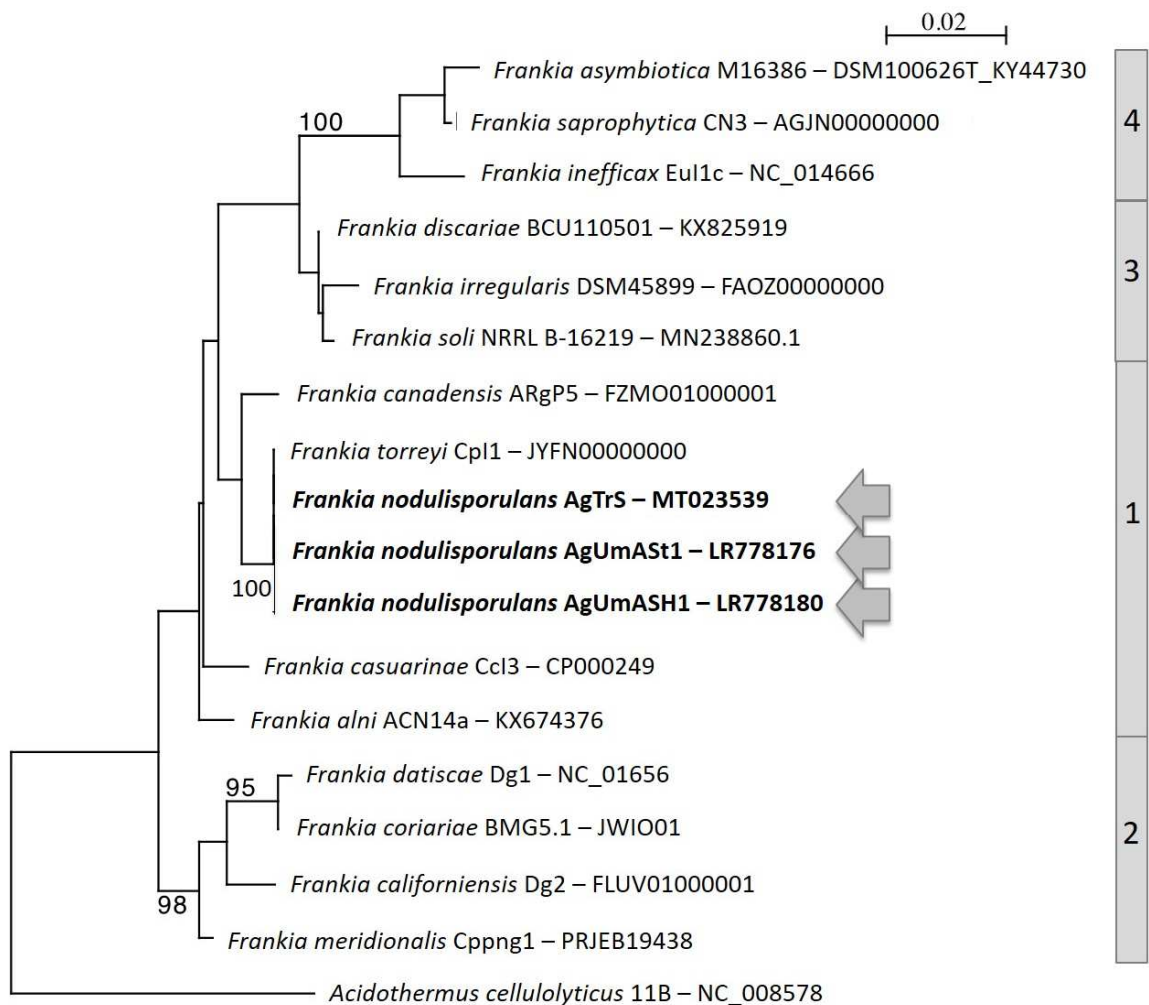


13 Figure 1. Semi-thin section of *Alnus glutinosa* nodule stained with lactophenol blue. Some
14 colonized cells mostly contain hyphae (H) and diazo-vesicles (V) while others are full of spores
15 gathered in sporangia (S). The scale bar represents 10 μ m. Similar observations were done from
16 nodules containing *Frankia* strain AgTrS^T, AgUmAH1 or AgUmAst1.

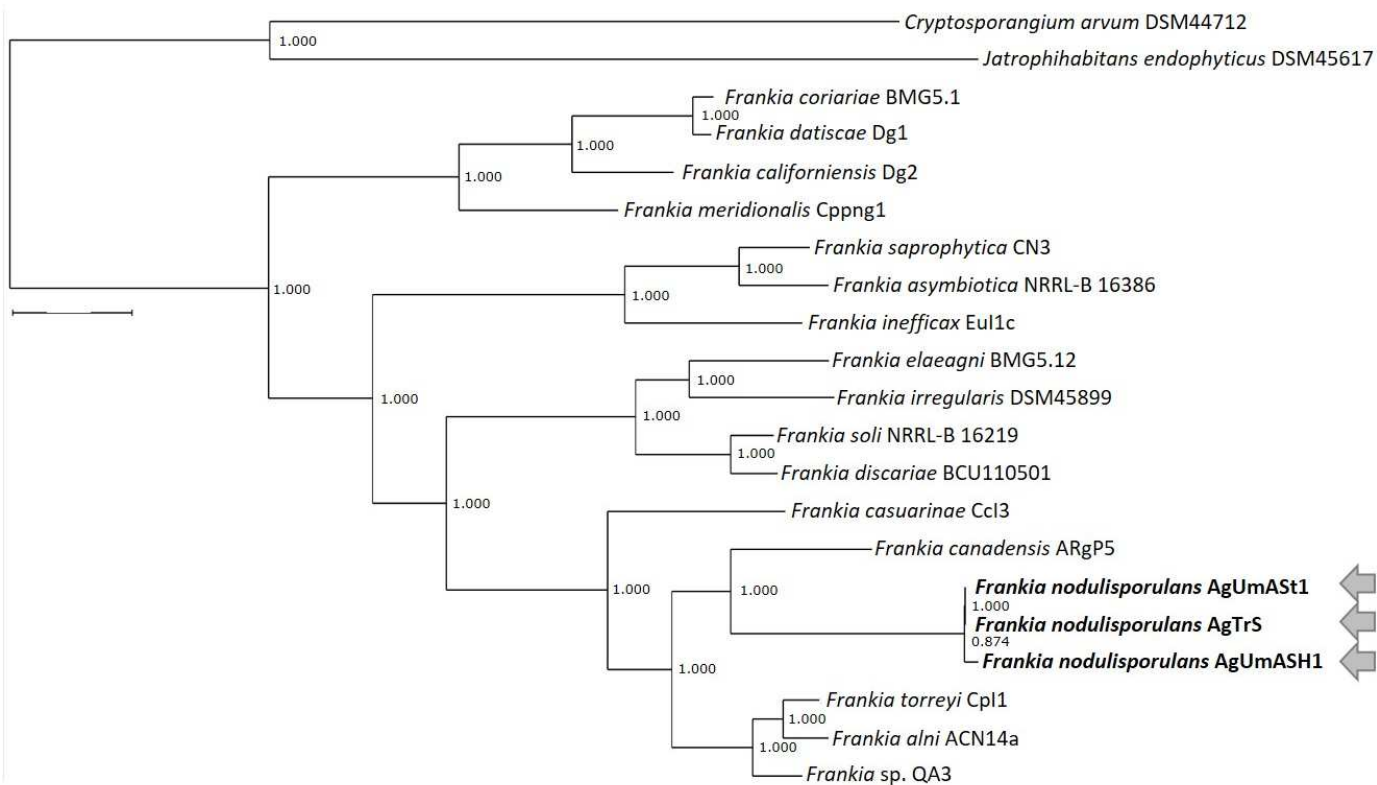
17
18
19

20

(A)



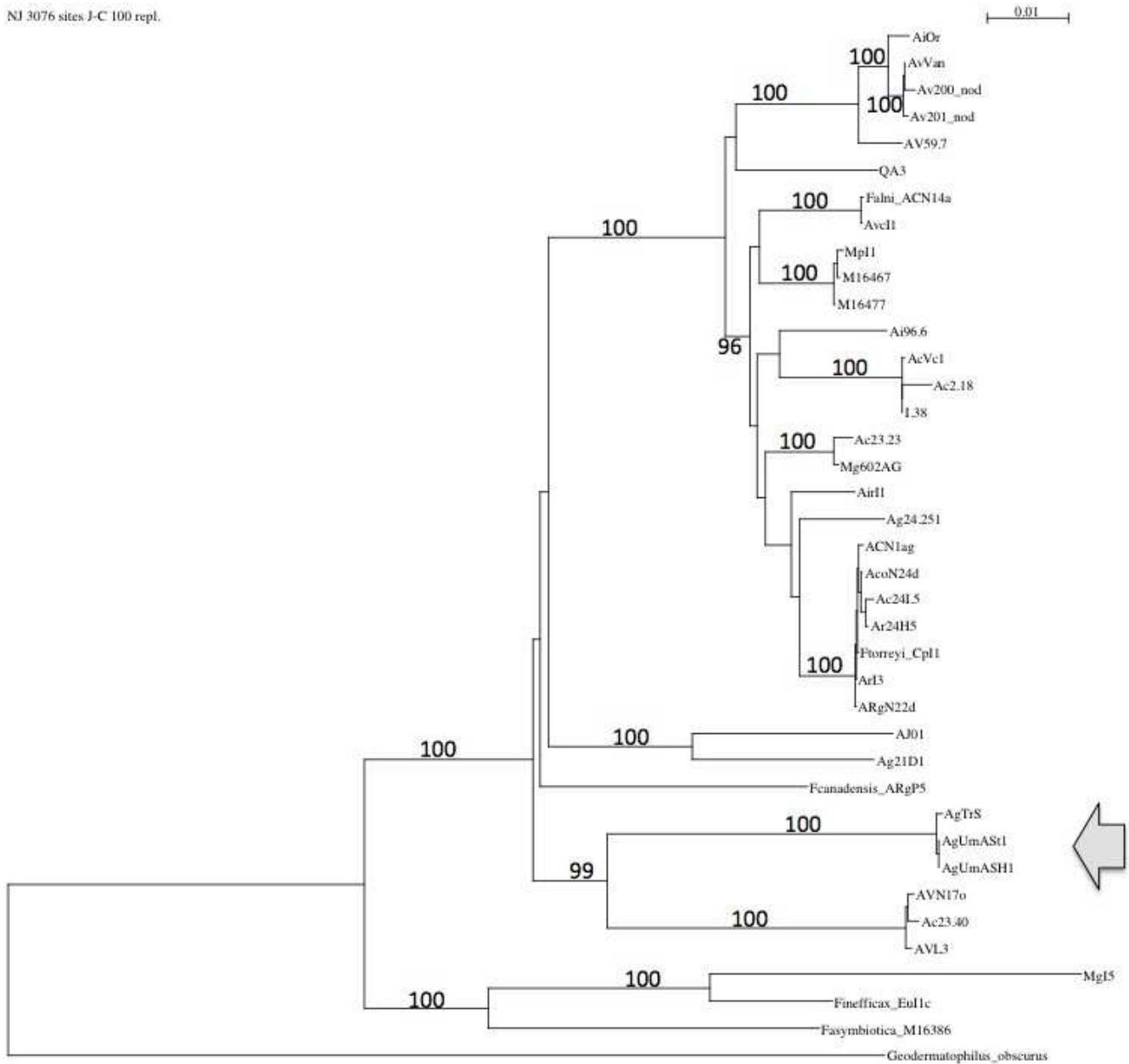
(B)



52 Figure 2. (A) Phylogenetic tree based on 16S rDNA sequences inferred using CLUSTAL X [15] as
53 implemented on SEAVIEW [9]. Phylogenetic trees were constructed by the Maximum Likelihood [9]
54 method. A bootstrap confidence analysis was performed with 1,000 replicates to determine the reliability
55 of the distance tree topology obtained [7], only those results above 95% are indicated. (B) Phylogenetic
56 tree for 20 genomes, built out of a core of 531 genes per genome, 10,620 in total, by EDGAR 2.0 [2, 3]
57 using the FastTree software (<http://www.microbesonline.org/fasttree/>) to generate approximately-
58 maximum-likelihood phylogenetic trees [37, 38]. The values at the branches show local support values
59 computed by FastTree using the Shimodaira-Hasegawa test. The core has 199469 amino acid residues
60 per genome, 3,989,380 in total. Genomes used were *Candidatus* Frankia datisca Dg1^T (NC_015656.1),
61 *Frankia coriariae* BMG5.1^T (NZ_JWIO00000000.1), *Candidatus* Frankia californiensis Dg2^T
62 (FLUV00000000.1), *Candidatus* Frankia meridionalis Cppng1^T (CADDZT010000001-
63 CADDZT010000101), *Frankia asymbiotica* M16386^T (NZ_MOMC00000000.1), *Frankia saprophytica*
64 CN3^T (NZ_AGJN00000000.2), *Frankia inefficax* Eul1c^T (NC_014666.1), *Frankia irregularis* DSM45899^T
65 (NZ_FAOZ00000000.1), *Frankia elaeagni* BMG5.12^T (NZ_ARFH00000000.1), *Frankia discariae*
66 BCU110501^T (NZ_ARDT00000000.1), *Frankia soli* NRRL B-16219^T (MAXA00000000.1), *Frankia*
67 *casuarinae* Ccl3^T (NC_007777.1), *Frankia canadensis* ARgP5^T (GCF_900241035.1), *Candidatus* Frankia
68 *nodulisporulans* AgTrS^T (NZ_CADCWS00000000.1), *Candidatus* Frankia *nodulisporulans* AgUmAst1
69 (CADDZU010000001-CADDZU010000211), *Candidatus* Frankia *nodulisporulans* AgUmASH1
70 (CADDZW010000001-CADDZW010000231), *Frankia alni* ACN14a^T (NC_008278.1), *Frankia torreyi*
71 Cpl1^T (NZ_JYFN00000000.1) and *Frankia sp.* QA3^T (AJWA00000000). The genome of *Cryptosporangium*
72 *arvum* DSM44712 (JFBT01000000) was used as the outgroup. ~~Bootstrap values are 100 for every branch;~~
73 ~~they were calculated in R (<http://www.R-project.org/>) [48] using the packages APE [35] and phangorn [41].~~
74 The size bar denotes 0.5 changes.

75 For both A and B panels, numbers in shaded boxes on the right correspond to the previously described
76 clusters in *Frankia* genus [26]. The arrows point at the three strains of *Candidatus* Frankia *nodulisporulans*.

77



79

80 Figure 3. *Alnus*-infective *Frankia* multilocus sequence analysis (MLSA) phylogeny. A Maximum
 81 Likelihood tree was inferred from the concatenation of 5 markers (*atpD*, *dnaA*, *ftsZ*, *pgk*, *rpoB*)
 82 into a super-alignment as previously described [37]. Only *Alnus*-infective strains were included in
 83 order to obtain optimal resolution for Cluster 1a. The bar denotes a 0.01 substitution/site distance.
 84 Numbers reported at nodes are the aLRT SH-like branch support. The arrow points at the three
 85 strains of *Candidatus* *Frankia nodulisporulans*.

86

87

REVISED VERSION**Tables****Table 1. Genomic description of AgTrS^T, AgUmASt1 and AgUmASH1 *Frankia* strains.**

Strains	AgTrS ^T	AgUmASt1	AgUmASH1
No. of contigs	281 (≥ 5 kb)	211 (≥ 3 kb)	231 (≥ 3 kb)
Genomic GC content (%)	71.6	71.3	71.2
Genomic length (bp)	4,882,652	4,315,411	4,261,328
BUSCO completeness (%)	94.1%	93.9 %	90.5 %
of 148 single core genes	140	139	134
No. of protein-encoding genes	4,388	3,709	3,681
No. of secondary metabolite clusters	16	12	10
Genome accession	NZ_CADCWS000000000.1	CADDZU010000001	CADDZW010000001
16S accession	MT023539	LR778176	LR778180
Host plant origin	<i>Alnus glutinosa</i>	<i>Alnus glutinosa</i> / <i>Alnus incana</i> *	<i>Alnus glutinosa</i> / <i>Alnus incana</i> *

* AgUmASt1 and AgUmASH1 were sequenced from nodules of *A. glutinosa* induced by the so-called “local source” of *Frankia*. The “local source” designated an inoculum made from crushed nodules of *A. incana* collected in Sweden and propagated for many years on *A. incana* as well as *A. glutinosa* plants grown under greenhouse conditions [14, 45].

11 **Table 2. Median ANI values between genomes of strains from *Frankia* Cluster 1 (including all published sequenced genomes in Cluster 1). All**
 12 **currently known *Frankia* Cluster 1 species are represented.**

13

	<i>Frankia nodulisporulans</i> AgTrS ^T	<i>Frankia nodulisporulans</i> AgUmAS1	<i>Frankia nodulisporulans</i> AgUmASH1	<i>Frankia alni</i> ACN14A ^T	<i>Frankia alni</i> Avcl1	<i>Frankia</i> sp. QA3 ^T	<i>Frankia torreyi</i> ACN1ag	<i>Frankia torreyi</i> Cp11 ^T	<i>Frankia canadensis</i> ARgP5 ^T	<i>Frankia casuarinae</i> Ccl6	<i>Frankia casuarinae</i> Ccl3 ^T	<i>Frankia casuarinae</i> Thr	<i>Frankia casuarinae</i> Allo2	<i>Frankia casuarinae</i> BR	<i>Frankia casuarinae</i> CeD	<i>Frankia casuarinae</i> KB5	
14	<i>Frankia nodulisporulans</i> AgTrS ^T	100.0	100.0	100.0	78.0	77.8	78.0	77.8	77.7	78.5	77.2	77.3	77.1	77.2	77.0	77.3	77.1
	<i>Frankia nodulisporulans</i> AgUmAS1	100.0	100.0	100.0	78.1	77.9	78.0	78.0	77.8	78.6	77.2	77.4	77.2	77.3	77.1	77.3	77.1
	<i>Frankia nodulisporulans</i> AgUmASH1	100.0	100.0	100.0	78.0	78.0	78.0	78.0	77.9	78.5	77.2	77.3	77.1	77.3	77.0	77.3	77.1
15	<i>Frankia alni</i> ACN14A ^T	78.8	79.0	78.9	100.0	99.7	90.0	92.1	92.1	79.7	79.9	79.9	79.7	79.8	79.8	79.8	79.5
	<i>Frankia alni</i> Avcl1	78.6	78.9	78.9	99.7	100.0	89.9	92.0	92.0	79.6	80.0	79.9	79.6	79.9	79.8	79.8	79.5
16	<i>Frankia</i> sp. QA3 ^T	79.0	79.2	79.2	89.9	89.8	100.0	90.8	90.7	80.2	80.4	80.5	80.2	80.4	80.5	80.4	80.2
	<i>Frankia torreyi</i> ACN1ag	78.8	78.8	79.0	92.1	92.0	90.8	100.0	99.9	79.6	79.9	80.0	79.7	79.8	79.9	79.8	79.5
	<i>Frankia torreyi</i> Cp11 ^T	78.9	78.7	78.9	92.2	92.0	90.9	99.9	100.0	79.5	80.0	80.1	79.9	79.9	79.9	79.9	79.7
17	<i>Frankia canadensis</i> ARgP5 ^T	79.3	79.5	79.5	79.4	79.4	79.9	79.6	79.4	100.0	78.4	78.3	78.3	78.4	78.3	78.4	78.1
	<i>Frankia casuarinae</i> Ccl6	77.2	77.2	77.4	79.1	79.0	79.8	79.3	79.3	77.7	100.0	100.0	100.0	99.9	99.9	99.8	99.8
18	<i>Frankia casuarinae</i> Ccl3 ^T	77.2	77.2	77.4	79.2	79.0	79.8	79.3	79.4	77.6	100.0	100.0	100.0	99.9	99.9	99.8	99.8
	<i>Frankia casuarinae</i> Thr	77.2	77.3	77.4	79.3	79.1	79.9	79.3	79.4	77.7	100.0	100.0	100.0	99.9	99.9	99.8	99.8
	<i>Frankia casuarinae</i> Allo2	77.1	77.2	77.4	79.1	79.0	79.8	79.2	79.2	77.6	99.9	99.9	99.9	100.0	99.8	99.8	99.7
19	<i>Frankia casuarinae</i> BR	77.0	77.2	77.3	79.1	79.0	79.8	79.2	79.2	77.6	99.9	99.9	99.9	99.8	100.0	99.7	99.7
	<i>Frankia casuarinae</i> CeD	77.1	77.1	77.4	79.2	79.0	79.8	79.2	79.3	77.6	99.8	99.8	99.8	99.8	99.7	100.0	99.7
	<i>Frankia casuarinae</i> KB5	77.1	77.1	77.4	79.1	78.8	79.7	79.1	79.2	77.6	99.8	99.8	99.8	99.7	99.7	99.7	100.0

20

21

22

23

24

25

26

27

Table 3. Descriptions of *Candidatus* Frankia **nodulisporulans** sp. nov.

Genus name	<i>Frankia</i>
Species name	<i>Candidatus</i> Frankia nodulisporulans
Specific epithet	nodulisporulans
Species status	sp. nov. [Candidatus]
Species etymology	no.du.li.spo'ru.lans. L. masc. n. nodulus a small knot; N.L. pres. part. sporulans sporulating; N.L. part. adj. nodulisporulans sporulating inside nodules.
Description of the new taxon and diagnostic traits	The species does not grow in the tested growth media, induces the formation of nodules when inoculated on <i>Alnus glutinosa</i> , <i>A. incana</i> or <i>A. alnobetula</i> under greenhouse conditions and sporulates in infected cells in cortical tissues. AgTrS is the type strain. Its genome (accession numbers: NZ_CADCWS000000000.1) has a G-C% of 71.2-71.6 and a length of 4.5-5 Mb. The species can be found in root nodules of alder in lowland regions of France and in shores of rivers and sea of northern Sweden.
Country of origin	France (and Sweden)
Region of origin	French Alps (and Swedish Västerbotten County)
Latitude (xx°xx'xx"N/S)	45°38' (63°50/63°50)
Longitude (xx°xx'xx"E/W)	5°52' (20°15/20°15)
Altitude (meters above sea level)	259 (10/12)
16S rRNA gene accession nr.	MT023539 (LR778176/LR778180)
Genome accession number [RefSeq; EMBL; ...]	NZ_CADCWS000000000.1 (CADDZU010000001/CADDZW010000001)
Genome status	Complete
Genome size	4,943 (4,315/4,261)
GC mol%	71.6 (71.3/71.2)
Number of strains in study	3
Source of isolation of non-type strains	The strains have not been isolated
Designation of the Type Strain	AgTrS ^T