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

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

Lorine Bethencourt[1], Hasna Boubakri[1], Najwa Taib[2], Philippe Normand[1], Jean Armengaud[3], Pascale Fournier[1], Céline Brochier-Armanet[2], Aude Herrera-Belaroussi[1*]

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

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

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

Author for correspondence:
Herrera-Belaroussi Aude
UMR 5557 Écologie Microbienne
43, Boulevard du 11 novembre 1918, F-69622 Villeurbanne Cedex, France
Tel: (33) 472 448 200
E-mail: aude.herrera-belaroussi@univ-lyon1.fr

Abstract

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

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

1. Introduction

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

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

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

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

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

2. Material and methods

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

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

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

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

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

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

Proteins were extracted from hyphae as follows: each replicate sample was dissolved in LDS 1X (Invitrogen, Carlsbad, Ca) with 100 µL of LDS 1X per 30 mg of pellet. The samples were then heated at 99°C for 5 min, subjected to sonication in an ultrasonic bath for 5 min, and bead-beated with a Precellys instrument (Bertin Technologies, Montigny-le-Bretonneux, France). Samples were centrifuged for 40 sec at 16,000 g and the resulting supernatants were heated for 10 min at 99°C. Proteins were extracted from spores with a recently optimized protocol consisting in a dilution with milliQ water into 1 mL final volume and bead-beating by means of a Precellys instrument (Bertin technologies) [30]. The sample was then centrifuged, and the resulting supernatant was precipitated with trichloroacetic acid (10% w/vol final). After centrifugation, the pellet was dissolved into 50 µL of LDS1X (Invitrogen), subjected to sonication in an ultrasonic bath for 5 min, and then heated for 10 min at 99°C. Hyphae and spore proteins were subjected to a short SDS-PAGE migration. The whole proteome from a single polyacrylamide band was treated and proteolyzed by trypsin, as previously described [31]. Tryptic peptides were analysed with a Q-Exactive HF high resolution tandem mass spectrometer (ThermoFisher Scientific, Waltham, Ma) incorporating an ultra-high-field orbitrap analyser as previously described [32]. MS/MS spectra were assigned with a protein sequence database derived from *Frankia alni* ACN14a strain genome with standard 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* CcI3, *Frankia*
184 EAN1pec and *F. inefficax* EuI1c). 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* EuI1c, CcI3, QA3, EUN1f, EAN1pec and CN3 strains, the 4 *Actinoplanes*, the

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

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

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

2.7. Proteomics data repository

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

3. Results

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

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

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

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

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

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

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

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

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

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

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

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

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

These 1,037 and 1,641 protein families reflected the core proteome shared between *Frankia* closely related sporangia-forming Actinobacteria proteomes, and between *Frankia* proteomes, respectively. They were compared to a third proteomic database (named “actino_non_spor.db”) gathering the protein sequences of seven Actinobacteria unable to sporulate (Table 1). Sequence similarity-based comparisons allowed identifying 39 and 221 protein families from “actino_sporange.db” and “14_Frankia.db” databases, respectively, without homologues in the “actino_non_spor.db” database and were considered as strong candidates potentially involved in *Frankia* sporulation (supplementary data 3). Among them, 25 and 139 (from “actino_sporange.db” and “14_Frankia.db” databases, respectively) presented functional annotations in ACN14a or CcI3 *Frankia* genomes (not “hypothetical proteins”).

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

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

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

3.4. Additional analyses of Frankia alni *ACN14a gene content using "Gene Phyloprofile" functionality*

In the present study, two alternative methods were performed to identify genes potentially involved in *Frankia* sporulation (*i.e.* well distributed in *Frankia* and other sporulating actinobacterial strains and absent in non-sporulating Actinobacteria). In parallel to the *in-silico* analyses based on proteome comparisons, additional and rapid analyses (using "Gene phyloprofile" functionality of the MicroScope Platform) were also conducted based this time on *Frankia alni* ACN14a gene contents, with respect to the genomes of other sporangia-forming (including *Frankia* strains EuI1c, CcI3, QA3, EUN1f, EAN1pec and CN3, 4 *Actinoplanes* strains, one *Geodermatophilus* and one *Streptosporangium* strains) and non-sporulating Actinobacteria (Table 1).

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

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

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

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

We systematically examined the genomic environment of the candidate genes, looking in particular for syntons or groups of neighbouring genes co-inherited across large evolutionary distances that presumably are involved in a related function. In particular, FRAAL6577 forms a synton comprising among others a putative septum site determining protein (FRAAL6565) and a DNA topoisomerase I (FRAAL6579). The protein encoded by FRAL6577 gene is downregulated when ACN14a *Frankia* strain is in symbiosis with actinorhizal host plant (FC=0.04), while those of the other ones are not modified [29], and downregulated in early steps of contact after 64h (FC=0.28; unpublished), while the other ones are not significantly modified.

3.6. Secondary metabolite gene clusters and Frankia *sporulation*

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

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

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

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

4. Discussion

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

4.1. Hypothetical pattern of Frankia *sporulation early stages*

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

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

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

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

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

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

4.2. Potential role of phosphorelay in Frankia *sporulation*

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

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

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

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

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

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

4.4. Conclusions

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

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

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

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

Acknowledgements

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

References

[1] Locci R, Petrolini Baldan B. On the spore formation process in actinomycetes. I. Morphology and development of *Streptoverticillium* species as examined by 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, Shliaha 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-Quiñónez 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.

[17] Ostash B, Yushchuk O, Tistechok S, Mutenko H, Horbal L, Muryn A, et al. The adpA-like regulatory gene from *Actinoplanes teichomyceticus*: in silico analysis and heterologous expression. World J Microbiol Biotechnol 2015; 31:1297-1301.

[18] Buttner M. Actinoplanes swims into the Molecular Age. J Bacteriol 2017; 25:199.

[19] Mouri Y, Jang M, Konishi K, Hirata A, Tezuka T, Ohnishi Y. Regulation of sporangium formation by the orphan response regulator TcrA in the rare actinomycete *Actinoplanes missouriensis*. Mol Microbiol 2018; 107:718-733.

[20] Normand P, Orso S, Cournoyer B, Jeannin P, Chapelon C, Dawson J, et al. Molecular phylogeny of the genus *Frankia* and related genera and emendation of the family *Frankiaceae*. Int J Syst Bacteriol 1996; 46:1-9.

[21] Nouioui I, Ghodhbane-Gtari F, Montero-Calasanz MD, Goker M, Meier-Kolthoff JP, Schumann P, et al. Proposal of a type strain for *Frankia alni* (Woronin 1866) Von Tubeuf 1895, emended description of *Frankia alni*, and recognition of *Frankia casuarinae* sp. nov. and *Frankia elaeagni* sp. nov. Int J Syst Evol Microbiol 2016; 66:5201-5210.

[22] Pozzi AC, Bautista-Guerrero HH, Nouioui I, Cotin-Galvan L, Pepin R, Fournier P, et al. In-planta sporulation phenotype: a major life history trait to understand the evolution of *Alnus*-infective *Frankia* strains. Environ Microbiol 2015; 17:3125-3138.

[23] Pozzi AC, Bautista-Guerrero HH, Abby SS, Herrera-Belaroussi A, Abrouk D, Normand P, et al. Multi-Locus Sequence Analysis and extensive sampling bring new insights on *Frankia* phylogeny and phylogeography. Syst Appl Microbiol 2018; S0723-2020(18)30113-9.

[24] van Dijk C, Merkus E. A microscopical study of the development of a spore-like stage in the life cycle of the root nodule endophyte of *Alnus glutinosa* (L.) Gaertn. New Phytol 1976; 77:73-91.

[25] Burleigh S, Dawson J. *In vitro* sporulation of *Frankia* strain HFPCcl3 from *Casuarina cunninghamiana*. Can J Microbiol 1991; 37:897-901.

[26] Dewedar A, Mansour S. Infection events in the establishment of *Casuarina-Frankia* symbioses using spore inoculation. Acta Oecolog 1992; 13:379-385.

[27] Normand P, Lapierre P, Tisa LS, Gogarten JP, Alloisio N, Bagnarol E, et al. Genome characteristics of facultatively symbiotic *Frankia* sp. strains reflect host range and host plant biogeography. Genome Res 2007; 17:7-15.

[28] Normand P, Lalonde M. Evaluation of *Frankia* strains isolated from provenances of two *Alnus* species. Can J Microbiol 1982; 28:1133-1142.

[29] Alloisio N, Queiroux C, Fournier P, Pujic P, Normand P, Vallenet D, et al. The *Frankia alni* symbiotic transcriptome. Mol Plant Microbe Interact 2010; 23:593-607.

[30] Mappa C, Pible O, Armengaud J, Alpha-Bazin B. Assessing the ratio of *Bacillus* spores and vegetative cells by shotgun proteomics. Environ Sci Pollut Res Int 2018; 10.1007/s11356-018-3341-z.

[31] Hartmann EM, Allain F, Gaillard JC, Pible O, Armengaud J. Taking the shortcut for high-throughput shotgun proteomic analysis of bacteria. Methods Mol Biol 2014; 1197:275-285.

[32] Ghedira K, Harigua-Souiai E, Benhamda C, Fournier P, Pujic P, Guesmi S, et al. The PEG-responding desiccome of the alder microsymbiont *Frankia alni*. Sci Rep 2018; 8:759.

[33] Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 1997; 25:3389-3402.

[34] Miele V, Penel S, Duret L. Ultra-fast sequence clustering from similarity networks with SiLiX. BMC Bioinformatics 2011; 12:116.

[35] Medema MH, Blin K, Cimermancic P, de Jager V, Zakrzewski P, Fischbach MA, et al. antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. Nucleic Acids Res 2011; 39:W339–W346.

[36] Vallenet D, Calteau A, Cruveiller S, Gachet M, Lajus A, Josso A, et al. MicroScope in 2017: an expanding and evolving integrated resource for community expertise of microbial genomes. Nucleic Acids Res 2017; 45:D517-D528.

[37] Hosoya S, Asai K, Ogasawara N, Takeuchi M, Sato T. Mutation in yaaT leads to significant inhibition of phosphorelay during sporulation in *Bacillus subtilis*. J Bacteriol 2002; 184:5545–5553.

[38] Tschowri N, Schumacher M, Schlimpert S, Chinnam N, Findlay K, Brennan R, et al. Tetrameric c-di-GMP mediates effective transcription factor dimerization to control *Streptomyces* development. Cell 2014; 158:1136-1147.

[39] Torrey J. Endophyte sporulation in root nodules of actinorhizal plants. Physiologia Plantarum 1987; 70:279-288.

[40] Kucho K, Hay A, Normand P. The determinants of the actinorhizal symbiosis. Microb Environ 2010; 25:241-252.

[41] Brana AF, Manzanal MB, Hardisson C. Occurrence of polysaccharide granules in sporulating hyphae of *Streptomyces viridochromogenes*. J Bacteriol 1980; 144:1139-1142.

[42] Ramade N, Vining LC. Accumulation of intracellular carbon reserves in relation to chloramphenicol biosynthesis by *Streptomyces venezuelae*. Can J Microbiol 1993; 39:377–383.

[43] Plaskitt KA, Chatter KF. Influences of developmental genes on localized glycogen deposition in colonies of a mycelial prokaryote, *Streptomyces coelicolor* A3(2): a possible interface between metabolism and morphogenesis. Phil Trans Roy Soc B 1995; 347:105-121.

[44] Xu H, Chater KF, Deng Z, Tao M. A Cellulose Synthase-Like Protein Involved in Hyphal Tip Growth and Morphological Differentiation in *Streptomyces*. J Bacteriol 2008; 190:4971–4978.

[45] Pujic P, Alloisio N, Fournier P, Roche D, Sghaier H, Armengaud J, et al. Omics of the early molecular dialogue between *Frankia alni* and *Alnus glutinosa* and the cellulase synton. Env Microbiol 2019; in press (EMI 14606).

[46] Hempel AM, Cantlay S, Molle V, Wang SB, Naldrett MJ, Parker JL, et al. The Ser/Thr protein kinase AfsK regulates polar growth and hyphal branching in the filamentous bacteria *Streptomyces*. Proc Natl Acad Sci U S A 2012; 109:E2371-2379.

[47] Kleinschnitz EM, Heichlinger A, Schirner K, Winkler J, Latus A, Maldener I, et al. Proteins encoded by the *mre* gene cluster in *Streptomyces coelicolor* A3(2) cooperate in spore wall synthesis. Mol Microbiol 2011; 79:1367-1379.

[48] Chandra G, Chater KF. Developmental biology of *Streptomyces* from the perspective of 100 actinobacterial genome sequences. FEMS Microbiol Rev 2014; 38:345-379.

[49] Tschowri N, Schumacher M, Schlimpert S, Chinnam N, Findlay K, Brennan R, et al. Tetrameric c-di-GMP mediates effective transcription factor dimerization to control *Streptomyces* development. Cell 2014 158:1136-1147.

[50] Mouri Y, Konishi K, Fujita A, Tezuka T, Ohnishi Y. Regulation of sporangium formation by BldD in the rare actinomycete *Actinoplanes missouriensis*. J Bacteriol 2017; 199:e00840-00816.

[51] Cotin-Galvan L. Relation plante-hôte / Frankia dans les symbioses actinorhiziennes : cas particulier des souches non-isolables capables de sporuler in-planta. Lyon1, Lyon. 2014.

[52] Hirono M, Mimura H, Nakanishi Y, Maeshima M. Expression of functional *Streptomyces coelicolor* H+-pyrophosphatase and characterization of its molecular properties. J Biochem 2005; 138:183–191.

[53] Russell JR, Cabeen MT, Wiggins PA, Paulsson J, Losick R. Noise in a phosphorelay 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. EuI1 (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

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

[68] Mansour SR, Oshone R, Hurst SGt, Morris K, Thomas WK, Tisa LS. Draft genome sequence of *Frankia* sp. strain CcI6, a salt-tolerant nitrogen-fixing actinobacterium isolated from the root nodule of *Casuarina cunninghamiana*. Genome Announc 2014; 2:e01205-01213.

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

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

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

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

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

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

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

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

[77] Ivanova N, Sikorski J, Jando M, Munk C, Lapidus A, Glavina Del Rio T, et al. Complete genome sequence of *Geodermatophilus obscurus* type strain (G-20). Stand Genomic Sci 2010; 2:158-167.

[78] Nolan M, Sikorski J, Jando M, Lucas S, Lapidus A, Glavina Del Rio T, et al. Complete genome sequence of *Streptosporangium roseum* type strain (NI 9100). Stand Genomic Sci 2010; 2:29-37.

[79] Yu DS, Jeong H, Lee DH, Kwon SK, Song JY, Kim BK, et al. Complete genome sequence of the probiotic bacterium *Bifidobacterium bifidum* strain BGN4. J Bacteriol 2012; 194:4757-4758.

[80] O'Connell Motherway M, Zomer A, Leahy SC, Reunanen J, Bottacini F, Claesson MJ, et al. Functional genome analysis of *Bifidobacterium breve* UCC2003 reveals type IVb tight adherence (Tad) pili as an essential and conserved host-colonization factor. Proc Natl Acad Sci USA 2011; 108:11217-11222.

[81] Lee JH, Karamychev VN, Kozyavkin SA, Mills D, Pavlov AR, Pavlova NV, et al. Comparative genomic analysis of the gut bacterium *Bifidobacterium longum* reveals loci susceptible to deletion during pure culture growth. BMC Genomics 2008; 9:247.

[82] MP. M, RL. W, WW. H, N. A, M. M, C. F, et al. The complete genome of *Rhodococcus* sp. RHA1 provides insights into a catabolic powerhouse. Proc Natl Acad Sci U S A 2006; 103:15582-15587.

[83] Sekine M, Tanikawa S, Omata S, Saito M, Fujisawa T, Tsukatani N, et al. Sequence analysis of three plasmids harboured in *Rhodococcus erythropolis* strain PR4. Environ Microbiol 2006; 8:334-346.

Figure legends

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

Form the outside in:

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

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

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

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

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

842 **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
***Frankia* strains**					
Frankia alni ACN14a	7497934	6775	6700	NC_008278	[27]
Frankia casuarinae CcI3	5433628	4618	4499	NC_007777	[27]
Frankia EAN1pec	8982042	7377	7191	NC_009921	[64]
Frankia inefficax EuI1c	8815781	7263	7083	NC_014666	[65]
Frankia discariae BCU110501	7887841	6047	5858	NZ_ARDT00000000	[66]
Frankia saprophytica CN3	9966592	7173	7010	NZ_AGJN00000000	[67]
Frankia casuarinae CcI6	5578620	4669	4524	NZ_AYTZ00000000	[68]
Frankia casuarinae CeD	5004459	4246	4112	NZ_JPGU00000000	[69]
Frankia casuarinae Thr	5309830	4560	4439	NZ_JENI00000000	[70]
Frankia casuarinae Allo2	5352110	4605	4465	NZ_JPHT00000000	[71]
Frankia QA3	7578953	6546	6031	NZ_AJWA00000000	[72]*
Frankia EUN1f	9352740	8236	8161	NZ_ADGX00000000	[64]*
Frankia elaeagni BMG5.12	7588936	5473	5320	NZ_ARFH00000000	[21]
Frankia BMG5.23	5352110	4478	4349	NZ_JDWE00000000	[73]
Sporangia-forming Actinobacteria					
Actinoplanes friuliensis DSM_7358	7376071	8648	8565	NC_022657	[74]
Actinoplanes missouriensis 431	8773466	8203	8124	NC_017093	[75]
Actinoplanes N902_109	9228054	8283	8212	NC_021191	Hu et al., unpublished
Actinoplanes SE50_110	9239851	8385	8247	NC_017803	[76]
Geodermatophilum obscurus DSM_43160	5322497	5155	4810	NC_013757	[77]
Streptosporangium roseum DSM_43021	10369518	9508	8975	NC_013595	[78]
Non-sporulating Actinobacteria					
Bifidobacterium bifidum BGN4	2223664	1902	1834	NC_017999	[79]
Bifidobacterium breve UCC2003	2422684	1985	1854	NC_020517	[80]
Bifidobacterium longum DJO10A	2389526	2074	2001	NC_010816	[81]
Rhodococcus jostii RHA1	9702737	9221	9145	NC_008271	[82]
Rhodococcus erythropolis PR4	6895538	6511	6437	NC_012490	[83]
Rhodococcus opacus B4	8830572	8253	8191	NC_012522	Takarada et al., unpublished
Micrococcus luteus NTC_2665	2501097	2345	2236	NC_012803	Lucas et al., unpublished

843 * = Normand, personal observations of sporulation

844 **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**
845 **hyphae)**
846

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	alkyldihydroxyacetonephosphate 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 PaaI	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	glutamyl-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 *Frankia alni* ACN14	Annotation in *Frankia alni* ACN14a genome	Locus_Tag *Frankia casuarinae* CcI3	Annotation in *Frankia casuarinae* CCI3 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 (Cysteinyl-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

853 **Table 4. Common genes identified from both comparative genomic and proteogenomic analyses**
854

Locus_Tag *Frankia alni* ACN14	**Annotation in *Frankia alni* ACN14a genome**	**Proteomic FC spore/hyphae**	***p-value***	**Group code in the comparative genomic analysis**
3 up-regulated proteins in spores:				
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
3 down-regulated proteins in spores:				
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

855
856

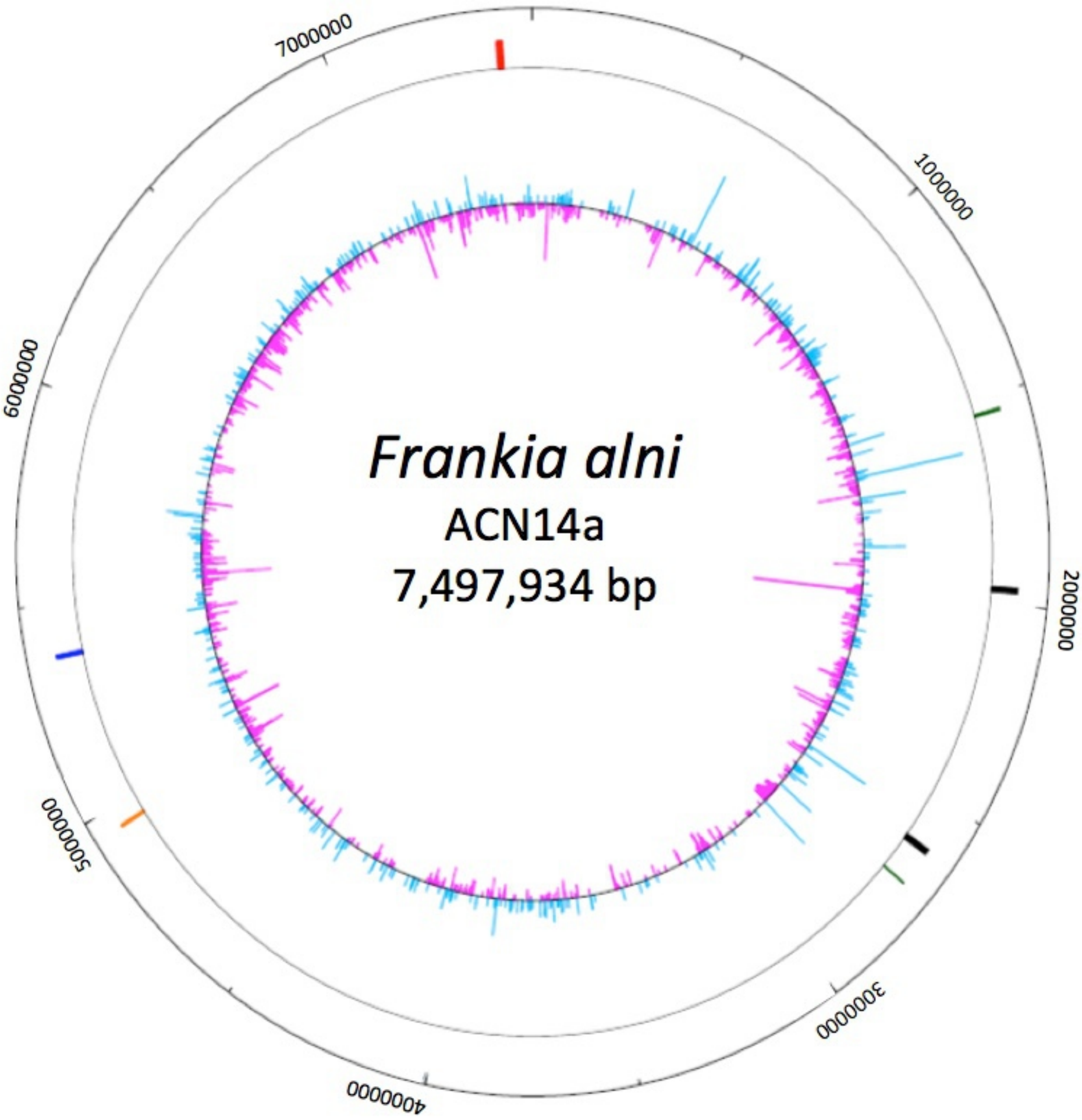
Table 5. Analyses of *Frankia alni* ACN14a gene content using "Gene Phyloprofile" functionality.

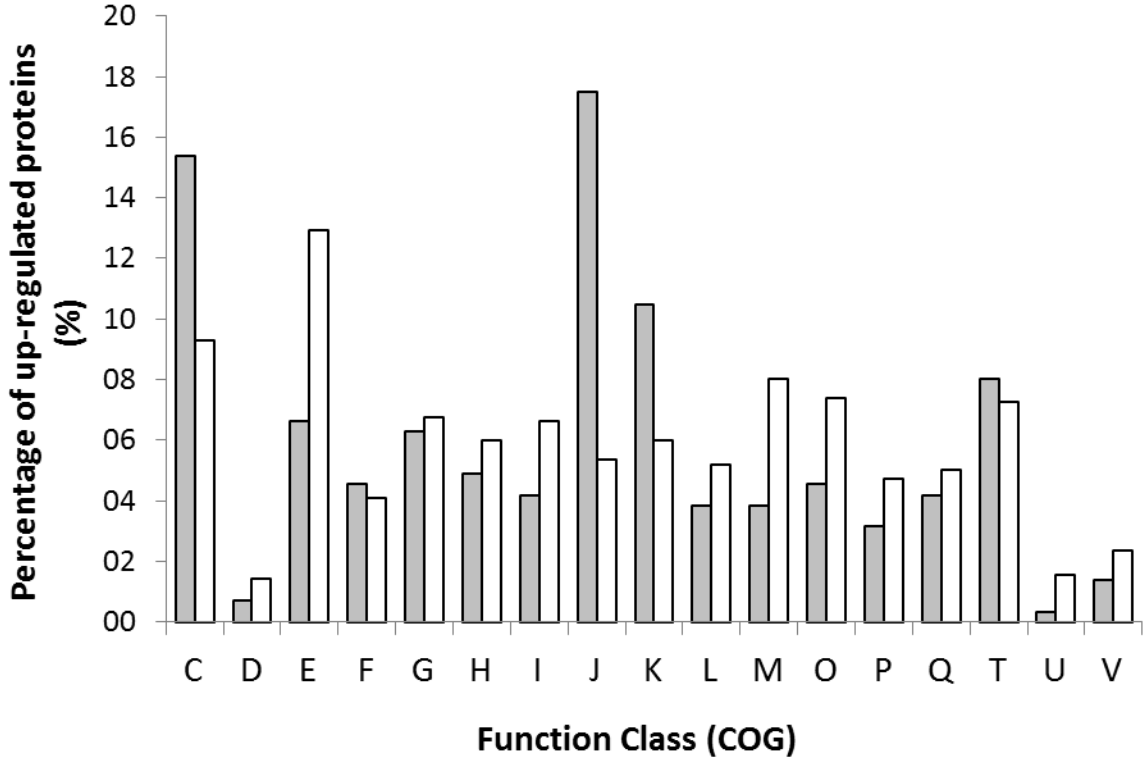
Frankia alni ACN14a genes with (i) homologs in the 12 genomes of other sporangia-forming actinobacterial (including *Frankia* strains EuI1c, CcI3, QA3, EUN1f, EAN1pec and CN3, 4 *Actinoplanes* strains, one *Geodermatophilus* and one *Streptosporangium* strains, given in Table 1) and (ii) without homologs 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	*pdxH*	pyridoxine 5'-phosphate oxidase
FRAAL0172	*metH*	B12-dependent homocysteine-N5-methyltetrahydrofolate transmethylase
FRAAL0399	_	conserved hypothetical protein; putative methyltransferase domain
FRAAL0617	_	conserved hypothetical protein
FRAAL0996	*fabH*	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	*hemG*	protoporphyrinogen oxidase
FRAAL2131	_	Elongation factor G
FRAAL2157	_	conserved hypothetical protein; putative membrane protein
FRAAL2168	*crtI*	Phytoene desaturase
FRAAL2207	*ileS*	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	*asnB*	Asparagine synthetase
FRAAL4561	*sufE*	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	*BldD*	Transcriptional regulator (DNA-binding protein)
FRAAL5320	_	Putative S-adenosyl-L-methionine-dependent methyltransferase
FRAAL5653	*sdhC*	Succinate dehydrogenase cytochrome B subunit
FRAAL5694	_	Hypothetical protein; putative NAD(P)-binding Rossmann-fold and GCN5-related N-acetyltransferase domains
FRAAL5699	*nrdJ*	Ribonucleoside-diphosphate reductase M1 chain (Vitamin B12-dependent ribonucleotide reductase)
FRAAL5734	*rimO*	Ribosomal protein S12 methylthiotransferase RimO
FRAAL5777	*cdsA*	Phosphatidate cytidylyltransferase
FRAAL5838	*leuA2*	2-isopropylmalate synthase 2
FRAAL5860	_	conserved hypothetical protein
FRAAL5909	*argK*	Lysine arginine ornithine transport system kinase
FRAAL5921	*meaA*	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	***hppA***	**Pyrophosphate-energized proton pump (H+-PPase)**
FRAAL6586	_	conserved hypothetical protein
FRAAL6700	*thrC-like*	Threonine synthase
FRAAL6731	_	Conserved protein of unknown function
FRAAL6737	_	putative Metallo-beta-lactamase; rhodanese-like domain protein
FRAAL6751	*pphA*	Serine/threonine specific protein phosphatase.

863


7000000
1000000
2000000
3000000
4000000
5000000
6000000
Frankia alni
ACN14a
7,497,934 bp



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