



HAL
open science

Comparative genomics and proteogenomics highlight key molecular players involved in *Frankia* sporulation

Lorine Bethencourt, Hasna Boubakri, Najwa Taïb, Philippe Normand, J. Armengaud, Pascale P Fournier, Céline Brochier-Armanet, Aude P Herrera-Belaroussi

► To cite this version:

Lorine Bethencourt, Hasna Boubakri, Najwa Taïb, Philippe Normand, J. Armengaud, et al.. Comparative genomics and proteogenomics highlight key molecular players involved in *Frankia* sporulation. *Research in Microbiology*, 2019, 170 (4-5), pp.202-213. 10.1016/j.resmic.2019.04.002 . hal-02309396

HAL Id: hal-02309396

<https://hal.science/hal-02309396v1>

Submitted on 25 Oct 2021

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 4.0 International License

1 **Comparative genomics and proteogenomics highlight key molecular players involved in**
2 ***Frankia* sporulation**

3 Lorine Bethencourt¹, Hasna Boubakri¹, Najwa Taib², Philippe Normand¹, Jean Armengaud³,
4 Pascale Fournier¹, Céline Brochier-Armanet², Aude Herrera-Belaroussi^{1*}

5
6 ¹ Écologie Microbienne, Centre National de la Recherche Scientifique UMR 5557, Université de
7 Lyon, Université Claude Bernard Lyon I, INRA, UMR 1418, Villeurbanne 69622 Cedex, France

8 ² Univ Lyon, Université Claude Bernard Lyon 1, CNRS, UMR5558, Laboratoire de Biométrie et
9 Biologie Évolutive, 43 bd du 11 novembre 1918, F-69622, Villeurbanne, France.

10 ³ Laboratoire Innovations Technologiques pour la Détection et le Diagnostic (Li2D), Service de
11 Pharmacologie et Immunoanalyse (SPI), CEA, INRA, Bagnols sur Cèze, F-30207, France

12
13 Author for correspondence:

14 Herrera-Belaroussi Aude

15 UMR 5557 Écologie Microbienne

16 43, Boulevard du 11 novembre 1918, F-69622 Villeurbanne Cedex, France

17 Tel: (33) 472 448 200

18 E-mail: aude.herrera-belaroussi@univ-lyon1.fr

19 **Abstract**

20 Sporulation is a microbial adaptive strategy to resist inhospitable conditions for vegetative
21 growth and to disperse to colonise more favourable environments. This microbial trait is
22 widespread in Actinobacteria. Among them, *Frankia* strains are able to differentiate sporangia in
23 pure culture, while others can sporulate even when in symbiosis with sporulation occurring within
24 host cells. The molecular determinants controlling *Frankia* sporulation have not been yet
25 described. In order to highlight, for the first time, the molecular players potentially involved in
26 *Frankia* sporulation, we conducted (i) a comparison of protein contents between *Frankia* spores
27 and hyphae and (ii) a comparative genomic analysis of *Frankia* proteomes with sporulating and
28 non-sporulating Actinobacteria. Among the main results, glycogen-metabolism related proteins, as
29 well as oxidative stress response and protease-like proteins were overdetected in hyphae,
30 recalling lytic processes that allow *Streptomyces* cells to erect sporogenic hyphae. Several genes
31 encoding transcriptional regulators, including GntR-like, appeared up-regulated in spores, as well
32 as tyrosinase, suggesting their potential role in mature spore metabolism. Finally, our results
33 highlighted new proteins potentially involved in *Frankia* sporulation, including a pyrophosphate-
34 energized proton pump and YaaT, described as involved in the phosphorelay allowing sporulation
35 in *Bacillus subtilis*, leading us to discuss the role of a phosphorelay in *Frankia* sporulation.

36

37

38 **Keywords:** *Frankia* sporulation, actinorhizal symbiosis, desiccation, starvation, secondary
39 metabolites

40

41 1. Introduction

42 During their life cycle, microorganisms are subjected to point and/or recurrent environmental
43 variations that exert strong selection pressures. To face such physico-chemical variations, some
44 microorganisms have developed dormancy strategies. These strategies generally include
45 morphological differentiation and physiological switches that result in cells with a robust cell wall to
46 prevent water loss and reduced activity allowing the organisms to preserve their genetic material
47 and survive until conditions become favourable again. Elaborate developmental processes have
48 for example been characterized in a limited number of bacterial lineages, leading to the formation
49 of highly resistant spores with a much reduced metabolism, in response to stresses such as
50 desiccation, starvation or DNA damage. Yet, spore differentiation (sporulation) has been
51 characterized in a limited number of bacterial lineages, including Actinobacteria. Indeed, many
52 Actinobacteria can produce, in parallel to their hyphal growth, a variety of spores, e.g. free isolated
53 spores (e.g. *Micromonospora*) or spores arranged in chains (e.g. *Streptomyces*), sporangia with
54 motile (e.g. *Actinoplanes*) or non-motile (e.g. *Frankia*) spores [1, 2]. Whether resulting in the
55 production of free spores or sporangia, the sporulation process represents a heavy energetic
56 investment which is why the developmental program is tightly controlled [3, 4]. This process has
57 been extensively studied in *Streptomyces* species [5-8]. In *Streptomyces*, nutrient deprivation and
58 the resulting growth impairment lead to autolytic degradation of the mycelium (probably involving a
59 programmed cell death (PCD)-like mechanism), resulting in the accumulation of cell wall-derived
60 metabolites (amino acids, amino sugars, nucleotides, and lipids) around the lysing substrate
61 mycelium. This accumulation represents a major trigger allowing cells to acquire the building
62 blocks needed to erect sporogenic structures called aerial hyphae [1, 2]. This is also coupled to
63 synthesis of antibiotics and other secondary metabolites [9]. This correlation between sporulation
64 and secondary metabolites production may exist because of the need to defend the colony when it
65 is undergoing PCD, and both could be under the control of the same regulators [10, 11]. Two major
66 classes of regulatory genes were found very early in *Streptomyces coelicolor* to be involved in the
67 formation of aerial hyphae and their differentiation to form chains of unigenomic spores: the *bld*
68 genes (for “bald” in reference to the “hairless” phenotype of mutants lacking the fluffy aerial
69 hyphae) controlling the formation of aerial hyphae that are the precursors of spores (as well as

70 antibiotic production), and *whi* (“white”) genes involved in spore formation, named from the aerial
71 mycelium yielding white colonies caused by the lack of pigmented antibiotics [8, 12-14]. High levels
72 of additional proteins, such as SsgA-like proteins (SALPs) [15], have also been described to
73 control formation of sporulation septa, chromosome segregation and spore maturation processes
74 like spore wall synthesis and separation of spores.

75 Little research has been reported on sporulation morphogenesis and genetics of other
76 filamentous Actinobacteria, even though some of them show ability to sporulate in unusual
77 environmental conditions [1, 2, 16-19]. Indeed, it is now well known that sporulation allows bacteria
78 to resist conditions that are inhospitable to their vegetative form and/or to disperse to colonise
79 more favourable environments for growth. However, a case of bacterial sporulation falling outside
80 the paradigm described above has been described in a symbiotic context. Actinobacteria related to
81 *Frankia* genus are able to associate with so-called “actinorhizal” plants such as *Alnus* spp. This
82 symbiotic association results in the formation of a new root organ, the nodule, where trophic
83 exchanges between plant and bacteria take place: bacteria fix atmospheric nitrogen in dedicated
84 structures, named vesicles or diazovesicles, and in turn the host plant provides photosynthetic
85 carbon substrates. Most isolated *Frankia* strains have been described as sporulating in liquid
86 culture medium [20] and to form non-motile spores [21]. These will be referred as “Sp-“. Yet,
87 certain strains, called “Sp+”, have the ability to form numerous and massive sporangia inside host
88 cells (unlike “Sp-” strains that do sporulate *in-vitro* but are unable of *in-planta* sporulation) [22, 23]
89 wherein non-motile spores are found [24]. The expression of sporulation by Sp+ strains in plant
90 cells thus remains undescribed in a symbiotic context where *Frankia* benefits from a highly
91 favourable habitat for its development and trophic interactions with its host. In the absence of Sp+
92 strains available in pure culture and since no Sp+ genome has been sequenced to date, our
93 understanding of the molecular mechanisms of *Frankia* sporulation in plant cells remains very
94 limited. Unlike *Streptomyces*, environmental signals triggering *Frankia* sporulation, including
95 sporulation in culture media, and the cellular system(s) involved have been poorly studied, and
96 only with a few isolates in pure culture. Furthermore, depending of the strains used contradictory
97 results were obtained. For instance, depletion in nitrogen and phosphorus of the culture medium
98 has been described as significantly increasing *in-vitro* sporulation on some strains, but without any

99 effects on others [25]. Cultivating strain at 33°C can inhibit sporulation [25], while Dewedar and
100 Mansour [26] reported that some other *Frankia* strains showed maximum spore formation at 35°C.
101 Thus, to date, environmental factors involved in *Frankia* sporulation are poorly described and seem
102 to vary according to the *Frankia* strains. In addition, genetic determinants controlling the
103 expression of *Frankia* sporulation are still unknown. A rapid search for protein functional
104 annotations related to sporulation in *F. alni* ACN14a strain genome suggested the presence of *whi*
105 (*whiA* and *whiB*) and *ssgA* genes previously described in *S. coelicolor* genome [27]. However,
106 such an approach based on gene annotation can be incomplete and miss the genetic sporulation
107 determinants specific to *Frankia* genus.

108 The aim of the present study was therefore to identify the molecular determinants involved in
109 *Frankia* sporulation using two complementary approaches: (i) a comparison of protein contents
110 between *Frankia* spores and hyphae (***in-vivo* proteogenomic approach**) and (ii) a comparative
111 genomic analysis of *Frankia* proteomes and genomes with sporulating and non-sporulating
112 Actinobacteria data (***in-silico* approach**). In addition, since a correlation between sporulation and
113 secondary metabolite production may exist [10, 11], we also investigated genes potentially
114 involved in secondary metabolism. These complementary approaches allowed us to retrieve for
115 the first time a large number of proteins up-regulated in spores and to identify new genes
116 potentially involved in *Frankia* sporulation.

117

118

119 **2. Material and methods**

120

121 *2.1. Frankia culture conditions for spore isolation and hyphae preparation*

122 All cultures of *Frankia alni* strain ACN14a [28] were prepared in liquid BAP+ medium
123 supplemented with 5 mM ammonium in order to repress differentiation of diazovesicles as
124 described previously [29].

125 For spore isolation, cultures were prepared in 250 mL medium and used after 3- to 5-months
126 incubation (28°C, without agitation). The cells were sedimented (1,500 × g for 10 min) to obtain 5
127 mL pellets that were then syringed 3 times with 21G needles. The syringed cells were filtered on

128 compacted cotton in order to recover spores that flowed through the cotton while hyphal fragments
129 were retained in the plug. A total of 3 spore suspension replicates were obtained, with fresh
130 biomass yield between 16 mg and 22 mg.

131 For hyphae sample preparation, 10-day old cultures were used. This incubation period was
132 chosen to recover hyphae at the beginning of the stationary phase, just before sporulation onset
133 and absence of spores in cultures was confirmed under the microscope before analyses. The cells
134 were sedimented (1,500 × g for 5 minutes) and the resulting pellets frozen until proteomics
135 analysis. A total of 3 hyphal sample replicates were obtained, with biomass between 350 mg and
136 500 mg.

137

138 *2.2. Proteomic sample preparation and nanoscale liquid chromatography coupled to tandem mass* 139 *spectrometry (nano LC-MS/MS) analysis of tryptic peptides*

140 Proteins were extracted from hyphae as follows: each replicate sample was dissolved in LDS
141 1X (Invitrogen, Carlsbad, Ca) with 100 µL of LDS 1X per 30 mg of pellet. The samples were then
142 heated at 99°C for 5 min, subjected to sonication in an ultrasonic bath for 5 min, and bead-beated
143 with a Precellys instrument (Bertin Technologies, Montigny-le-Bretonneux, France). Samples were
144 centrifuged for 40 sec at 16,000 g and the resulting supernatants were heated for 10 min at 99°C.
145 Proteins were extracted from spores with a recently optimized protocol consisting in a dilution with
146 milliQ water into 1 mL final volume and bead-beating by means of a Precellys instrument (Bertin
147 technologies) [30]. The sample was then centrifuged, and the resulting supernatant was
148 precipitated with trichloroacetic acid (10% w/vol final). After centrifugation, the pellet was dissolved
149 into 50 µL of LDS1X (Invitrogen), subjected to sonication in an ultrasonic bath for 5 min, and then
150 heated for 10 min at 99°C. Hyphae and spore proteins were subjected to a short SDS-PAGE
151 migration. The whole proteome from a single polyacrylamide band was treated and proteolyzed by
152 trypsin, as previously described [31]. Tryptic peptides were analysed with a Q-Exactive HF high
153 resolution tandem mass spectrometer (ThermoFisher Scientific, Waltham, Ma) incorporating an
154 ultra-high-field orbitrap analyser as previously described [32]. MS/MS spectra were assigned with a
155 protein sequence database derived from *Frankia alni* ACN14a strain genome with standard
156 parameters.

157

158 *2.3. Analyses of differentially expressed proteins: Homology-based functional analysis, Clusters of*
159 *Orthologous Groups of proteins (COGs) prediction and genome mapping*

160 Proteins were quantified based on their spectral counts. Proteome comparison between hyphae
161 and spores was done taking into account the six replicates with the TFold module from the
162 PatternLab software, as previously described [31]. Differentially detected proteins between spores
163 and hyphae were categorized into 4 classes: i) Blue class proteins for which identifications
164 satisfied both, the fold change (*i.e.* ratio of protein quantities) between spores and hyphae (≥ 1.5)
165 and Tfold statistical criteria (t-test, p -value ≤ 0.05); ii) Orange class proteins for which
166 identifications did not meet the fold criterion but have low p -values; iii) Green class proteins for
167 which identifications satisfied the fold change criterion but not the statistical criterion; iv) and finally
168 Red class for which identifications did not meet the fold and p -value criteria.

169 The sequences of the 1,385 proteins of *F. alni* belonging to the blue class (for which
170 identifications satisfied both the fold >1.5 and p -value < 0.05) were uploaded in the STRING
171 database with *F. alni* chosen as a query microorganism, in order to access sequence identity.
172 These sequences were also assigned to COGs using MicroScope (Microbial Genome Annotation
173 & Analysis) Platform from Genoscope
174 (<http://www.genoscope.cns.fr/agc/microscope/home/index.php>).

175 A circular genome map of the genome sequence *F. alni* was drawn using GenVision software
176 from DNASTar (DNASTAR, Madison, WI).

177

178 *2.4. Construction of actinobacterial predicted proteome databases and identification of candidate*
179 *genes involved in Frankia sporulation by comparative proteome analysis*

180 Three databases were constructed from actinobacterial predicted proteomes downloaded from
181 NCBI website (<http://www.ncbi.nih.gov>). A first database, named “14_Frankia.db”, included the
182 protein sequences of 14 *Frankia* strains described as sporulating in pure culture (Table 1), among
183 which only four had complete genomes (*Frankia alni* ACN14a, *F. casuarinae* Ccl3, *Frankia*
184 EAN1pec and *F. inefficax* Eul1c). These four proteomes were used to construct a second
185 database referred as “actino_sporange.db” together with six other sporulating actinobacteria out of

186 the *Frankia* genus, chosen for their ability to sporulate by producing sporangia: four *Actinoplanes*
187 strains, one *Geodermatophilus* strain and one *Streptosporangium* strain (Table 1). The third
188 database (“actino_non_spor.db”) grouped the predicted proteomes of seven Actinobacteria unable
189 to differentiate spores, related to *Bifidobacterium*, *Rhodococcus* and *Micrococcus* genera (Table
190 1).

191 All-against-all blasts of protein sequences were performed for each database “14_Frankia.db”
192 and “actino_sporange.db” using the BLASTP software (version 2.2.26, with default parameters,
193 [33]). Protein families gathering homologous sequences were assembled using the SiLiX software
194 (version 1.2.9) [34], based on 45% identity over 80% of the length of the shortest sequence.
195 Several identity thresholds ranging from 35 to 95% were tested and 45% was found to be the best
196 compromise between false positives and false negatives within protein families (supplementary
197 data 1).

198 Based on the assumption that sporangia formation relied on similar (homologous) mechanisms
199 in sporangia-forming Actinobacteria, we searched for protein families largely distributed in *Frankia*
200 or sporangia-producing strains and absent in non-sporulating Actinobacteria. To do so, we
201 selected for each database homologous protein families represented in at least 80% of genomes
202 included in the respective database. The longest protein sequence of each protein family was then
203 used to query the “actino_non_spor.db” database using BLASTP (45% identity, e-value < 10⁻⁴), in
204 order to select groups without homologous protein sequences encoded in the genomes of non-
205 sporulating Actinobacteria.

206

207 2.5. Phyloprofile analysis of *Frankia alni* ACN14a gene content

208 *In-silico* analyses based on proteome comparisons of several *Frankia* strains with sporulating
209 and non-sporulating Actinobacteria were completed with analyses based this time on gene
210 contents using the “Gene phyloprofile” functionality of the MicroScope Platform from Genoscope
211 (<http://www.genoscope.cns.fr/agc/microscope/home/index.php>). This functionality (using pre-
212 computed homologies and synteny groups) allowed to find genes in the query genome of *Frankia*
213 *alni* ACN14a (i) with homologues in the genomes of 12 other sporangia-forming strains previously
214 used (*Frankia* Eul1c, Ccl3, QA3, EUN1f, EAN1pec and CN3 strains, the 4 *Actinoplanes*, the

215 *Geodermatophilus* and the *Streptosporangium* strains cited in Table 1), and (ii) without
216 homologues in the genomes of 7 non-sporulating actinobacterial strains (3 *Bifidobacterium* strains,
217 3 *Rhodococcus* strains and one *Micrococcus* strain cited in Table 1), using an identity threshold
218 >45% over 70% of the length of the shortest sequence.

219

220 2.6. Search for secondary metabolite gene clusters in *Frankia* genome

221 A total of 27 gene clusters encoding enzymes involved in production of secondary metabolites
222 have been identified in *Frankia* ACN14a genome (accession number NC_008278) using
223 AntiSMASH [35] running on the MicroScope Platform [36], including a total of 1,000 genes. These
224 clusters were searched for among (i) protein sequences differentially detected between *Frankia*
225 ACN14a spores and hyphae proteomes and (ii) families of interest selected from the comparative
226 genomic analyses (using AntiSMASH).

227

228 2.7. Proteomics data repository

229 The mass spectrometry raw files and interpreted proteomic data were deposited at the
230 ProteomeXchange Consortium (<http://proteomecentral.proteomexchange.org>) via the PRIDE
231 partner repository with the data set identifier PXD011620.

232

233

234 3. Results

235

236 3.1. Differentially expressed proteins between *Frankia alni* ACN14a spores and hyphae

237 High-throughput shotgun proteomic analyses were performed to compare protein contents
238 between *Frankia alni* ACN14a spores and hyphae. Proteins were quantified based on their spectral
239 counts in 3 biological replicates of spores and 3 biological replicates of hyphae. A total of 2,193
240 and 2,070 proteins were detected in hyphae and spores, respectively, when compiling the
241 replicates. When merging both spores and hyphae datasets, a total of 2,336 proteins was detected
242 (1,927 proteins were detected in both conditions). Using a Fold Change (FC) threshold of 1.5 and
243 a *p*-value below 0.05, 1,385 (59%) differentially detected proteins were identified (Blue class),

244 including 432 proteins more abundant in spores compared to hyphae and 953 up-regulated in
245 hyphae (supplementary data 2).

246 The genes corresponding to up-regulated proteins in spores and hyphae were scattered
247 throughout the genome of *Frankia* ACN14a and did not form gene clusters or genomic islands
248 (Fig. 1). Similarly, no correlation between protein abundances in spores *versus* hyphae and gene
249 location could be established (Fig. 1). Finally, these genes did not cluster with symbiotic genes
250 previously described: clockwise from the top squalene and phytoene biosynthesis genes in green,
251 the two hydrogenase uptake clusters in black, the iron-sulfur biosynthesis cluster in orange, the
252 nitrogenase cluster in red [29] and the cellulase-cellulose synthase cluster in blue [45].

253 The COG distribution of up-regulated proteins in spores compared to hyphae revealed (i) the
254 importance of COG related to “information storage and processing” (29% of proteins against 17%
255 in hyphae), and (ii) a decreased number of COG related to “cellular processes and signalling” and
256 “metabolism” (19% and 49% of proteins against 28% and 55% in hyphae, respectively). The
257 detailed distribution of proteins into COG functional categories is illustrated in Fig. 2. The most
258 overabundant proteins in spores compared to hyphae included COG-J “Translation, ribosomal
259 structure and biogenesis (17.5% against 5.4% in hyphae), COG-C “Energy production and
260 conversion” (15.4% against 9.3% in hyphae) and COG-K “Transcription” (10.5% against 6.0% in
261 hyphae). In contrast, the most overdetected proteins in hyphae were categorized into COG-E
262 “Amino acid transport and metabolism” (12.9% against 6.6% in spores), COG-M “Cell
263 wall/membrane/envelope biogenesis” (8.0% against 3.8% in spores) and COG-O “Posttranslational
264 modification, protein turnover, chaperones” (7.4% against 4.5% in spores).

265 Table 2 lists the most 30 up- and down-regulated proteins in *Frankia* ACN14a spores compared
266 to hyphae, respectively. Among the top proteins enriched in spores, 9 proteins were at least 10
267 times more abundant compared to hyphae (Table 2), including a putative phytase (FRAAL1518,
268 FC = 30.167) and a tyrosinase (FRAAL0531, FC = 20.333). Several transcriptional regulators were
269 also observed as PadR-like (FRAAL2570, FC = 15,571), GntR-like (FRAAL1728, FC = 11.571)
270 and TetR/AcrR-like (FRAAL2521, FC = 11.071).

271 Conversely, the 30 top proteins over-detected in hyphae were at least 10 times more abundant
272 compared to spores (Table 2). They included among others: (i) some wall-related proteins and cell

273 division determinants (e.g. FRAAL1858 aminopeptidase N, FC = 46.380 and FRAAL2205 cell
274 division protein DivIVA 5, FC = 16.900), (ii) proteins involved in glucose metabolism (e.g.
275 FRAAL0052 glucose-6-phosphate isomerase, FC = 22.710; FRAAL1567 6-phosphogluconate
276 dehydrogenase, FC = 17.75; FRAAL4578 6-phosphogluconolactonase, FC = 10.670; FRAAL5091
277 glucokinase, FC = 10.080), (iii) proteins involved in glutamine metabolism (e.g. FRAAL4683
278 carbamoyl phosphate synthase, FC = 20.000; FRAAL5855 glutamyl-tRNA amidotransferase, FC =
279 10.250; FRAAL5165 glutamine synthetase, FC = 10.000), and (iv) proteins involved in oxidative
280 stress response (e.g. FRAAL5116 rubrerythrin, FC = 29.286 and FRAAL1783 glutathione
281 peroxidase, FC = 12.353).

282 Proteins involved in glycogen metabolism were also found overabundant in hyphae compared
283 to spores such as FRAAL2118 glycogen debranching enzyme (FC = 10.375) (Table 2) or
284 FRAAL2116 malto-oligosyltrehalose trehalohydrolase (FC = 4.060), FRAAL2117 trehalose
285 synthase (FC = 2.030) and FRAAL5902 glycogen branching enzyme (FC = 1.850) (supplementary
286 data 2).

287

288 3.2. Selection of genes potentially involved in the *Frankia* sporulation by comparative proteome 289 analysis

290 This approach was based on the hypotheses that proteins involved in *Frankia* sporangia
291 differentiation would be part of a core proteome shared between sporulating *Frankia* strains and
292 other phylogenetically close Actinobacteria able to differentiate sporangia, and that they would be
293 encoded by homologous genes (hypothesis 1), and would not have homologues in Actinobacteria
294 unable to sporulate (hypothesis 2).

295 Under these hypotheses, two different databases were constructed from the predicted
296 proteomes of *Frankia* and other actinobacterial strains selected for their ability to differentiate
297 sporangia (Table 1). The first database included only *Frankia* proteins (named "14_Frankia.db",
298 81,518 protein sequences). This database could be queried to identify core protein families in
299 *Frankia*. The second (named "actino_sporange.db", 72,406 protein sequences) could be used to
300 identify proteins families shared between *Frankia* and sporangia-forming Actinobacteria. It
301 gathered the four *Frankia* predicted proteomes annotated as complete and six other sporangia-

302 forming actinobacterial predicted proteomes outside genus *Frankia*. For each database,
303 homologous protein families were delineated using SILIX program: a total of 1,037 and 1,641
304 protein families were obtained for the “actino_sporange.db” and the “14_Frankia.db” databases,
305 respectively.

306 These 1,037 and 1,641 protein families reflected the core proteome shared between *Frankia*
307 closely related sporangia-forming Actinobacteria proteomes, and between *Frankia* proteomes,
308 respectively. They were compared to a third proteomic database (named “actino_non_spor.db”)
309 gathering the protein sequences of seven Actinobacteria unable to sporulate (Table 1). Sequence
310 similarity-based comparisons allowed identifying 39 and 221 protein families from
311 “actino_sporange.db” and “14_Frankia.db” databases, respectively, without homologues in the
312 “actino_non_spor.db” database and were considered as strong candidates potentially involved in
313 *Frankia* sporulation (supplementary data 3). Among them, 25 and 139 (from “actino_sporange.db”
314 and “14_Frankia.db” databases, respectively) presented functional annotations in ACN14a or Ccl3
315 *Frankia* genomes (not “hypothetical proteins”).

316 Fourteen protein families with functional annotations were found in common from both
317 “actino_sporange.db” and “14_Frankia.db” databases (Table 3). Two of them included sequences
318 with functional annotations undoubtedly related to bacterial sporulation. The protein family #7
319 contained sequences annotated "Sporulation and cell division protein SsgA" (e.g. FRAAL2127,
320 FRAAL4594), with SsgA-like proteins (SALPs) previously described to control spore differentiation
321 in *S. coelicolor* [15]. The protein family #8 included "Stage 0 sporulation protein YaaT" annotated
322 proteins, YaaT protein being involved in the phosphorelay cascade allowing the activation of the
323 DNA binding protein regulator Spo0A in *B. subtilis* [37]. It is worth noting that these sequences
324 were annotated "hypothetical protein" in all *Frankia* proteomes used in this study and could only be
325 identified as YaaT protein sequences based on *Actinoplanes* and *Streptosporangium* protein
326 annotations. The other 12 groups of homologous proteins corresponded to enzyme categories with
327 very broad cell functions (e.g. decarboxylase, phosphatase, phosphotransacetylase,
328 metallophosphoesterase ...) and without evident or specific involvement in sporulation.

329

330 **3.3. Common genes identified from both comparative proteome and proteogenomic analyses**

331 Six proteins present in *Frankia* ACN14a proteome were found to be potentially involved in
332 sporulation based on both comparative genomic and proteogenomic approaches (Table 4). Among
333 them, 3 proteins were 1.7 to 4.6 times overabundant in spores compared to hyphae, including:
334 FRAAL1006 3-polyprenyl-4-hydroxybenzoate decarboxylase (UbiD), FRAAL6577 pyrophosphate-
335 energized proton pump (H⁺-PPase) and FRAAL6045 hypothetical protein. Three genes were 1.8
336 to 4.6 times under-expressed in spores: FRAAL2019 glycyl-tRNA synthetase, FRAAL3803 putative
337 sulfotransferase and FRAAL1171 putative secreted protein phosphodiesterase domain.

338

339 3.4. Additional analyses of *Frankia alni* ACN14a gene content using “Gene Phyloprofile” 340 functionality

341 In the present study, two alternative methods were performed to identify genes potentially
342 involved in *Frankia* sporulation (*i.e.* well distributed in *Frankia* and other sporulating actinobacterial
343 strains and absent in non-sporulating Actinobacteria). In parallel to the *in-silico* analyses based on
344 proteome comparisons, additional and rapid analyses (using “Gene phyloprofile” functionality of
345 the MicroScope Platform) were also conducted based this time on *Frankia alni* ACN14a gene
346 contents, with respect to the genomes of other sporangia-forming (including *Frankia* strains Eul1c,
347 Ccl3, QA3, EUN1f, EAN1pec and CN3, 4 *Actinoplanes* strains, one *Geodermatophilus* and one
348 *Streptosporangium* strains) and non-sporulating Actinobacteria (Table 1).

349 The results showed that 57 genes in *Frankia* ACN14 sequenced genome (Table 5) had
350 homologues in the 12 genomes of other sporangia-forming actinobacterial (Table 1), but not in the
351 genomes of seven actinobacterial strains unable to sporulate. Among them, we found four genes
352 that were also recovered following our previous comparative genomic analyses: FRAAL3803
353 annotated “putative sulfotransferase”, FRAAL5155 annotated “peptidase S51-dipeptidase E”,
354 FRAAL6080 annotated “putative Phosphotransacetylase” and FRAAL6577 annotated
355 “pyrophosphate-energized proton pump (H⁺-PPase)”. We also noticed the presence of
356 FRAAL5240 annotated “transcriptional regulator BldD”, which was described as a master regulator
357 able, when bound to tetrameric cyclic-di-GMP, to repress the transcription of genes involved in
358 *Streptomyces* sporulation [2, 38].

359 It's worth to note that when the genome of *S. coelicolor* A3 strain was included in the
360 phyloprofile analysis as the most studied model of sporulating Actinobacteria, 10 genes among the
361 57 were not recovered (Table 5), suggesting that the differentiation of sporangia in *Frankia*,
362 *Actinoplanes* and *Geodermatophilus* strains would involve slightly different genetic mechanisms
363 from those controlling production of spores in *Streptomyces*.

364

365 *3.5. Further investigation of the FRAAL6577 pyrophosphate-energized proton pump (H⁺-PPase) as*
366 *a potentially Frankia sporulation involved protein*

367 One protein systematically emerged as a strong candidate for *Frankia* sporulation from all *in-*
368 *vivo* and *in-silico* performed analyses (proteogenomics, comparative genomics and phyloprofile
369 analyses): the FRAAL6577 pyrophosphate-energized proton pump (H⁺-PPase).

370 We systematically examined the genomic environment of the candidate genes, looking in
371 particular for syntons or groups of neighbouring genes co-inherited across large evolutionary
372 distances that presumably are involved in a related function. In particular, FRAAL6577 forms a
373 synton comprising among others a putative septum site determining protein (FRAAL6565) and a
374 DNA topoisomerase I (FRAAL6579). The protein encoded by FRAAL6577 gene is downregulated
375 when ACN14a *Frankia* strain is in symbiosis with actinorhizal host plant (FC=0.04), while those of
376 the other ones are not modified [29], and downregulated in early steps of contact after 64h
377 (FC=0.28; unpublished), while the other ones are not significantly modified.

378

379 *3.6. Secondary metabolite gene clusters and Frankia sporulation*

380 Twenty-seven secondary metabolite clusters, encompassing 1,000 genes, have been predicted
381 in the *Frankia* ACN14a genome. None of these genes were found in the 14 protein families of
382 interest identified through the comparative proteome analysis (Table 3). In contrast, 50 proteins
383 (11.6%) belonging to predicted secondary metabolites clusters were found among the 432
384 upregulated proteins in spores, and 119 (12.5%) among the 953 upregulated proteins in hyphae
385 (supplementary data 4). Furthermore, some of the most upregulated proteins in both spores and
386 hyphae were found related to secondary metabolite biosynthesis, such as the third and fourth most
387 upregulated proteins in spores FRAAL2348 (hypothetical protein) and FRAAL2570 (Putative PadR

388 family transcriptional regulator), respectively, or the first and the third most upregulated proteins in
389 hyphae, FRAAL1858 (aminopeptidase N) and FRAAL 6507 (Terpene cyclase), respectively.

390 Of all the 27 secondary metabolite clusters predicted in *Frankia* ACN14a genome, only cluster
391 #17, involved in bacteriocin biosynthesis, was not represented among upregulated proteins in
392 spores or hyphae. Proteins included in the 26 other clusters were found overdetected in spores or
393 hyphae. The most represented secondary metabolite clusters were involved in the biosynthesis of
394 terpene (41% of the specified proteins were overdetected in spores or hyphae): cluster #27, with 4
395 and 3 upregulated proteins in spores and hyphae, respectively, and cluster #4, with 1 and 8
396 upregulated proteins in spores and hyphae, respectively.

397 Proteins related to Type I Polyketide synthase (t1pks) biosynthesis (cluster #13) were seen only
398 in spores, while proteins involved in other terpene biosynthesis (secondary metabolite clusters #3
399 and #10), t1pks (secondary metabolite clusters #1 and #14), lantipeptide (secondary metabolite
400 cluster #12) and other pks (secondary metabolite clusters #6 and #15) were found only in hyphae.

401

402

403 **4. Discussion**

404 Sporulation is a microbial adaptive trait that is widespread in Actinobacteria, that live for the
405 most part in soils where they are subject to cycles of desiccation and wetness, and abundance and
406 scarcity of nutrients. Different patterns and molecular mechanisms of sporulation are distributed
407 among the various Actinobacteria genera [1], revealing a wide diversity among this phylum.
408 Although sporulation morphogenesis and genetics have been largely investigated in *Streptomyces*,
409 sporulation in other filamentous Actinobacteria remained overlooked, even though some of them
410 show original ability to sporulate in unusual environmental contexts, such as *Frankia* strains [39]. In
411 the case of *Frankia*, studies on sporulation mainly aimed at understanding environmental factors
412 influencing *in-vitro* sporulation, spore germination, and spore ability to infect actinorhizal plants [25,
413 26]. But the molecular mechanisms and genetic determinants controlling *Frankia* sporulation
414 remain largely misunderstood, in large part due to the lack of a genetic transformation system [40].
415 In the present study, we investigated this issue by coupling comparative genomic and
416 proteogenomic approaches.

417

418 *4.1. Hypothetical pattern of Frankia sporulation early stages*

419 *Frankia* hyphae analysed through proteogenomics were recovered from 10-days old cultures at
420 the beginning of the stationary phase. This time point just before sporulation onset (the absence of
421 spores in cultures was confirmed under the microscope before analyses) was chosen in order to
422 understand which mechanisms could play during the slowing down of vegetative growth,
423 eventually followed by the initiation of sporogenic structure formation.

424 A large number of proteins involved in cell division and glucose and glutamine metabolisms
425 were more abundant in hyphae compared to spores, suggesting that even at the beginning of the
426 stationary phase, hyphae would still be growing and metabolically active. However, several
427 observations also suggested that cells would be getting ready to sporulate. Regarding glucose
428 metabolism, glycogen deposition has previously been detected in several *Streptomyces* spp. (*S.*
429 *antibioticus*, *S. fluorescens*, *S. griseus* and *S. viridochromogenes*: [41], and *S. venezuelae*: [42]),
430 localized in two particular cell-types: (i) in the hyphal region from which aerial branches emerge (up
431 to 20% of total cellular dry weight, [41] and (ii) in the tips of aerial hyphae undergoing sporulation.
432 More precisely, sections of *S. coelicolor* aerial hyphae including the junction between sporulating
433 and non-sporulating parts showed that glycogen was present only in the sporulating parts [43]. The
434 abundance of proteins involved in glycogen-metabolism in hyphae suggests that glycogen
435 accumulation and degradation play an important part in morphological differentiation of *Frankia*
436 spores as previously observed for *Streptomyces*. Second, several proteins involved in oxidative
437 stress response (e.g. rubrerythrin, the second most overabundant protein in hyphae, and
438 glutathione peroxidase) or associated to protease activities (e.g. FRAAL1858 “aminopeptidase” or
439 FRAAL5033 “serine protease”) were more abundant in hyphae. Genes with peptidase-related
440 annotation were also highlighted in *Frankia* proteomes as potentially involved in sporulation, based
441 on comparative genomic analyses of sporulating and non-sporulating Actinobacteria (e.g.
442 FRAAL5155 “peptidase S51-dipeptidase E” corresponding to the protein family #16). All these
443 proteins related to oxidative stress and protease activities recall the lytic process and the
444 extracellular protease cascade linked to starvation that allow *Streptomyces* cells to acquire the
445 building blocks needed to erect aerial hyphae [1, 2]. In *Frankia* strains, oxidative stress response

446 and protease activities occurring in the dying zones of the hyphal network could trigger, similarly to
447 *Streptomyces*, the sporogenic process, although aerial hyphae have never been observed in
448 *Frankia*. It is worth to note that the rodlin and chaplin determinants associated in *Streptomyces*
449 with the formation of aerial hyphae have not been found in the *Frankia* genome. However, the
450 cellulose synthase also associated with aerial hyphae formation [44] has been identified in the
451 *Frankia alni* genome and upregulated by plants exudates [45] but it is not overabundant in the
452 fractions. Up to now, only few studies have been focused on *Frankia* sporulation [25, 26], none of
453 them have been focussed on the early stages of this process and the sporogenic structures still
454 need to be investigated.

455 Interestingly, a conserved gene cluster (FRAAL2190-2212) comprising murEFXCDG-ftsIWZ-
456 sepF-divIVA-ispA that has 8 proteins identified, all of which overabundant in hyphae. This is
457 confirmatory for the proteogenomic approach since all these determinants are important for polar
458 growth of hyphal tips, which is a hallmark of hyphae [46]. Two distant homologues of polar growth
459 determinants known in *Streptomyces* (Scy, FilP) have also been identified as more abundant in
460 hyphae than in spores (Scy, FRAAL3687; FilP, FRAAL5914 - supplementary data 2). Conversely,
461 the mreBCD region associated with lateral wall formation in rod-shaped bacteria but with
462 sporulation in *Streptomyces* [47] was found to be overabundant in spores (FRAAL1917) as well as
463 several neighbouring genes that form a highly conserved synton in sporulating actinobacteria
464 (FRAAL1910-1921).

465 In parallel to the proteomic analyses of *Frankia* spores and hyphae, *in-silico* comparisons of
466 *Frankia* proteomes and genomes with sporulating and non-sporulating Actinobacteria data allowed
467 us to highlight several genes potentially involved in the early stages of *Frankia* sporangia
468 differentiation. For instance, phyloprofile analyses revealed the presence of the transcriptional
469 regulator BldD in *Frankia* genomes. This regulator was described in *Streptomyces* as a master
470 regulator able, when bound to tetrameric cyclic-di-GMP, to repress the transcription of genes for
471 many key developmental regulatory proteins, including early stage sporulation genes *whiB*, *whiG*,
472 *ssgA* and *ssgB* that were also present in *Frankia* genomes [2, 48, 49]. In *Streptomyces*, BldD-
473 cyclic di-GMP therefore functions as a 'brake' to prevent sporulation. This regulator was also
474 described as involved in sporangia differentiation in *Actinoplanes*: a *bldD* mutant formed

475 morphologically abnormal sporangia and earlier than in the wild type (supporting the role of BldD
476 as a repressor) [50]. In our experiments, BldD was not detected in *Frankia* spores or hyphae
477 through proteomic analyses (it could be absent in hyphae at the beginning of the stationary phase
478 and in mature spores). Another gene also highlighted as potentially involved in early stages of
479 *Frankia* sporangia differentiation was the gene encoded “sporulation and cell division SsgA
480 protein” (FRAAL2127 and FRAAL4594). SsgA-like proteins (SALPs) were largely described in
481 *Streptomyces* to control formation of sporulation septa, chromosome segregation and spore
482 maturation processes [8, 15]. Since the presence of several *ssgA* genes was previously described
483 in *F. alni* ACN14a genome [27, 51], its detection was expected in our study, proving the relevance
484 of both comparative genomic and proteogenomic approaches.

485

486 4.2. Potential role of phosphorelay in *Frankia* sporulation

487 One of the main results in this study is the potential role of a *Frankia* gene annotated
488 “pyrophosphate-energized proton pump” (H⁺-PPase, FRAAL6577, gene name *hppa*), highlighted
489 by both comparative genomic and proteomic analyses. H⁺-PPase has been described in
490 *Streptomyces* to convert energy from pyrophosphate hydrolysis into active H⁺ transport across the
491 plasma membrane [52]. The energy of phosphate hydrolysis used as a driving force for proton
492 movement across cell membrane could play an important role in *Frankia* sporulation. In addition to
493 FRAAL6577, the most abundant protein in spores compared to hyphae was described as a
494 putative phytase (FRAAL1518, FC = 30.167), a subgroup of phosphatases that catalyse the
495 hydrolysis of phytate through a series of myo-inositol phosphate intermediates. In parallel, kinase-
496 and phosphate transporter- related proteins were significantly overdetected in spores compared to
497 hyphae, such as FRAAL0753 “serine/threonine kinase” (FC = 7.667) and FRAAL4942 “phosphate
498 ABC transporter substrate-binding protein” (FC = 4.440). All these elements could indicate that
499 phosphorelay-like mechanisms could be involved in *Frankia* sporulation, as it happens in other
500 sporulating bacteria. Indeed, entry into sporulation in *Bacillus subtilis* is governed by a
501 phosphorelay in which phosphoryl groups from a histidine kinase are successively transferred via
502 relay proteins to the response regulator Spo0A [53]. The presence of *yaaT*-like gene (FRAAL6582)
503 in *Frankia* genome (absent in non-sporulating Actinobacteria), described as involved in the

504 phosphorelay cascade allowing the activation of Spo0A in *B. subtilis* [37], supports this hypothesis
505 (although no homologue of *spoA* was found in *Frankia* genomes). Under this hypothesis, it is also
506 interesting to point out that FRAAL6577 forms a synton comprising a putative septum site
507 determining protein (FRAAL6565) and a DNA topoisomerase I (FRAAL6579), two types of proteins
508 which could be involved in spore differentiation and whose regulation could be depending on
509 FRAAL6577. This discovery calls for new proteomics experiments devoted to a comprehensive
510 characterization of the phosphoproteome dynamics during sporulation, a study requiring
511 enrichment of phosphopeptides and their precise quantitation.

512

513 4.3. *Frankia* mature spores: metabolic activity and secondary metabolites

514 Overabundant proteins in *Frankia* spores compared to hyphae were examined to understand
515 molecular mechanisms occurring in the later stages of *Frankia* sporulation (well after the septation
516 of the sporogenic structures to form sporangia containing mature spores). It is worth to note that
517 proteins categorized in COG-C "metabolism" were detected among the most overabundant
518 proteins in *Frankia* mature spores, suggesting that these spores were not totally dormant (although
519 less active than hyphae). This result is consistent with previous descriptions of *Frankia* spores
520 exhibiting low levels of endogenous respiration at least ten-fold lower than the endogenous
521 respiration rate of vegetative cells [54].

522 A GntR family regulator was observed among the most overexpressed proteins in spores
523 (FRAAL1728, FC = 11.571). GntR family regulator is a poorly characterized transcriptional
524 regulator. GntR-like *whiH* gene of *S. coelicolor* has been described to play a crucial role in the
525 septation of aerial hyphae during sporulation [55]. In *Frankia alni* ACN14a genome, the GntR
526 family regulator was homologous to a FtsK/SpoIIIE-like protein from *Bacillus* involved in protein
527 secretion [56] and was found in a plasmid replication synton. This GntR family regulator, which is
528 overabundant in spores, could play in *Frankia* an important role, either (i) in mature spore
529 metabolism, although its exact function would need to be investigated, or (ii) in earlier stages of
530 sporangia differentiation (for instance to permit protein provision to spore wall) and then it would
531 persist in mature spores for specific reasons that should be investigated.

532 In many Actinobacteria, among them *Streptomyces*, antibiotics and a wide range of other
533 secondary metabolites can be produced alongside spore production. Some of the regulators of
534 morphological differentiation also affect secondary metabolite production, indicating regulatory
535 links between these processes [10, 11]. In this study, we did not observe clear evidence of a
536 higher involvement of secondary metabolites in *Frankia* spores or hyphae (11.6% of upregulated
537 proteins in spores belonged to predicted secondary metabolites clusters, against 12.5% in
538 hyphae). However, few elements caught our attention. For instance, the presence of tyrosinase as
539 the second most overdetected protein in spores (FRAAL0531, FC = 20.333) was quite unexpected.
540 Indeed, in *Streptomyces*, tyrosinase is involved in the biosynthesis of melanin by transforming the
541 tyrosine into L-DOPA (3, 4-dihydroxy phenyl-L-alanine), which is further converted into
542 dopachrome and spontaneously oxidized to indole-5, 6-quinone. The later polymerizes
543 spontaneously into DOPA-melanin, a dark brown pigment [57] thought to protect against UV and
544 free radicals [58]. No black pigment in *Frankia alni* ACN14a has ever been reported *in-vitro*,
545 suggesting no melanin is present. In *Frankia* spores, tyrosinase could therefore be involved in
546 other processes such as auxin transformation [59] or reactive quinone derivatives synthesis [60,
547 61].

548

549 4.4. Conclusions

550 The aim of the present study was to set up a comprehensive and consistent database of
551 sporulation-associated determinants. It is important to keep in mind that the link between candidate
552 genes and proteins discussed here and their effective role in *Frankia* sporulation has to be
553 confirmed using functional studies, for instance through gene complementation and tracking their
554 expression during sporulation through RT-qPCR, as functional shifts may have occurred during
555 Actinobacteria diversification. Indeed, in *Streptomyces*, the spore production process is initiated by
556 the production of aerial hyphae with a hardened tip by the action of a cellulose-like synthase [44]
557 and the synthesis of a lipid-rich envelope comprising an aminated hopanoid to render the spores
558 mildly hydrophobic and permit them to float [62]. Some of these processes linked to sporulation in
559 *Streptomyces* appear to have been diverted in other lineages such as in *Frankia* where the
560 cellulose synthase is upregulated in response to contact and the hopanoid machinery is geared to

561 formation of nitrogen-fixing vesicles. More precisely, the role of the hopanoid layers surrounding
562 *Frankia* vesicles is to provide a barrier to oxygen diffusion that would otherwise destroy the
563 oxygen-labile nitrogenase [63]. It is thus quite strikingly different from its role in *Streptomyces*.
564 Similar molecular determinants can therefore be found in different actinobacterial lineages, but the
565 fulfilled functions of these genes may have been diverted from their initial role.

566 It might also be interesting to complete this study with metabolic analyses of *Frankia* spores and
567 hyphae (e.g. primary and secondary metabolic extractions followed by HPLC and/or GC analyses),
568 allowing for instance to detect differences in secondary metabolite contents between both types of
569 cells. Additional shotgun proteomics and phosphoproteomics analyses performed at different
570 stages of *Frankia* development (e.g. including hyphae recovered at stationary phase or young
571 spores) could also provide a better understanding of molecular mechanisms involved in *Frankia*
572 sporulation at different intermediate stages.

573 Finally, once the role in *Frankia* sporulation of the different proteins highlighted in this study
574 will be better understood, it would be interesting to further investigate their expression in a
575 symbiotic context, when *Frankia* sporulation occurs inside host plant cells. It would help to
576 elucidate what molecular factors could suppress the sporulation capacity of Sp- *Frankia* strains *in-*
577 *planta* and allow in Sp+ strains the expression of sporulation inside nodules.

578

579

580 **Acknowledgements**

581 The authors thank Danis ABROUK (University of Lyon, UMR-CNRS 5557, IBIO platform) for
582 his help with drawing of the circular genome map.

583

584

585 **References**

586 [1] Locci R, Petrolini Baldan B. On the spore formation process in actinomycetes. I.
587 Morphology and development of *Streptoverticillium* species as examined by
588 scanning electron microscopy. Friulian J Sci 2010; 13:17-31.

- 589 [2] Barka EA, Vatsa P, Sanchez L, Gaveau-Vaillant N, Jacquard C, Meier-Kolthoff JP, et al.
590 Taxonomy, physiology, and natural products of Actinobacteria. *Microbiol Mol Biol*
591 *Rev* 2015; 80:1-43.
- 592 [3] de Hoon MJL, Eichenberger P, Vitkup D. Hierarchical evolution of the bacterial
593 sporulation network. *Curr Biol* 2010; 20:R735–R745.
- 594 [4] Larsson J. *Streptomyces* sporulation - Genes and regulators involved in bacterial cell
595 differentiation. Lund University, Lund. 2010.
- 596 [5] Horinouchi S, Beppu T. Hormonal control by A-factor of morphological development
597 and secondary metabolism in *Streptomyces*. *Proc Jpn Acad Ser B Phys Biol Sci* 2007;
598 83:277-295.
- 599 [6] Flärdh K, Buttner MJ. *Streptomyces* morphogenetics: dissecting differentiation in a
600 filamentous bacterium. *Nat Rev Microbiol* 2009; 7:36-49.
- 601 [7] McCormick J, Flärdh K. Signals and regulators that govern *Streptomyces*
602 development. *FEMS Microbiol Rev* 2012; 36:206-231.
- 603 [8] Chater KF. Recent advances in understanding *Streptomyces*. *F1000 Res* 2016; 5:2795.
- 604 [9] Rioseras B, Shliha PV, Gorshkov V, Yagüe P, López-García MT, Gonzalez-Quiñonez N,
605 et al. Quantitative proteome and phosphoproteome analyses of *Streptomyces*
606 *coelicolor* reveal proteins and Pphosphoproteins modulating differentiation and
607 secondary metabolism. *Mol Cell Proteomics* 2018 17:1591-1611.
- 608 [10] Elliot M, Damji F, Passantino R, Chater K, Leskiw BK. The *bldD* gene of *Streptomyces*
609 *coelicolor* A3(2): a regulatory gene involved in morphogenesis and antibiotic
610 production. *J Bacteriol* 1998; 180:1549-1555.
- 611 [11] López-García MT, Yagüe P, González-Quiñonez N, Rioseras B, Manteca A. The
612 SCO4117 ECF Sigma factor pleiotropically controls secondary metabolism and
613 morphogenesis in *Streptomyces coelicolor*. *Front Microbiol* 2018; 9:312.
- 614 [12] Hopwood DA, Wildermuth H, Palmer HM. Mutants of *Streptomyces coelicolor*
615 defective in sporulation. *J Gen Microbiol* 1970; 61:397-408.
- 616 [13] Chater KF. A morphological and genetic mapping study of white colony mutants of
617 *Streptomyces coelicolor*. *J Gen Microbiol* 1972; 72:9-28.
- 618 [14] Chater KF. Regulation of sporulation in *Streptomyces coelicolor* A3(2): a checkpoint
619 multiplex? *Curr Opin Microbiol* 2001; 4:667-673.
- 620 [15] van Wezel G, van der Meulen J, Kawamoto S, Luiten R, Koerten H, Kraal B. *ssgA* is
621 essential for sporulation of *Streptomyces coelicolor* A3(2) and affects hyphal
622 development by stimulating septum formation. *J Bacteriol* 2000; 182:5653–5662.
- 623 [16] Stevens R. Fine structure of sporogenesis and septum formation in *Micromonospora*
624 *globosa* Kriss and *M. fusca* Jensen. *Can J Microbiol* 1975; 21:1081-1088.

- 625 [17] Ostash B, Yushchuk O, Tistechok S, Mutenko H, Horbal L, Muryl A, et al. The adpA-like
626 regulatory gene from *Actinoplanes teichomyeticus*: in silico analysis and
627 heterologous expression. *World J Microbiol Biotechnol* 2015; 31:1297-1301.
- 628 [18] Buttner M. *Actinoplanes* swims into the Molecular Age. *J Bacteriol* 2017; 25:199.
- 629 [19] Mouri Y, Jang M, Konishi K, Hirata A, Tezuka T, Ohnishi Y. Regulation of sporangium
630 formation by the orphan response regulator TcrA in the rare actinomycete
631 *Actinoplanes missouriensis*. *Mol Microbiol* 2018; 107:718-733.
- 632 [20] Normand P, Orso S, Cournoyer B, Jeannin P, Chapelon C, Dawson J, et al. Molecular
633 phylogeny of the genus *Frankia* and related genera and emendation of the family
634 *Frankiaceae*. *Int J Syst Bacteriol* 1996; 46:1-9.
- 635 [21] Nouioui I, Ghodhbane-Gtari F, Montero-Calasanz MD, Goker M, Meier-Kolthoff JP,
636 Schumann P, et al. Proposal of a type strain for *Frankia alni* (Woronin 1866) Von
637 Tubeuf 1895, emended description of *Frankia alni*, and recognition of *Frankia*
638 *casuarinae* sp. nov. and *Frankia elaeagni* sp. nov. *Int J Syst Evol Microbiol* 2016;
639 66:5201-5210.
- 640 [22] Pozzi AC, Bautista-Guerrero HH, Nouioui I, Cotin-Galvan L, Pepin R, Fournier P, et al.
641 In-plant sporulation phenotype: a major life history trait to understand the
642 evolution of *Alnus*-infective *Frankia* strains. *Environ Microbiol* 2015; 17:3125-3138.
- 643 [23] Pozzi AC, Bautista-Guerrero HH, Abby SS, Herrera-Belaroussi A, Abrouk D, Normand P,
644 et al. Multi-Locus Sequence Analysis and extensive sampling bring new insights on
645 *Frankia* phylogeny and phylogeography. *Syst Appl Microbiol* 2018; S0723-
646 2020(18)30113-9.
- 647 [24] van Dijk C, Merkus E. A microscopical study of the development of a spore-like stage
648 in the life cycle of the root nodule endophyte of *Alnus glutinosa* (L.) Gaertn. *New*
649 *Phytol* 1976; 77:73-91.
- 650 [25] Burleigh S, Dawson J. *In vitro* sporulation of *Frankia* strain HFPCc13 from *Casuarina*
651 *cunninghamiana*. *Can J Microbiol* 1991; 37:897-901.
- 652 [26] Dewedar A, Mansour S. Infection events in the establishment of *Casuarina-Frankia*
653 symbioses using spore inoculation. *Acta Oecolog* 1992; 13:379-385.
- 654 [27] Normand P, Lapierre P, Tisa LS, Gogarten JP, Alloisio N, Bagnarol E, et al. Genome
655 characteristics of facultatively symbiotic *Frankia* sp. strains reflect host range and
656 host plant biogeography. *Genome Res* 2007; 17:7-15.
- 657 [28] Normand P, Lalonde M. Evaluation of *Frankia* strains isolated from provenances of
658 two *Alnus* species. *Can J Microbiol* 1982; 28:1133-1142.
- 659 [29] Alloisio N, Queiroux C, Fournier P, Pujic P, Normand P, Vallenet D, et al. The *Frankia*
660 *alni* symbiotic transcriptome. *Mol Plant Microbe Interact* 2010; 23:593-607.

- 661 [30] Mappa C, Pible O, Armengaud J, Alpha-Bazin B. Assessing the ratio of *Bacillus* spores
662 and vegetative cells by shotgun proteomics. *Environ Sci Pollut Res Int* 2018;
663 10.1007/s11356-018-3341-z.
- 664 [31] Hartmann EM, Allain F, Gaillard JC, Pible O, Armengaud J. Taking the shortcut for
665 high-throughput shotgun proteomic analysis of bacteria. *Methods Mol Biol* 2014;
666 1197:275-285.
- 667 [32] Ghedira K, Harigua-Souiai E, Benhamda C, Fournier P, Pujic P, Guesmi S, et al. The
668 PEG-responding desiccome of the alder microsymbiont *Frankia alni*. *Sci Rep* 2018;
669 8:759.
- 670 [33] Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST
671 and PSI-BLAST: a new generation of protein database search programs. *Nucleic
672 Acids Res* 1997; 25:3389-3402.
- 673 [34] Miele V, Penel S, Duret L. Ultra-fast sequence clustering from similarity networks with
674 SiLiX. *BMC Bioinformatics* 2011; 12:116.
- 675 [35] Medema MH, Blin K, Cimermancic P, de Jager V, Zakrzewski P, Fischbach MA, et al.
676 antiSMASH: rapid identification, annotation and analysis of secondary metabolite
677 biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic
678 Acids Res* 2011; 39:W339–W346.
- 679 [36] Vallenet D, Calteau A, Cruveiller S, Gachet M, Lajus A, Josso A, et al. MicroScope in
680 2017: an expanding and evolving integrated resource for community expertise of
681 microbial genomes. *Nucleic Acids Res* 2017; 45:D517-D528.
- 682 [37] Hosoya S, Asai K, Ogasawara N, Takeuchi M, Sato T. Mutation in *yaaT* leads to
683 significant inhibition of phosphorelay during sporulation in *Bacillus subtilis*. *J Bacteriol*
684 2002; 184:5545–5553.
- 685 [38] Tschowri N, Schumacher M, Schlimpert S, Chinnam N, Findlay K, Brennan R, et al.
686 Tetrameric c-di-GMP mediates effective transcription factor dimerization to control
687 *Streptomyces* development. *Cell* 2014; 158:1136-1147.
- 688 [39] Torrey J. Endophyte sporulation in root nodules of actinorhizal plants. *Physiologia
689 Plantarum* 1987; 70:279-288.
- 690 [40] Kucho K, Hay A, Normand P. The determinants of the actinorhizal symbiosis. *Microb
691 Environ* 2010; 25:241-252.
- 692 [41] Brana AF, Manzanal MB, Hardisson C. Occurrence of polysaccharide granules in
693 sporulating hyphae of *Streptomyces viridochromogenes*. *J Bacteriol* 1980;
694 144:1139-1142.

- 695 [42] Ramade N, Vining LC. Accumulation of intracellular carbon reserves in relation to
696 chloramphenicol biosynthesis by *Streptomyces venezuelae*. *Can J Microbiol* 1993;
697 39:377–383.
- 698 [43] Plaskitt KA, Chatter KF. Influences of developmental genes on localized glycogen
699 deposition in colonies of a mycelial prokaryote, *Streptomyces coelicolor* A3(2): a
700 possible interface between metabolism and morphogenesis. *Phil Trans Roy Soc B*
701 1995; 347:105-121.
- 702 [44] Xu H, Chater KF, Deng Z, Tao M. A Cellulose Synthase-Like Protein Involved in Hyphal
703 Tip Growth and Morphological Differentiation in *Streptomyces*. *J Bacteriol* 2008;
704 190:4971–4978.
- 705 [45] Pujic P, Alloisio N, Fournier P, Roche D, Sghaier H, Armengaud J, et al. Omics of the
706 early molecular dialogue between *Frankia alni* and *Alnus glutinosa* and the
707 cellulase synton. *Env Microbiol* 2019; in press (EMI 14606).
- 708 [46] Hempel AM, Cantlay S, Molle V, Wang SB, Naldrett MJ, Parker JL, et al. The Ser/Thr
709 protein kinase AfsK regulates polar growth and hyphal branching in the filamentous
710 bacteria *Streptomyces*. *Proc Natl Acad Sci U S A* 2012; 109:E2371-2379.
- 711 [47] Kleinschnitz EM, Heichlinger A, Schirner K, Winkler J, Latus A, Maldener I, et al. Proteins
712 encoded by the *mre* gene cluster in *Streptomyces coelicolor* A3(2) cooperate in
713 spore wall synthesis. *Mol Microbiol* 2011; 79:1367-1379.
- 714 [48] Chandra G, Chater KF. Developmental biology of *Streptomyces* from the
715 perspective of 100 actinobacterial genome sequences. *FEMS Microbiol Rev* 2014;
716 38:345-379.
- 717 [49] Tschowri N, Schumacher M, Schlimpert S, Chinnam N, Findlay K, Brennan R, et al.
718 Tetrameric c-di-GMP mediates effective transcription factor dimerization to control
719 *Streptomyces* development. *Cell* 2014 158:1136-1147.
- 720 [50] Mouri Y, Konishi K, Fujita A, Tezuka T, Ohnishi Y. Regulation of sporangium formation
721 by BldD in the rare actinomycete *Actinoplanes missouriensis*. *J Bacteriol* 2017;
722 199:e00840-00816.
- 723 [51] Cotin-Galvan L. Relation plante-hôte / *Frankia* dans les symbioses actinorhiziennes :
724 cas particulier des souches non-isolables capables de sporuler in-planta. Lyon1,
725 Lyon. 2014.
- 726 [52] Hirono M, Mimura H, Nakanishi Y, Maeshima M. Expression of functional *Streptomyces*
727 *coelicolor* H⁺-pyrophosphatase and characterization of its molecular properties. *J*
728 *Biochem* 2005; 138:183–191.
- 729 [53] Russell JR, Cabeen MT, Wiggins PA, Paulsson J, Losick R. Noise in a phosphorelay
730 drives stochastic entry into sporulation in *Bacillus subtilis*. *EMBO J* 2017; 36:2856-2869.

- 731 [54] Krumholz GD, Chval MS, McBride MJ, Tisa LS. Germination and physiological
732 properties of *Frankia* spores. *Plant and Soil* 2003; 254:57-67.
- 733 [55] Ryding NJ, Kelemen GH, Whatling CA, Flardh K, Buttner MJ, Chater KF. A
734 developmentally regulated gene encoding a repressor-like protein is essential for
735 sporulation in *Streptomyces coelicolor* A3(2). *Mol Microbiol* 1998; 29:343-357.
- 736 [56] Garufi G, Butler E, Missiakas D. ESAT-6-like protein secretion in *Bacillus anthracis*. *J*
737 *Bacteriol* 2008; 190:7004-7011.
- 738 [57] Dastager SG, Wen-Jun L, Dayanand A, Shu-Kun T, Xin-Peg TX-YZ, Li-Hua X, et al.
739 Separation, Identification and analysis of pigment (Melanin) production in
740 *Streptomyces*. *Afr J Biotechnol* 2006; 5:1131-1134.
- 741 [58] Li C, Ji C, Tang B. Purification, characterisation and biological activity of melanin
742 from *Streptomyces* sp. *FEMS Microbiol Lett* 2018; 365:10.1093/femsle/fny1077
- 743 [59] Briggs WR, Ray PM. An auxin inactivation system involving tyrosinase. *Plant Physiol*
744 1956; 31:165-167.
- 745 [60] Rescigno A, Rinaldi AC, Sanjust E. Some aspects of tyrosine secondary metabolism.
746 *Biochem Pharmacol* 1998; 56:1089–1096.
- 747 [61] Dias FRF, Novais JS, Devillart T, da Silva W, Ferreira M, Loureiro R, et al. Synthesis and
748 antimicrobial evaluation of amino sugar-based naphthoquinones and isoquinoline-
749 5,8-diones and their halogenated compounds. *Eur J Med Chem* 2018; 156:1-12.
- 750 [62] Poralla K, Muth G, Hartner T. Hopanoids are formed during transition from substrate to
751 aerial hyphae in *Streptomyces coelicolor* A3(2). *FEMS Microbiology Letters* 2000;
752 189:93-95.
- 753 [63] Berry AM, Harriott OT, Moreau RA, Osman SF, Benson DR, Jones AD. Hopanoid lipids
754 compose the *Frankia* vesicle envelope, presumptive barrier of oxygen diffusion to
755 nitrogenase. *Proc Natl Acad Sci USA* 1993; 90:6091-6094.
- 756 [64] Lalonde M, Calvert H, Pine S. Isolation and use of *Frankia* strains in actinorhizae
757 formation, p. 296-299. *In* AH Gibson WN (ed.), *Current Perspectives in Nitrogen*
758 *Fixation*. Australian Academy of Science, Canberra. 1981.
- 759 [65] Baker D, Newcomb W, Torrey J. Characterization of an ineffective actinorhizal
760 microsymbiont, *Frankia* sp. Eu11 (Actinomycetales). *Canadian Journal of*
761 *Microbiology* 1980; 26:1072-1089.
- 762 [66] Wall LG, Beauchemin N, Cantor MN, Chaia E, Chen A, Detter JC, et al. Draft genome
763 sequence of *Frankia* sp. strain BCU110501, a nitrogen-fixing actinobacterium
764 isolated from nodules of *Discaria trinevis*. *Genome Announc* 2013; 1:e00503-00513.
- 765 [67] Ghodhbane-Gtari F, Beauchemin N, Bruce D, Chain P, Chen A, Walston Davenport
766 K, et al. Draft genome sequence of *Frankia* sp. strain CN3, an atypical, noninfective

767 (Nod-) ineffective (Fix-) isolate from *Coriaria nepalensis*. Genome Announc 2013;
768 1:e0008513.

769 [68] Mansour SR, Oshone R, Hurst SGt, Morris K, Thomas WK, Tisa LS. Draft genome
770 sequence of *Frankia* sp. strain Ccl6, a salt-tolerant nitrogen-fixing actinobacterium
771 isolated from the root nodule of *Casuarina cunninghamiana*. Genome Announc
772 2014; 2:e01205-01213.

773 [69] Ngom M, Oshone R, Hurst St, Abebe-Akele F, Simpson S, Morris K, et al. Permanent
774 draft genome sequence for *Frankia* sp. strain CeD, a nitrogen-fixing
775 actinobacterium isolated from the root nodules of *Casuarina equistifolia* grown in
776 Senegal. Genome Announc 2016; 4:e00265-00216.

777 [70] Hurst SGt, Oshone R, Ghodhbane-Gtari F, Morris K, Abebe-Akele F, Thomas WK, et al.
778 Draft genome sequence of *Frankia* sp. strain Thr, a nitrogen-fixing actinobacterium
779 isolated from the root nodules of *Casuarina cunninghamiana* grown in Egypt.
780 Genome Announc 2014; 2:e00493-00414.

781 [71] Oshone R, Ngom M, Abebe-Akele F, Simpson S, Morris K, Sy MO, et al. Permanent
782 draft genome sequence of *Frankia* sp. strain Allo2, a salt-tolerant nitrogen-fixing
783 actinobacterium isolated from the root nodules of *Allocasuarina*. Genome
784 Announc 2016; 4:e00388-00316.

785 [72] Sen A, Beauchemin N, Bruce D, Chain P, Chen A, Walston Davenport K, et al. Draft
786 genome sequence of *Frankia* sp. strain QA3, a nitrogen-fixing actinobacterium
787 isolated from the root nodule of *Alnus nitida*. Genome Announc 2013; 1:e0010313.

788 [73] Ghodhbane-Gtari F, Hurst SGt, Oshone R, Morris K, Abebe-Akele F, Thomas WK, et al.
789 Draft genome sequence of *Frankia* sp. strain BMG5.23, a salt-tolerant nitrogen-fixing
790 actinobacterium isolated from the root nodules of *Casuarina glauca* grown in
791 Tunisia. Genome Announc 2014; 2:e00520-00514.

792 [74] Ruckert C, Szczepanowski R, Albersmeier A, Goesmann A, Fischer N, Steinkamper A,
793 et al. Complete genome sequence of the actinobacterium *Actinoplanes friuliensis*
794 HAG 010964, producer of the lipopeptide antibiotic friulimycin. J Biotechnol 2014;
795 178:41-42.

796 [75] Yamamura H, Ohnishi Y, Ishikawa J, Ichikawa N, Ikeda H, Sekine M, et al. Complete
797 genome sequence of the motile actinomycete *Actinoplanes missouriensis* 431 (T) (=
798 NBRC 102363(T)). Stand Genomic Sci 2012; 7:294-303.

799 [76] Schwientek P, Neshat A, Kalinowski J, Klein A, Rückert C, Schneiker-Bekel S, et al.
800 Improving the genome annotation of the acarbose producer *Actinoplanes* sp.
801 SE50/110 by sequencing enriched 5'-ends of primary transcripts. J Biotechnol 2014;
802 190:85-95.

- 803 [77] Ivanova N, Sikorski J, Jando M, Munk C, Lapidus A, Glavina Del Rio T, et al. Complete
804 genome sequence of *Geodermatophilus obscurus* type strain (G-20). *Stand*
805 *Genomic Sci* 2010; 2:158-167.
- 806 [78] Nolan M, Sikorski J, Jando M, Lucas S, Lapidus A, Glavina Del Rio T, et al. Complete
807 genome sequence of *Streptosporangium roseum* type strain (NI 9100). *Stand*
808 *Genomic Sci* 2010; 2:29-37.
- 809 [79] Yu DS, Jeong H, Lee DH, Kwon SK, Song JY, Kim BK, et al. Complete genome
810 sequence of the probiotic bacterium *Bifidobacterium bifidum* strain BGN4. *J*
811 *Bacteriol* 2012; 194:4757-4758.
- 812 [80] O'Connell Motherway M, Zomer A, Leahy SC, Reunanen J, Bottacini F, Claesson MJ,
813 et al. Functional genome analysis of *Bifidobacterium breve* UCC2003 reveals type
814 IVb tight adherence (Tad) pili as an essential and conserved host-colonization
815 factor. *Proc Natl Acad Sci USA* 2011; 108:11217-11222.
- 816 [81] Lee JH, Karamychev VN, Kozyavkin SA, Mills D, Pavlov AR, Pavlova NV, et al.
817 Comparative genomic analysis of the gut bacterium *Bifidobacterium longum*
818 reveals loci susceptible to deletion during pure culture growth. *BMC Genomics*
819 2008; 9:247.
- 820 [82] MP. M, RL. W, WW. H, N. A, M. M, C. F, et al. The complete genome of *Rhodococcus*
821 sp. RHA1 provides insights into a catabolic powerhouse. *Proc Natl Acad Sci U S A*
822 2006; 103:15582-15587.
- 823 [83] Sekine M, Tanikawa S, Omata S, Saito M, Fujisawa T, Tsukatani N, et al. Sequence
824 analysis of three plasmids harboured in *Rhodococcus erythropolis* strain PR4. *Environ*
825 *Microbiol* 2006; 8:334-346.
- 826

827

Figure legends

828

829

830 **Fig. 1. Circular map of the genome of *Frankia alni* ACN14a spores and hyphae proteins.**

831 Form the outside in:

832 1-Scale with ticks at 1,000,000nt.

833 2-Selected genes that play a role in symbiosis with clockwise from the top squalene and phytoene
834 biosynthesis genes (green), the two hydrogenase uptake clusters (black), the iron-sulfur
835 biosynthesis cluster (orange), the cellulase cluster (blue) and the nitrogenase cluster (red).

836 3-Up-regulated genes (proteins that are more abundant in spores than in hyphae) in light blue.

837 4-Down-regulated genes (proteins that are more abundant in hyphae than in spores) in pink.

838

839

840 **Fig. 2. Distribution into COG functional categories of the 432 and 953 up-regulated proteins**
841 **in *Frankia alni* ACN14a spores (grey bars) and hyphae (white bars), respectively.**

Table 1. List of *Frankia* and other Actinobacteria predicted proteomes used in this study.

| Strains | Genome size | Number of genes | Number of proteins | Accession number | References |
|--|-------------|-----------------|--------------------|------------------|------------------------------|
| <u>Frankia strains</u> | | | | | |
| <i>Frankia alni</i> ACN14a | 7497934 | 6775 | 6700 | NC_008278 | [27] |
| <i>Frankia casuarinae</i> Ccl3 | 5433628 | 4618 | 4499 | NC_007777 | [27] |
| <i>Frankia</i> EAN1pec | 8982042 | 7377 | 7191 | NC_009921 | [64] |
| <i>Frankia inefficax</i> Eul1c | 8815781 | 7263 | 7083 | NC_014666 | [65] |
| <i>Frankia discariae</i> BCU110501 | 7887841 | 6047 | 5858 | NZ_ARDT00000000 | [66] |
| <i>Frankia saprophytica</i> CN3 | 9966592 | 7173 | 7010 | NZ_AGJN00000000 | [67] |
| <i>Frankia casuarinae</i> Ccl6 | 5578620 | 4669 | 4524 | NZ_AYTZ00000000 | [68] |
| <i>Frankia casuarinae</i> CeD | 5004459 | 4246 | 4112 | NZ_JPGU00000000 | [69] |
| <i>Frankia casuarinae</i> Thr | 5309830 | 4560 | 4439 | NZ_JENI00000000 | [70] |
| <i>Frankia casuarinae</i> Allo2 | 5352110 | 4605 | 4465 | NZ_JPHT00000000 | [71] |
| <i>Frankia</i> QA3 | 7578953 | 6546 | 6031 | NZ_AJWA00000000 | [72]* |
| <i>Frankia</i> EUN1f | 9352740 | 8236 | 8161 | NZ_ADGX00000000 | [64]* |
| <i>Frankia elaeagni</i> BMG5.12 | 7588936 | 5473 | 5320 | NZ_ARFH00000000 | [21] |
| <i>Frankia</i> BMG5.23 | 5352110 | 4478 | 4349 | NZ_JDWE00000000 | [73] |
| <u>Sporangia-forming Actinobacteria</u> | | | | | |
| <i>Actinoplanes friuliensis</i> DSM_7358 | 7376071 | 8648 | 8565 | NC_022657 | [74] |
| <i>Actinoplanes missouriensis</i> 431 | 8773466 | 8203 | 8124 | NC_017093 | [75] |
| <i>Actinoplanes</i> N902_109 | 9228054 | 8283 | 8212 | NC_021191 | Hu et al., unpublished |
| <i>Actinoplanes</i> SE50_110 | 9239851 | 8385 | 8247 | NC_017803 | [76] |
| <i>Geodermatophilum obscurus</i> DSM_43160 | 5322497 | 5155 | 4810 | NC_013757 | [77] |
| <i>Streptosporangium roseum</i> DSM_43021 | 10369518 | 9508 | 8975 | NC_013595 | [78] |
| <u>Non-sporulating Actinobacteria</u> | | | | | |
| <i>Bifidobacterium bifidum</i> BGN4 | 2223664 | 1902 | 1834 | NC_017999 | [79] |
| <i>Bifidobacterium breve</i> UCC2003 | 2422684 | 1985 | 1854 | NC_020517 | [80] |
| <i>Bifidobacterium longum</i> DJO10A | 2389526 | 2074 | 2001 | NC_010816 | [81] |
| <i>Rhodococcus jostii</i> RHA1 | 9702737 | 9221 | 9145 | NC_008271 | [82] |
| <i>Rhodococcus erythropolis</i> PR4 | 6895538 | 6511 | 6437 | NC_012490 | [83] |
| <i>Rhodococcus opacus</i> B4 | 8830572 | 8253 | 8191 | NC_012522 | Takarada et al., unpublished |
| <i>Micrococcus luteus</i> NTC_2665 | 2501097 | 2345 | 2236 | NC_012803 | Lucas et al., unpublished |

* = Normand, personal observations of sporulation

Table 2. List of the 30 most up- and down-regulated proteins in *Frankia alni* ACN14a spores (with the fold change FC between spores and hyphae)

| FRAAL accession | Functional description | FC (Sp/Hy) | p-value |
|-----------------------------|---|------------|----------|
| UPREGULATED PROTEINS | | | |
| FRAAL1518 | putative Phytase | 30.167 | 6.08E-11 |
| FRAAL0531 | tyrosinase | 20.333 | 5.09E-10 |
| FRAAL2348 | hypothetical protein | 18.625 | 4.14E-08 |
| FRAAL2570 | PadR family transcriptional regulator | 15.571 | 1.24E-06 |
| FRAAL1575 | alkylidihydroxyacetonephosphate synthase | 12.667 | 2.83E-08 |
| FRAAL1728 | GntR family transcriptional regulator | 11.571 | 4.45E-07 |
| FRAAL2521 | TetR/AcrR family transcriptional regulator | 11.071 | 7.59E-08 |
| FRAAL1401 | peroxidase | 10.385 | 7.22E-05 |
| FRAAL5299 | two-component system response regulator | 10.167 | 3.64E-04 |
| FRAAL3583 | hypothetical protein | 9.667 | 7.42E-04 |
| FRAAL6677 | hypothetical protein | 8.833 | 2.73E-09 |
| FRAAL5293 | oxidoreductase | 8.167 | 3.90E-05 |
| FRAAL0753 | serine/threonine protein kinase | 7.667 | 5.00E-06 |
| FRAAL0319 | hypothetical protein | 7.563 | 1.19E-12 |
| FRAAL4129 | enoyl-CoA hydratase | 6.500 | 1.06E-05 |
| FRAAL1434 | aromatic compound degradation protein Paal | 6.333 | 1.01E-07 |
| FRAAL1094 | MULTISPECIES: 30S ribosomal protein S14 type Z | 6.167 | 1.36E-07 |
| FRAAL0756 | C4-dicarboxylate ABC transporter substrate-binding protein | 6.167 | 1.51E-05 |
| FRAAL6513 | formylmethionine deformylase | 6.167 | 8.39E-06 |
| FRAAL2378 | hypothetical protein | 6.000 | 3.69E-06 |
| FRAAL3391 | branched-chain amino acid ABC transporter substrate-binding protein | 5.800 | 2.28E-07 |
| FRAAL2465 | levansucrase | 5.667 | 8.72E-07 |
| FRAAL4168 | iron ABC transporter | 5.636 | 8.81E-07 |
| FRAAL2867 | SAM-dependent methyltransferase involved in tRNA-Met maturation | 5.593 | 2.08E-06 |
| FRAAL3790 | amidohydrolase | 5.571 | 7.84E-06 |
| FRAAL0837 | hypothetical protein | 5.371 | 1.06E-08 |
| FRAAL1618 | hypothetical protein | 5.353 | 1.66E-06 |
| FRAAL6016 | GTPase | 5.333 | 1.06E-05 |
| FRAAL0773 | acyl-CoA dehydrogenase | 5.283 | 2.89E-10 |
| FRAAL5006 | methyltransferase | 5.222 | 7.35E-05 |

DOWN REGULATED PROTEINS

| | | | |
|-----------|--|---------|----------|
| FRAAL1858 | aminopeptidase N | -46.375 | 1.02E-10 |
| FRAAL5116 | rubrerythrin | -29.286 | 4.11E-09 |
| FRAAL6507 | Terpene cyclase | -22.833 | 1.76E-09 |
| FRAAL0052 | glucose-6-phosphate isomerase | -22.714 | 3.47E-10 |
| FRAAL4683 | carbamoyl phosphate synthase large subunit | -20.000 | 6.82E-09 |
| FRAAL1567 | 6-phosphogluconate dehydrogenase | -17.750 | 2.04E-10 |
| FRAAL2205 | cell division protein DivIVA | -16.900 | 3.44E-11 |
| FRAAL0413 | hypothetical protein | -15.857 | 8.82E-08 |
| FRAAL4574 | transaldolase | -15.643 | 1.03E-09 |
| FRAAL0943 | thiosulfate sulfurtransferase | -14.938 | 6.29E-11 |
| FRAAL2220 | hypothetical protein | -14.000 | 2.25E-10 |
| FRAAL6633 | glycosyl hydrolase | -14.000 | 9.95E-10 |
| FRAAL5004 | histidinol dehydrogenase | -13.500 | 2.77E-09 |
| FRAAL1783 | glutathione peroxidase | -12.353 | 2.08E-11 |
| FRAAL5778 | ribosome-recycling factor | -12.222 | 2.32E-08 |
| FRAAL1069 | MULTISPECIES: 50S ribosomal protein L7/L12 | -12.034 | 1.29E-09 |
| FRAAL5846 | 2-isopropylmalate synthase | -11.571 | 9.41E-09 |
| FRAAL5337 | (2Fe-2S)-binding protein | -11.556 | 1.87E-06 |
| FRAAL6527 | hypothetical protein | -11.500 | 2.70E-10 |
| FRAAL5111 | cytochrome C | -11.273 | 2.80E-08 |
| FRAAL6634 | hypothetical protein | -11.000 | 3.17E-06 |
| FRAAL4578 | 6-phosphogluconolactonase | -10.667 | 5.32E-10 |
| FRAAL5151 | dihydrolipoamide dehydrogenase | -10.576 | 2.64E-12 |
| FRAAL2118 | glycogen debranching protein | -10.375 | 1.05E-09 |
| FRAAL5855 | glutamy-tRNA amidotransferase | -10.250 | 5.20E-09 |
| FRAAL1424 | membrane protein | -10.231 | 8.28E-09 |
| FRAAL3174 | endoglucanase | -10.167 | 2.46E-07 |
| FRAAL5091 | glucokinase | -10.083 | 8.45E-10 |
| FRAAL5704 | ATP-binding protein | -10.000 | 8.15E-09 |
| FRAAL5165 | MULTISPECIES: glutamine synthetase | -10.000 | 6.46E-10 |

Table 3. List of the 14 protein families (group of homologous protein sequences) with relevant annotations, shared between the 4 selected sporulating *Frankia* proteomes and the 6 other sporangia-forming Actinobacteria proteomes and without homologous protein sequences in the proteomes of non-sporulating Actinobacteria.

| Group of homologous proteins | Locus_Tag <i>Frankia alni</i> ACN14 | Annotation in <i>Frankia alni</i> ACN14a genome | Locus_Tag <i>Frankia casuarinae</i> Ccl3 | Annotation in <i>Frankia casuarinae</i> CCl3 genome |
|------------------------------|---|---|---|---|
| 8 (YaaT) | FRAAL6582 | hypothetical protein; putative Signal peptidase-like protein | Francci3_4308 | PSP1 |
| 7 (SsgA) | FRAAL2127 | Putative regulator (partial) | Francci3_1359 | Sporulation and cell division protein SsgA |
| | FRAAL4594 | Sporulation and cell division protein SsgA | Francci3_1632 | Hypothetical protein |
| 2 | FRAAL1006 | 3-polyprenyl-4-hydroxybenzoate decarboxylase (UbiD) | Francci3_0515 | 4Hydroxybenzoate decarboxylase |
| 16 | FRAAL5155 | Peptidase S51, dipeptidase E | Francci3_3137 | Peptidase S51, dipeptidase E |
| 9 | FRAAL6577 | Pyrophosphate-energized proton pump (H ⁺ -PPase) | Francci3_4304 | Membrane-bound proton-translocating pyrophosphatase |
| 17 | FRAAL6682 | Putative Metallo-dependent phosphatase domain | Francci3_4384 | Metallophosphoesterase |
| 12 | FRAAL6080 | Conserved hypothetical protein; putative Phosphotransacetylase | Francci3_3823 | Hypothetical protein |
| 13 | FRAAL6045 | Conserved hypothetical protein | Francci3_3798 | Endonuclease/exonuclease/phosphatase |
| 11 | FRAAL3803 | Putative sulfotransferase | Francci3_2395 | Sulfotransferase |
| 6 | FRAAL2173 | Putative monooxygenase | Francci3_1397 | Spheroidene monooxygenase |
| 4 | FRAAL0930 | L-seryl-tRNA(Sec) selenium transferase (CysteinyI-tRNA(Sec) selenium transferase) | Francci3_0439 | Selenocysteine synthase |
| 18 | FRAAL6833 | ABC-transporter polyamine-binding lipoprotein | not present | |
| 10 | FRAAL2019 | Glycyl-tRNA synthetase | Francci3_1275 | Glycine-Trna Ligase |
| 15 | FRAAL1171 | Putative secreted protein, putative phosphodiesterase domain | Francci3_0663 | Ribonuclease BN |

Table 4. Common genes identified from both comparative genomic and proteogenomic analyses

| Locus_Tag <i>Frankia alni</i> ACN14 | Annotation in <i>Frankia alni</i> ACN14a genome | Proteomic FC spore/hyphae | p-value | Group code in the comparative genomic analysis |
|--|--|------------------------------|----------|--|
| <u>3 up-regulated proteins in spores:</u> | | | | |
| FRAAL1006 | 3-polyprenyl-4-hydroxybenzoate decarboxylase | 4.600 | 2.34E-09 | 2 |
| FRAAL6577 | Pyrophosphate-energized proton pump (H ⁺ -PPase) | 4.431 | 8.78E-07 | 9 |
| FRAAL6045 | Conserved hypothetical protein | 1.714 | 3.99E-02 | 13 |
| <u>3 down-regulated proteins in spores:</u> | | | | |
| FRAAL2019 | Glycyl-tRNA synthetase | -4.538 | 1.41E-10 | 10 |
| FRAAL3803 | Putative sulfotransferase | -2.375 | 2.23E-03 | 11 |
| FRAAL1171 | Putative secreted protein. putative phosphodiesterase domain | -1.857 | 2.14E-02 | 15 |

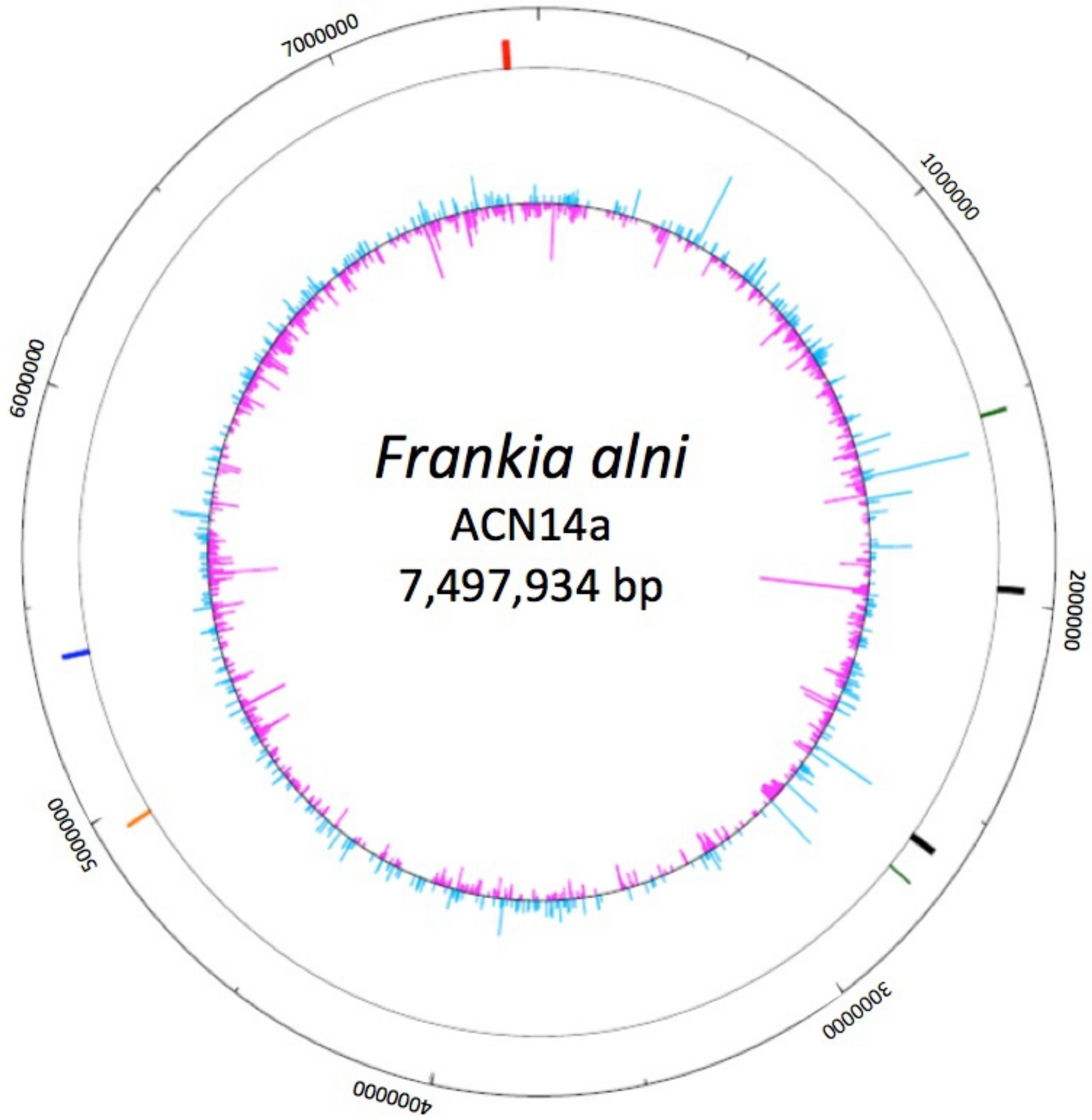
857
858
859
860
861
862

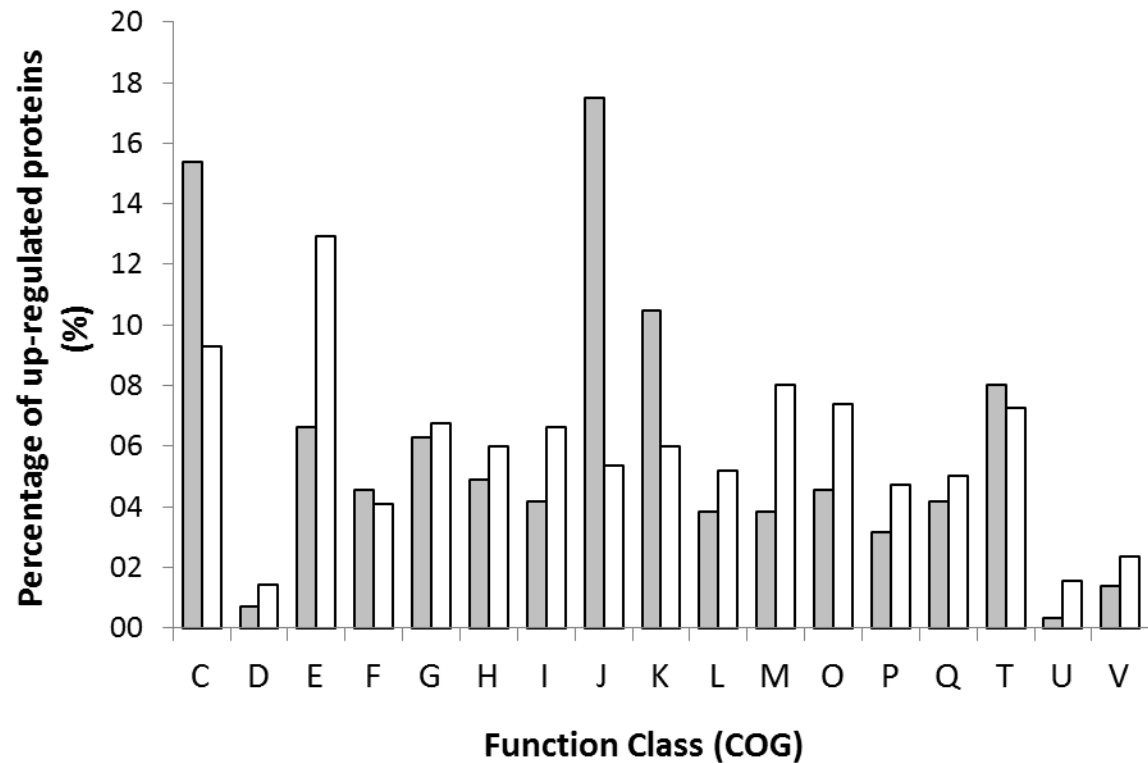
Table 5. Analyses of *Frankia alni* ACN14a gene content using “Gene Phyloprofile” functionality.

Frankia alni ACN14a genes with (i) homologs in the 12 genomes of other sporangia-forming actinobacterial (including *Frankia* strains Eu11c, Ccl3, QA3, EUN1f, EAN1pec and CN3, 4 *Actinoplanes* strains, one *Geodermatophilus* and one *Streptosporangium* strains, given in Table 1) and (ii) without homologs in the genomes of 7 actinobacterial strains unable of sporulation (Table1). FRAAL highlighted in gray were not recovered when genome of *Streptomyces coelicolor* A3 strain was included in the phyloprofile analysis among sporulating strains (as the most studied model of sporulating actinobacteria). FRAAL shown in bold were also highlighted in our comparative genomic analyses previously described (Table 3).

| Label | Gene | Product |
|-----------|--------------|---|
| FRAAL0106 | <i>pdxH</i> | pyridoxine 5'-phosphate oxidase |
| FRAAL0172 | <i>methH</i> | B12-dependent homocysteine-N5-methyltetrahydrofolate transmethylase |
| FRAAL0399 | – | conserved hypothetical protein; putative methyltransferase domain |
| FRAAL0617 | – | conserved hypothetical protein |
| FRAAL0996 | <i>fabH</i> | 3-oxoacyl-[acyl-carrier-protein] synthase III; acetylCoA ACP transacylase |
| FRAAL1269 | – | conserved hypothetical protein |
| FRAAL1272 | – | conserved hypothetical protein |
| FRAAL1303 | – | conserved hypothetical protein; putative Ribosome-associated factor |
| FRAAL1314 | – | Putative HTH-type transcriptional repressor, XRE family |
| FRAAL1538 | – | GroES-family molecular chaperone |
| FRAAL1606 | – | putative aminopeptidase |
| FRAAL1638 | – | conserved hypothetical protein; putative S-adenosyl-L-methionine-dependent methyltransferase domain |
| FRAAL1771 | – | conserved hypothetical protein |
| FRAAL1877 | – | ABC transporter ATP-binding protein |
| FRAAL1921 | – | Putative Fe-S oxidoreductase |
| FRAAL1922 | – | conserved hypothetical protein |
| FRAAL2082 | – | conserved hypothetical protein; putative nucleic acid binding, OB-fold, tRNA/helicase-type domain |
| FRAAL2102 | <i>hemG</i> | protoporphyrinogen oxidase |
| FRAAL2131 | – | Elongation factor G |
| FRAAL2157 | – | conserved hypothetical protein; putative membrane protein |
| FRAAL2168 | <i>crtI</i> | Phytoene desaturase |
| FRAAL2207 | <i>ileS</i> | isoleucyl-tRNA synthetase |
| FRAAL2220 | – | conserved hypothetical protein; putative coiled-coil domain |
| FRAAL2714 | – | conserved hypothetical protein |
| FRAAL2881 | – | conserved hypothetical protein |
| FRAAL2957 | – | conserved hypothetical protein |

| | | |
|------------------|--------------------|--|
| FRAAL3172 | – | conserved hypothetical protein |
| FRAAL3801 | – | hypothetical protein; putative alpha-L-arabinanase domain |
| FRAAL3803 | – | Sulfotransferase |
| FRAAL4035 | – | conserved hypothetical protein |
| FRAAL4488 | <i>asnB</i> | Asparagine synthetase |
| FRAAL4561 | <i>sufE</i> | ferredoxin containing [2Fe-2S] Rieske domain |
| FRAAL4581 | – | Putative rubredoxin, iron-binding domain |
| FRAAL5130 | – | Potassium channel beta chain |
| FRAAL5155 | – | Peptidase S51, dipeptidase E |
| FRAAL5187 | – | putative Pyridoxal phosphate phosphatase |
| FRAAL5240 | <i>BldD</i> | Transcriptional regulator (DNA-binding protein) |
| FRAAL5320 | – | Putative S-adenosyl-L-methionine-dependent methyltransferase |
| FRAAL5653 | <i>sdhC</i> | Succinate dehydrogenase cytochrome B subunit |
| FRAAL5694 | – | Hypothetical protein; putative NAD(P)-binding Rossmann-fold and GCN5-related N-acetyltransferase domains |
| FRAAL5699 | <i>nrdJ</i> | Ribonucleoside-diphosphate reductase M1 chain (Vitamin B12-dependent ribonucleotide reductase) |
| FRAAL5734 | <i>rimO</i> | Ribosomal protein S12 methyltransferase RimO |
| FRAAL5777 | <i>cdsA</i> | Phosphatidate cytidylyltransferase |
| FRAAL5838 | <i>leuA2</i> | 2-isopropylmalate synthase 2 |
| FRAAL5860 | – | conserved hypothetical protein |
| FRAAL5909 | <i>argK</i> | Lysine arginine ornithine transport system kinase |
| FRAAL5921 | <i>meaA</i> | Methylmalonyl-CoA mutase, coenzyme B12-dependent mutase |
| FRAAL5982 | – | Conserved hypothetical protein |
| FRAAL6032 | – | Conserved protein of unknown function |
| FRAAL6080 | – | conserved hypothetical protein; putative Phosphotransacetylase |
| FRAAL6235 | – | Putative pyrophosphatase |
| FRAAL6577 | <i>hppA</i> | Pyrophosphate-energized proton pump (H⁺-PPase) |
| FRAAL6586 | – | conserved hypothetical protein |
| FRAAL6700 | <i>thrC-like</i> | Threonine synthase |
| FRAAL6731 | – | Conserved protein of unknown function |
| FRAAL6737 | – | putative Metallo-beta-lactamase; rhodanese-like domain protein |
| FRAAL6751 | <i>pphA</i> | Serine/threonine specific protein phosphatase. |





| | |
|---|---|
| C | Energy production and conversion |
| D | Cell cycle control, cell division, chromosome partitioning |
| 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 | Translation, ribosomal structure and biogenesis |
| K | Transcription |
| L | Replication, recombination and repair |
| M | Cell wall/membrane/envelope biogenesis |
| O | Posttranslational modification, protein turnover, chaperones |
| P | Inorganic ion transport and metabolism |
| Q | Secondary metabolites biosynthesis, transport and catabolism |
| T | Signal transduction mechanisms |
| U | Intracellular trafficking, secretion, and vesicular transport |
| V | Defense mechanisms |
