



## OPEN ACCESS

## EDITED BY

Masaaki Konishi,  
Kitami Institute of Technology, Japan

## REVIEWED BY

María Mercedes Zambrano,  
Corporación Corpogen, Colombia  
Ken-ichi Kucho,  
Kagoshima University, Japan

## \*CORRESPONDENCE

Maier Gtari  
✉ maier.gtari@insat.rnu.tn

RECEIVED 08 January 2024

ACCEPTED 04 July 2024

PUBLISHED 31 July 2024

## CITATION

Gtari M, Maaoui R, Ghodhbane-Gtari F,  
Ben Slama K and Sbissi I (2024) MAGs-centric  
crack: how long will, spore-positive *Frankia*  
and most *Protofrankia*, microsymbionts  
remain recalcitrant to axenic growth?  
*Front. Microbiol.* 15:1367490.  
doi: 10.3389/fmicb.2024.1367490

## COPYRIGHT

© 2024 Gtari, Maaoui, Ghodhbane-Gtari,  
Ben Slama and Sbissi. This is an open-access  
article distributed under the terms of the  
[Creative Commons Attribution License  
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction  
in other forums is permitted, provided the  
original author(s) and the copyright owner(s)  
are credited and that the original publication  
in this journal is cited, in accordance with  
accepted academic practice. No use,  
distribution or reproduction is permitted  
which does not comply with these terms.

# MAGs-centric crack: how long will, spore-positive *Frankia* and most *Protofrankia*, microsymbionts remain recalcitrant to axenic growth?

Maier Gtari<sup>1\*</sup>, Radhi Maaoui<sup>1</sup>, Faten Ghodhbane-Gtari<sup>1,2</sup>,  
Karim Ben Slama<sup>3</sup> and Imed Sbissi<sup>4</sup>

<sup>1</sup>Department of Biological and Chemical Engineering, USCR Molecular Bacteriology and Genomics, National Institute of Applied Sciences and Technology, University of Carthage, Tunis, Tunisia, <sup>2</sup>Higher Institute of Biotechnology Sidi Thabet, University of La Manouba, Tunisia, <sup>3</sup>LR Bioresources, Environment, and Biotechnology (LR22ES04), Higher Institute of Applied Biological Sciences of Tunis, University of Tunis El Manar, Tunis, Tunisia, <sup>4</sup>LR Pastoral Ecology, Arid Regions Institute, University of Gabes, Medenine, Tunisia

Nearly 50 years after the ground-breaking isolation of the primary *Comptonia peregrina* microsymbiont under axenic conditions, efforts to isolate a substantial number of *Protofrankia* and *Frankia* strains continue with enduring challenges and complexities. This study aimed to streamline genomic insights through comparative and predictive tools to extract traits crucial for isolating specific *Frankia* in axenic conditions. Pangenome analysis unveiled significant genetic diversity, suggesting untapped potential for cultivation strategies. Shared metabolic strategies in cellular components, central metabolic pathways, and resource acquisition traits offered promising avenues for cultivation. Ecological trait extraction indicated that most uncultured strains exhibit no apparent barriers to axenic growth. Despite ongoing challenges, potential caveats, and errors that could bias predictive analyses, this study provides a nuanced perspective. It highlights potential breakthroughs and guides refined cultivation strategies for these yet-uncultured strains. We advocate for tailored media formulations enriched with simple carbon sources in aerobic environments, with atmospheric nitrogen optionally sufficient to minimize contamination risks. Temperature adjustments should align with strain preferences—28–29°C for *Frankia* and 32–35°C for *Protofrankia*—while maintaining an alkaline pH. Given potential extended incubation periods (predicted doubling times ranging from 3.26 to 9.60 days, possibly up to 21.98 days), patience and rigorous contamination monitoring are crucial for optimizing cultivation conditions.

## KEYWORDS

*Frankia*, *Protofrankia*, obligate/facultative microsymbionts, metagenome-assembled genomes, ecological traits, axenic conditions, phenotype predictive tools, trait-based model

## Introduction

Actinobacterial taxa within the four genera - *Frankia*, *Protofrankia*, *Parafrankia*, and *Pseudofrankia*, of the family *Frankiaceae* (Gtari, 2022), colonize the nitrogen-fixing root nodules in actinorhizal plants, mainly engaging in mutualistic symbiosis (Huss-Danell, 1997). Initially classified as obligate microsymbionts with elusive identities (Becking, 1970), efforts

to isolate and characterize these microsymbionts faced challenges such as slow growth rates, undefined media, and contamination issues (Wheeler et al., 2008). A pivotal breakthrough in 1978 led to the first axenic cultivation of a *Frankia* strain from *Comptonia peregrina* nodules (Callaham et al., 1978). Despite ongoing successful isolations, there has been limited success in culturing *Protofrankia* strains (Persson et al., 2011, 2015; Gtari et al., 2015; Nguyen et al., 2016, 2019; Gueddou et al., 2019; Berckx et al., 2022). Moreover, *Frankia* strains exhibiting the Sp+ (spore-positive phenotype), which can be distinguished from Sp- (spore-negative phenotype) types that produce fewer or no multilocular sporangia in planta (Schwintzer, 1990; Schwob et al., 2018), have not yet been cultured despite numerous attempts (Normand et al., 2017; Herrera-Belaroussi et al., 2020; Pozzi et al., 2020). These yet-uncultured *Frankia* microsymbionts have been extensively studied using various phylogenetic markers, amplified by PCR directly from DNA extracted from the root nodules of their host species. Key markers include 16S rRNA genes (Nazaret et al., 1991; Simonet et al., 1994), ITS (Internal Transcribed Spacer) rRNA gene regions (Ghodhbane-Gtari et al., 2010), the *glnA* gene (glutamine synthetase), the *dnaA* gene (chromosome replication initiator), and the *nifDK* IGS (intergenic spacer between *nifD* and *nifK* genes) (Nouioui et al., 2014). More recently, Metagenome-Assembled Genomes (MAGs) have been employed to study these microsymbionts, uncovering the genetic diversity and functional capabilities of these elusive organisms. Additionally, four candidate species have been defined to accommodate uncultured taxa: *Candidatus* *Protofrankia* *daticae* (Persson et al., 2011) and *Candidatus* *Protofrankia* *californiensis* (Normand et al., 2017), as well as *Candidatus* *Frankia* *alpina* (Pozzi et al., 2020) and *Candidatus* *Frankia* *nodulisporulans* (Herrera-Belaroussi et al., 2020).

Leveraging data from MAGs and single-cell amplified genomes (SAGs), efforts are underway to challenge the perception of certain bacteria as “uncultivable” (Gutleben et al., 2018; Lewis and Ettema, 2019; Lewis et al., 2021; Xie et al., 2021; Liu et al., 2022; Laugier, 2023). Through exploration of diverse culture techniques such as co-cultivation strategies, microfluidics, and synthetic biology, the goal is to establish culturomics as a pivotal complement to metagenomics and single-cell genomics, enhancing our understanding of historically challenging-to-culture microbial communities (Lagier et al., 2012; Nowrotek et al., 2019).

If some of *Frankia* strains are genuinely uncultivable, their exclusive symbiotic lifestyle, distinct from their cultured counterparts, is expected to prompt a relaxation of selection on various metabolic functions that become obsolete within the stress-buffered host cells where the microsymbiont is shielded by host stress responses (Wernegreen, 2015; Lajoie and Parfrey, 2022). Consequently, this could lead to genome streamlining, characterized by a gradual loss of functions from the microsymbiont genome that are also present in the host genome (Moran et al., 2008; Lo et al., 2016). To distill traits pertinent to cultivability, genomic information is organized into a metabolomic blueprint that encapsulates ecological traits (Lajoie and Kembel, 2019; Gtari et al., 2024). This integrative approach, permitted to unveil metabolic profiles, exposing unforeseen requirements and guiding the design of tailored conditions for the growth and isolation of previously uncultured microorganisms (Giovannoni et al., 2014; Lewis et al., 2021). Cultivation strategies span specific growth conditions to innovative methods bridging natural habitats and laboratory settings, including environmental condition replication,

dormancy release, co-cultivation, and specialized *in situ* cultivation devices (Riva et al., 2022; Kapinusova et al., 2023; Schultz et al., 2023; Yan et al., 2023). Advanced techniques such as genome editing and adaptive evolution further contribute to narrowing the gap between a microorganism's native environment and controlled laboratory conditions (Liu et al., 2022; Laugier, 2023). A notable example for *Frankia* research involves the achievement of employing a dual approach, integrating comparative genomics with physiological assays on nodule tissues. This methodology facilitated the axenic cultivation of a previously elusive *Protofrankia* microsymbiont associated with *Coriaria* spp. (Gtari et al., 2015).

The main objective of this study is to delve into the ecological and evolutionary imprints of uncultured *Frankia* genomes. Leveraging comparative genomics and advanced genome predictive tools, we seek to unravel the intricate details of their genome content, organization, and predictive functions that could be instrumental in unlocking their growth potential in axenic cultures. Despite persistent challenges, our findings offer a nuanced perspective, laying the groundwork for potential breakthroughs and steering the development of cultivation strategies.

## Materials and methods

### Genome selection

For this study, a comprehensive set of 30 genomes was obtained from the NCBI database, encompassing diverse strains from the four *Frankia* genera, including both uncultured microsymbionts from *Protofrankia* and *Frankia* genera and their closely related cultivated counterparts (Table 1). As per the Minimum Information about a MAG (MIMAG) standards (Bowers et al., 2017), MAGs are classified as ‘high-quality’ if they exhibit >90% completeness and less than 5% contamination. In this study, the completeness of the used MAGs ranges from 82.45 to 97.39%, and contamination levels range from 0.25 to 5.2%. These values indicate that the MAGs meet the criteria necessary for conducting comparative analyses with isolate-assembled genomes (IAGs), thereby substantiating rigorous downstream functional analysis and ensuring the validity of potential conclusions.

### Genome comparison

Pairwise Mash distances were calculated using Mash v2.1 (Ondov et al., 2016). Digital DNA–DNA hybridization (dDDH) values and confidence intervals were calculated using GGDC version 3.0 (Meier-Kolthoff et al., 2013). Pairwise average nucleotide identity (ANI) were calculated at <https://www.ezbiocloud.net> (Yoon et al., 2017).

### Genomic evolutionary signatures

For the presence of insertion sequences, ISfinder (Siguiet et al., 2006, 2014) was employed. Pseudogene prediction was conducted utilizing Pseudofinder (Syberg-Olsen et al., 2022), and CRISPR elements were identified using the CRISPR Recognition Tool (Bland et al., 2007). Data concerning plasmid-like and virus-like components within a genome were sourced from IMG-M (Chen et al., 2023),

incorporating geNomad for the automated identification of both virus-like and plasmid-like entities. Putative horizontal gene transfer (HGT) events were identified using data from IMG-M (Chen et al., 2023).

## Comparative analysis of cultured and uncultured strains

Pangenome analysis which permitted to categorize genes into core genes include strict-core (>99 to 100% prevalence) and soft-core genes (>95 to >96% prevalence), and cloud genes, was performed using Panaroo v1.3.4 (Tonkin-Hill et al., 2020) with default settings (sequence identity threshold 95%, protein family sequence identity threshold 70%, length difference threshold 95%). Gene gain and loss events were assessed using Panstripe (Tonkin-Hill et al., 2023) alongside the phylogenetic tree from the core gene alignment constructed using IQ-TREE (Nguyen et al., 2015), along with the Panaroo gene presence-absence matrix.

To investigate the evolution of gene content, ancestral gene numbers and events were estimated using the Dollo parsimony method implemented in COUNT v9.1106 program, which does not require an out-group species for the phylogenetic tree (Csurös, 2010). The final tree was generated using the web application chiplot (Xie et al., 2023).

To explore genomic differences in uncultured and cultured strains, a detailed comparative analysis was conducted using Orthovenn3 using the OrthoFinder algorithm, with an  $e$ -value of  $1e^{-2}$  and an inflation value of 1.50, was employed for orthologous analyses and GO term enrichment analysis was also performed using Orthovenn3 (Sun et al., 2023). Pairwise non-synonymous and synonymous substitution ratios (dN/dS) for each single-copy orthologous gene (OG) were estimated with the Codeml module of the PAML v4.10.0 package using the following settings (model=0 and codon frequency=2) (Yang, 2007).

Functional annotations of uncultured *Frankia* and *Protofrankia*, along with their closely related cultured counterparts, were carried out utilizing the Clusters of Orthologous Genes (COG) database (Galperin et al., 2019). The completeness of metabolic pathways was assessed using the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (Kanehisa et al., 2002; Touchon and Roche, 2007). The KEGG metabolic pathways were examined through BlastKOALA (Kanehisa et al., 2016), aligning against the prokaryotes database to elucidate pathway functionalities and completeness.

Cell Wall and Capsule, Dormancy and Sporulation, and Regulation and Cell signaling, were obtained using SEED server for Rapid Annotation Subsystem Technology server (RAST) (Overbeek et al., 2014). The heatmaps were created with ComplexHeatMap package (Gu, 2022) on RStudio.

## Ecological traits extraction

Genomes of uncultured and cultured strains were submitted to the PhenDB web server,<sup>1</sup> using recommended default settings with a balanced accuracy cutoff of 0.75 and a prediction confidence cutoff of

0.6 (Feldbauer et al., 2015) were employed to predict obligate or facultative intracellular lifestyles. GapMind was used to detect gene pathways for essential carbon sources (Price et al., 2021) and amino acid biosynthesis (Price et al., 2020). Functional traits extraction was mainly performed using MicroTrait, an R package streamlining the extraction of fitness traits from microbial genome sequences (Karaoz and Brodie, 2022). This tool employs profile hidden Markov models (profile-HMM) and logical operations to predict and map the protein family content within genome sequences to fitness traits. The microTrait framework depends on meticulously curated HMMs, known as microtrait-hmms, which capture the sequence diversity of protein families sourced from IMG/M. These HMMs undergo benchmarking using KEGG orthologs for trusted cutoff (TC) scores. Software dependencies include HMMER (Potter et al., 2018) and Prodigal (Hyatt et al., 2010), with additional features utilizing Infernal (Nawrocki and Eddy, 2013), tRNAscan-SE (Schattner et al., 2005) and bedtools (Quinlan, 2014). Essential data dependencies encompass microtrait-hmm (gene level profile-HMM database) and dbCAN (Huang et al., 2018)-HMMdb (domain level profile-HMM database for Carbohydrate-active enzymes). Additionally, secondary metabolite clusters were determined using AntiSMASH 7.0 (Blin et al., 2023).

## Results and discussion

Cultivating microsymbionts in axenic conditions has historically posed challenges due to specific growth requirements, nutrient dependencies, and slow growth rates (Alain and Querellou, 2009; Stewart, 2012; Pham and Kim, 2016). This study thoroughly examines previously labeled uncultured *Frankia* strains, representing a significant proportion of *Protofrankia* strains and sporulating *Frankia* microsymbionts (Sp+ types) (Persson et al., 2011, 2015; Nguyen et al., 2016, 2019; Normand et al., 2017; Herrera-Belaroussi et al., 2020; Pozzi et al., 2020; Berckx et al., 2022), to determine if they exhibit barriers to isolation in axenic conditions.

## Phylogeny and taxonomy of uncultured *Frankia*

The 560 single-copy gene phylogenomic trees (Figure 1A) revealed a robust phylogenetic radiation within the *Frankiaceae* family supported by robust bootstraps and posterior probability values. This analysis clearly delineated the four genera—*Pseudofrankia*, *Protofrankia*, *Parafrankia*, and *Frankia* (Gtari, 2022). Mash-based analyses, utilizing MinHash sketches and Jaccard indices for shared k-mers, enabled cohesive clustering of the 30 genomes (Supplementary Figure S1), ensuring a robust alignment of type strain species classifications within their respective genera. Significantly, the most elevated similarities observed for *Frankia umida* Ag45\_Mut15<sup>T</sup> and *Candidatus* *Frankia* nodulisporulans members were 97.15% with AgTrS, 96.80% with AgUmASH1, and 96.82% with AgUmAS1 (Supplementary Figure S1). Digital DNA:DNA hybridization (dDDH) between *Candidatus* *Frankia* nodulisporulans members and Ag45\_Mut15<sup>T</sup> yielded 73.7% ([70.7–76.5]), designating AgTrS, AgUmASH1, and AgUmAS1 as members of *Frankia umida* species (Chun et al., 2018). *Candidatus* *Protofrankia* datisciae Dg1 and *Protofrankia* symbiont of *Coriaria nepalensis* Dg1\_Cn\_nod exhibited genomic

<sup>1</sup> <http://phenadb.org/>

TABLE 1 Genome Assembly and Quality Metrics for Bacterial strains used in the present study.

| Organism name                                       | Genome Accession | Axenic culture status | Genome assembled origin     | Assembly level | Contig N50 | Scaffold N50 | Assembly sequencing tech | CheckM marker set  | CheckM completeness | CheckM contamination |
|---|------------------|-----------------------|-----------------------------|----------------|------------|--------------|--------------------------|--------------------|---------------------|----------------------|
| <i>Frankia casuarinae</i> Cc13 <sup>T</sup>         | CP000249         | +                     | Isolate-assembled Genome    | 1 Scaffold     | 5,433,628  | 5,433,628    |                          | <i>Frankia</i>     | 99.99               | 1.52                 |
| <i>Frankia alni</i> ACN14a <sup>T</sup>             | CT5732132        | +                     | Isolate-assembled Genome    | 1 Scaffold     | 7,497,934  | 7,497,934    |                          | <i>Frankia</i>     | 99.96               | 0.56                 |
| <i>Frankia torreyi</i> Cp11 <sup>T</sup>            | JYFN00000000     | +                     | Isolate-assembled Genome    | 153 Scaffold   | 99,379     | 107,928      | Illumina                 | <i>Frankia</i>     | 99.57               | 1.9                  |
| <i>Frankia canadensis</i> ARgP5 <sup>T</sup>        | FZMO00000000     | +                     | Isolate-assembled Genome    | 568 Contig     | 27,216     |              |                          | <i>Frankia</i>     | 98.87               | 2.33                 |
| <i>Frankia umida</i> Ag45/Mut15 <sup>T</sup>        | JALKFT0000000000 | +                     | Isolate-assembled Genome    | 157 Scaffold   | 112,473    | 188,571      | Illumina                 | <i>Frankia</i>     | 91.62               | 0.5                  |
| <i>Frankia gtarii</i> Agncl-4 <sup>T</sup>          | JANEZS0000000000 | +                     | Isolate-assembled Genome    | 442 Contig     | 38,429     |              | Illumina                 | <i>Frankia</i>     | 98.86               | 1.47                 |
| <i>Frankia</i> sp. QA3                              | AJWA0000000000   | +                     | Isolate-assembled Genome    | 1 Scaffold     | 130,116    | 7,590,853    | Illumina                 | <i>Frankia</i>     | 94.69               | 1.36                 |
| <i>Candidatus</i> Frankia nodulisporulans AgUmASt1  | CADDZU0000000000 | –                     | Metagenome-assembled genome | 211 Scaffold   | 21,949     | 30,049       | Illumina                 | <i>Frankia</i>     | 82.45               | 0.86                 |
| <i>Candidatus</i> Frankia alpina AvVan              | SSXH0000000000   | –                     | Metagenome-assembled genome | 1,228 Contig   | 6,553      |              | Illumina                 | <i>Frankia</i>     | 86.66               | 0.34                 |
| <i>Candidatus</i> Frankia alpina AiOr               | CADCWT0000000000 | –                     | Metagenome-assembled genome | 669 contig     | 17,359     |              | Illumina                 | <i>Frankia</i>     | 87.87               | 0.38                 |
| <i>Candidatus</i> Frankia nodulisporulans AgTrS     | CADCWS0000000000 | –                     | Metagenome-assembled genome | 612 Contig     | 15,284     |              | Illumina                 | <i>Frankia</i>     | 87.26               | 0.29                 |
| <i>Candidatus</i> Frankia nodulisporulans AgUmASH1  | CADDZW0000000000 | –                     | Metagenome-assembled genome | 231 Scaffold   | 17,387     | 26,450       | Illumina                 | <i>Frankia</i>     |                     |                      |
| <i>Parafrankia elaeagni</i> BMG5.12 <sup>T</sup>    | ARFH0000000000   | +                     | Isolate-assembled genome    | 135 Scaffold   | 162,237    | 162,237      | Illumina                 | <i>Parafrankia</i> | 99.59               | 0.19                 |
| <i>Parafrankia irregularis</i> G2 <sup>T</sup>      | FAOZ0000000000   | +                     | Isolate-assembled genome    | 83 Scaffold    | 189,407    | 227,129      |                          | <i>Parafrankia</i> | 99.64               | 1.03                 |
| <i>Parafrankia discariae</i> BCU110501 <sup>T</sup> | ARDT0000000000   | +                     | Isolate-assembled genome    | 194 Scaffold   | 127,450    | 132,179      | Illumina                 | <i>Parafrankia</i> | 99.48               | 0.97                 |

(Continued)

TABLE 1 (Continued)

| Organism name  | Genome Accession | Axenic culture status | Genome assembled origin     | Assembly level | Contig N50 | Scaffold N50 | Assembly sequencing tech | CheckM marker set    | CheckM completeness | CheckM contamination |
|--|------------------|-----------------------|-----------------------------|----------------|------------|--------------|--------------------------|----------------------|---------------------|----------------------|
| <i>Parafrankia colletiae</i> Cc1 17 <sup>T</sup>                       | MBLM00000000     | +                     | Isolate-assembled genome    | 195 Contig     | 118,488    |              | Illumina                 | <i>Parafrankia</i>   | 96.86               | 2.51                 |
| <i>Parafrankia soli</i> Cj <sup>T</sup>                                | MAXA00000000     | +                     | Isolate-assembled genome    | 289 Contig     | 88,066     |              | Illumina                 | <i>Parafrankia</i>   | 95.59               | 0.89                 |
| <i>Protofrankia coriariae</i> BMG5.1 <sup>T</sup>                      | JWIO00000000     | +                     | Isolate-assembled genome    | 116 Scaffold   | 41,204     | 105,614      | Illumina                 | <i>Protofrankia</i>  | 88.17               | 0.76                 |
| <i>Protofrankia</i> sp. BMG5.30  | MOME00000000     | +                     | Isolate assembled genome    | 94 Contig      | 124,729    |              | Illumina                 | <i>Protofrankia</i>  | 98.52               | 0.44                 |
| <i>Candidatus</i> Protofrankia datisciae Dg1                           | CP002801         | –                     | Metagenome-assembled genome | 3 Scaffold     | 5,323,186  | 5,323,186    | Illumina                 | <i>Protofrankia</i>  | 95.83               | 0.25                 |
| <i>Protofrankia</i> symbiont of <i>Coriaria ruscifolia</i> Cv1_Ct_nod1 | CAAAFR00000000   | –                     | Metagenome-assembled genome | 203 Scaffold   | 32,417     | 57,864       | Illumina                 | <i>Protofrankia</i>  | 97.39               | 0.9                  |
| <i>Candidatus</i> Protofrankia californiensis Dg2                      | FLUV00000000     | –                     | Metagenome-assembled genome | 2,738 Contig   | 3,103      |              |                          | <i>Protofrankia</i>  | 84.78               | 5.2                  |
| <i>Protofrankia</i> symbiont of <i>Coriaria nepalensis</i> Dg1_Cn_nod  | CAAAHA00000000   | –                     | Metagenome-assembled genome | 1,126 Contig   | 8,186      |              | Illumina                 | <i>Protofrankia</i>  | 84.78               | 5.2                  |
| <i>Candidatus</i> Protofrankia meridionalis Cppng1_Ca_nod              | CAAAFQ00000000   | –                     | Metagenome-assembled genome | 101 Scaffold   | 7,881      | 87,705       | Illumina                 | <i>Protofrankia</i>  |                     |                      |
| <i>Candidatus</i> Protofrankia meridionalis Cppng1                     | CADDZT00000000   | –                     | Metagenome-assembled genome | 101 Scaffold   | 7,881      | 87,705       | Illumina                 | <i>Protofrankia</i>  |                     |                      |
| <i>Pseudofrankia inefficax</i> Eu11c <sup>T</sup>                      | CP002299         | +                     | Isolate-assembled genome    | 1 Scaffold     | 8,815,781  | 8,815,781    | Illumina                 | <i>Pseudofrankia</i> | 100                 | 1.36                 |
| <i>Pseudofrankia saprophytica</i> CN3 <sup>T</sup>                     | AGJN00000000     | +                     | Isolate-assembled genome    | 2 Scaffold     | 221,503    | 8,190,446    | Illumina                 | <i>Pseudofrankia</i> | 99.95               | 1.83                 |

(Continued)

TABLE 1 (Continued)

| Organism name  | Genome Accession | Axenic culture status | Genome assembled origin  | Assembly level | Contig N50 | Scaffold N50 | Assembly sequencing tech | CheckM marker set    | CheckM completeness | CheckM contamination |
|--|------------------|-----------------------|--------------------------|----------------|------------|--------------|--------------------------|----------------------|---------------------|----------------------|
| <i>Pseudofrankia asymbiotica</i> M16386 <sup>T</sup> | MOMC000000000    | +                     | Isolate-assembled genome | 174 Contig     | 117,461    |              | Illumina                 | <i>Pseudofrankia</i> | 99.85               | 1.99                 |
| <i>Pseudofrankia</i> sp. DC12                        | LANG000000000    | +                     | Isolate-assembled genome | 1 Scaffold     | 837,743    | 6,884,336    | Illumina                 | <i>Pseudofrankia</i> | 99.41               | 1.27                 |
| <i>Pseudofrankia</i> sp. BMG5.36                     | MBLO000000000    | +                     | Isolate-assembled genome | 280 Scaffold   | 84,949     |              | Illumina                 | <i>Pseudofrankia</i> | 95.39               | 2.75                 |

\* The symbols (+) and (-) indicate the cultivability in axenic conditions; (+) denotes strains isolated in axenic conditions, while (-) indicates strains that have not yet been cultured despite several attempts (Persson et al., 2011, 2015; Nguyen et al., 2016, 2019; Normand et al., 2017; Herrera-Belaroussi et al., 2020; Pozzi et al., 2020; Berckx et al., 2022).

similarities of 96.72 and 96.65%, respectively, with *Protofrankia coriariae* BMG5.1<sup>T</sup> (Supplementary Figure S2). dDDH values ranged from 66.5% ([63.5–69.3]) to 67.5% ([64.5–70.3]), classifying them as a borderline taxon. *Candidatus* *Frankia alpina* strains AiOr and AvVan displayed values of 95.58 and 95.86%, respectively, with *Frankia gtarii* strain Agnc1-4<sup>T</sup>. However, dDDH values of 56.90% ([54.1–59.6%]) for AiOr and 4.50% ([51.8–57.2%]) for AvVan suggest genomic distinctions.

This observation aligns with findings in microbial ecology, where the uncultivability of certain taxa is not solely attributed to evolutionary divergence but may involve a complex interplay of ecological adaptations and genomic features (Stewart, 2012). It challenges the conventional notion that uncultured microbes represent entirely novel lineages, emphasizing the need for a nuanced understanding of microbial diversity and ecology (Connon and Giovannoni, 2002). The potential genomic continuity hints at shared ancestry, possibly driven by environmental factors influencing microbial evolution (Shade et al., 2014).

## Genome structure and evolution of uncultured *Frankia*

As stated previously, an obligate intracellular lifestyle may lead to genome reduction, making microorganisms uncultivable in the lab (Casadevall, 2008; Rodríguez-Gijón et al., 2022). Genomic comparisons between cultured and uncultured *Frankia* strains revealed consistent genome size differences (0.924–3.427 Mb) (Figure 1A), a trend also observed among cultivated strains associated with host plant speciation (Normand et al., 2007; Tisa et al., 2016). Insertion sequences, horizontal gene transfer, viral interactions, plasmid dynamics, and CRISPR systems collectively drive genome evolution (Casadevall, 2008; Sloan and Moran, 2013; Brito, 2021; Arnold et al., 2022). These elements play a role in rearranging genomes, introducing novel genetic material, and the CRISPR systems serve as a defense mechanism against integration (Nussenzweig and Marraffini, 2020). Diverse adaptive strategies were evident in gene density (ratio of the number of genes per number of base pairs) variations, with some uncultured *Frankia* strains exhibiting high gene densities (Figure 1A), possibly indicating specialized adaptations (Nakabachi et al., 2006; Kuo et al., 2009; Kuo and Ochman, 2010). Pseudogenes identified in the entire set of 30 genomes predominantly displayed features such as incompleteness, frameshifts, multiple issues, internal stop codons, with fewer instances of ambiguous residues (Supplementary Table S1). The heightened presence of pseudogenes in Dg2 implies continuous evolutionary dynamics, likely influenced by specific environmental pressures. Furthermore, the KEGG distribution of pseudogenes primarily associated them with “carbohydrate metabolism,” followed by “Genetic information processing,” and “Unclassified genetic information processing” (Supplementary Figure S3). The examination of gene events in the 30 analyzed *Frankiaceae* genomes unveiled a predominant sequence of occurrences, primarily involving loss, followed by duplication, and subsequently, gain (Figure 1A). Notably, in the case of uncultured *Frankia*, the total number of gene gains and losses was

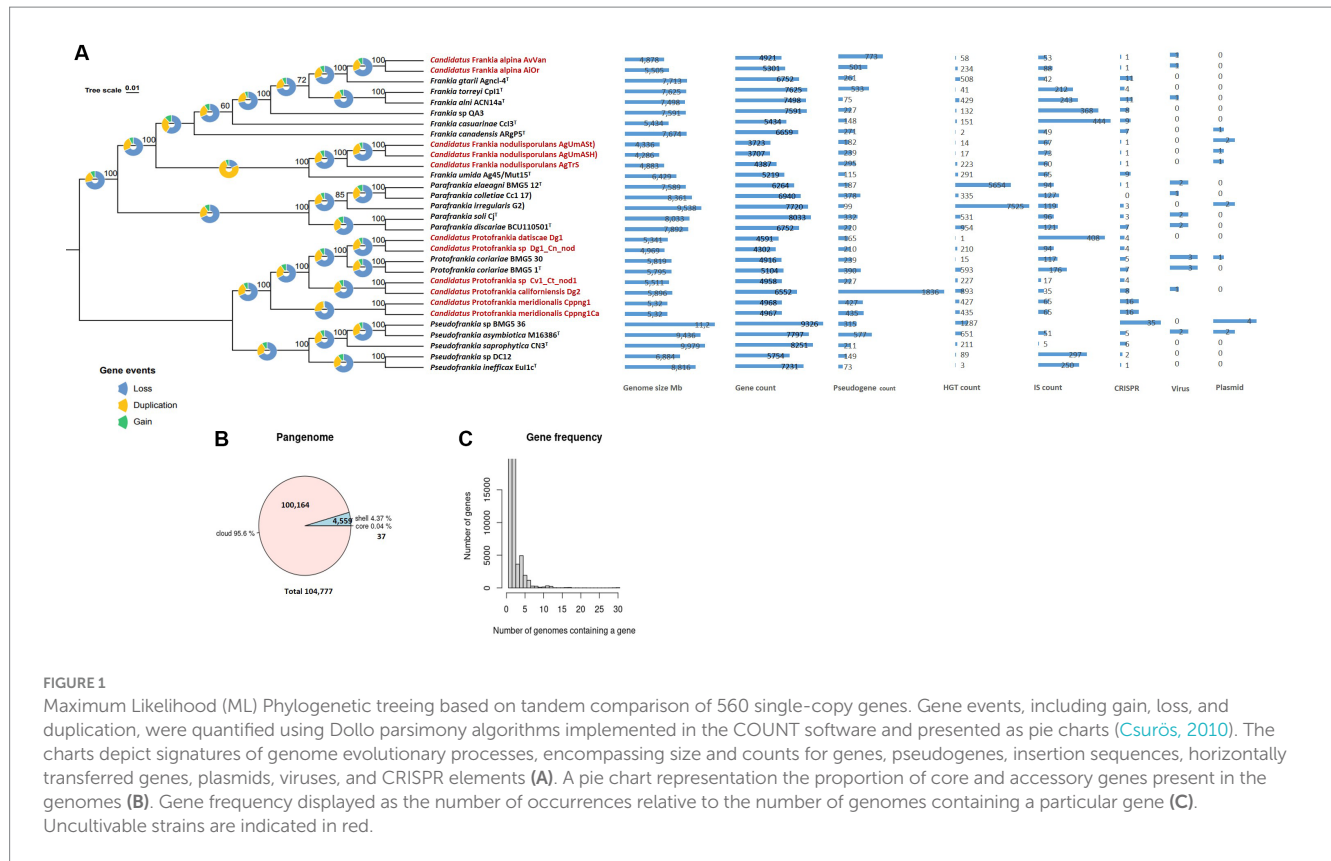


FIGURE 1

Maximum Likelihood (ML) Phylogenetic tree based on tandem comparison of 560 single-copy genes. Gene events, including gain, loss, and duplication, were quantified using Dollo parsimony algorithms implemented in the COUNT software and presented as pie charts (Csurös, 2010). The charts depict signatures of genome evolutionary processes, encompassing size and counts for genes, pseudogenes, insertion sequences, horizontally transferred genes, plasmids, viruses, and CRISPR elements (A). A pie chart representation the proportion of core and accessory genes present in the genomes (B). Gene frequency displayed as the number of occurrences relative to the number of genomes containing a particular gene (C). Uncultivable strains are indicated in red.

comparatively lower than in their cultured counterparts (Figure 2A). Specifically, the counts range from 1,000 to 2,000 for uncultured *Frankia*, whereas they range from 3,000 to 4,000 for cultured strains. Conversely, in the *Protofrankia* group, cultured strains exhibit similar or lower counts (1,000–2,000) compared to their uncultured counterparts (1,000–5,000) (Figure 3A).

Gene-gain events are contingent on the rates of horizontal gene transfer and gene duplication, while gene-loss events hinge on the rates of gene inactivation and deletion (Bordenstein and Reznikoff, 2005). The occurrence of various horizontal gene transfers events, regardless of the cultivation status, underscores the inherent ability of *Frankia* strains, whether cultured or uncultured, to facilitate the exchange of genetic material. This phenomenon is intricately linked to the symbionts' ability to establish contact not only with conspecific strains but also with other bacterial species, a possibility that extends even within the confines of actinorhizal nodules (Ghodhbane-Gtari et al., 2021).

## Pangenome analysis

The pangenome categorizes genes into three groups: core genes, shared universally across all strains (subdivided into core and soft core genes); accessory genes, found in subsets of strains (subdivided into shell and cloud genes); and unique/singleton genes, exclusive to individual strains (Tettelin et al., 2005). Core genes are essential for basic functions (Tettelin et al., 2005; Kim et al., 2020), while accessory and unique genes reflect adaptations to environments,

hosts, or lifestyles (Zhang and Sievert, 2014; Nelson and Stegen, 2015), which may influence their propensity for growth under axenic conditions.

Our investigation of 30 *Frankiaceae* genomes revealed a diverse genomic landscape with 37 universally present core genes, 17 soft core genes in a substantial subset, and 4,559 shell genes (Figure 1B). Additionally, 100,164 cloud genes indicate a dynamic genetic reservoir within this taxonomic group, forming a pangenome of 104,777 genes that offers insights into genetic diversity, adaptability, and evolutionary dynamics.

Analysis of *Frankia* and *Protofrankia* revealed distinct genetic landscapes. The pangenome analysis of *Frankia* strains revealed a total of 19,414 genes (Figure 2B). Among these, 588 core genes were present in 99–100% of the strains, indicating their essential role. The pangenome includes 7,076 shell genes (15–95% presence), contributing to functional diversity, and 11,750 cloud genes (less than 15% presence), highlighting extensive genetic variability.

The pangenome analysis of *Protofrankia* strains revealed a total of 18,411 genes (Figure 3B). Among these, 194 core genes were present in 99–100% of the strains, indicating their essential role. The pangenome includes 12,949 shell genes (15–95% presence), contributing to functional diversity, and 5,268 cloud genes (less than 15% presence), highlighting extensive genetic variability.

No soft-core genes (95–99% presence) were found in the analyzed *Frankia* and *Protofrankia* strains. Soft-core genes are particularly relevant in analyses involving draft genomes as some genes may be missing in draft genomes due to their lower assembly quality than completely assembled genomes (Nelson and Stegen,

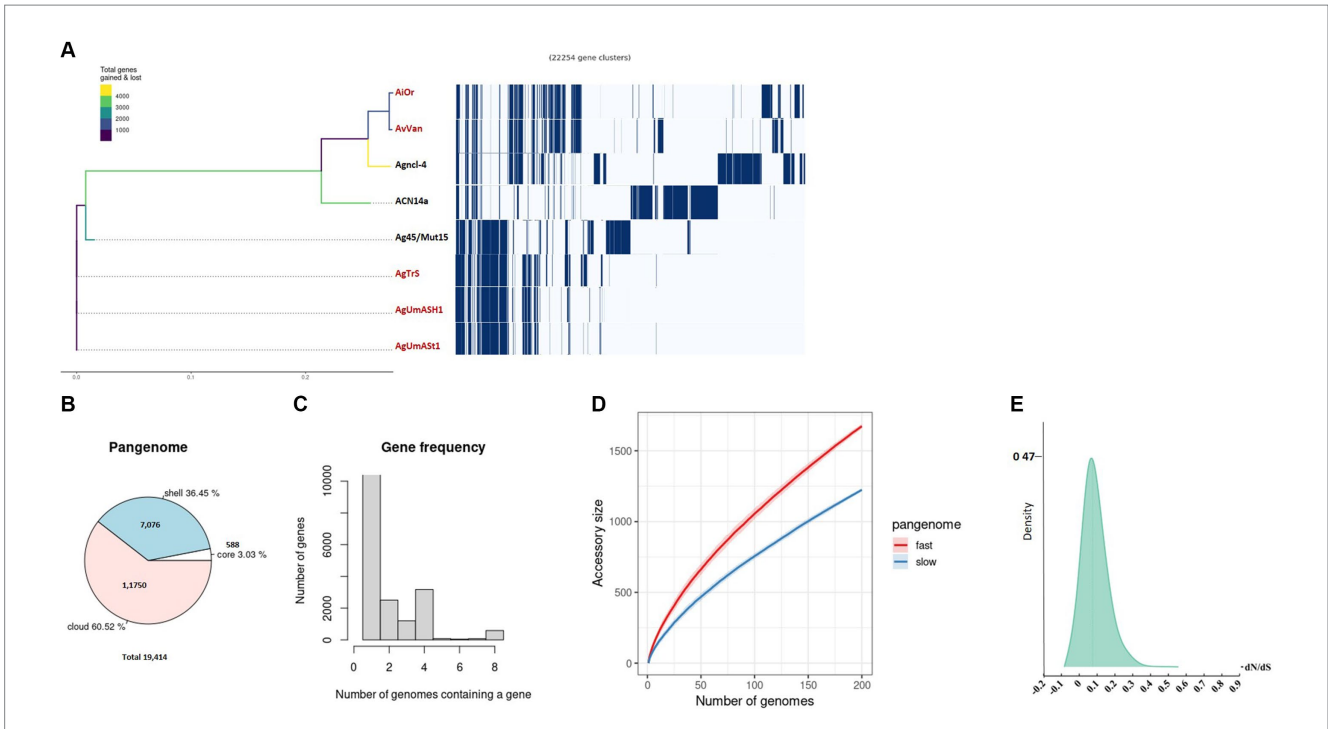


FIGURE 2

Pangenome analysis is conducted on uncultured *Frankia* strains and their closely related cultured counterparts. (A) The phylogenetic branching of uncultured *Frankia* is depicted, accompanied by the gene presence/absence matrix resulting from pangenome analysis, where each row represents the gene profile of an individual genome. A pie chart illustrates the proportion of core and accessory genes across the analyzed genomes (B). Gene frequency is presented as the number of occurrences relative to the number of genomes containing a particular gene (C). Pangenome accumulation curves for accessory genes, showcasing both fast and slow accumulation dynamics, are shown (D). Density plot of dN/dS values (E). Uncultivable strains are indicated in red.

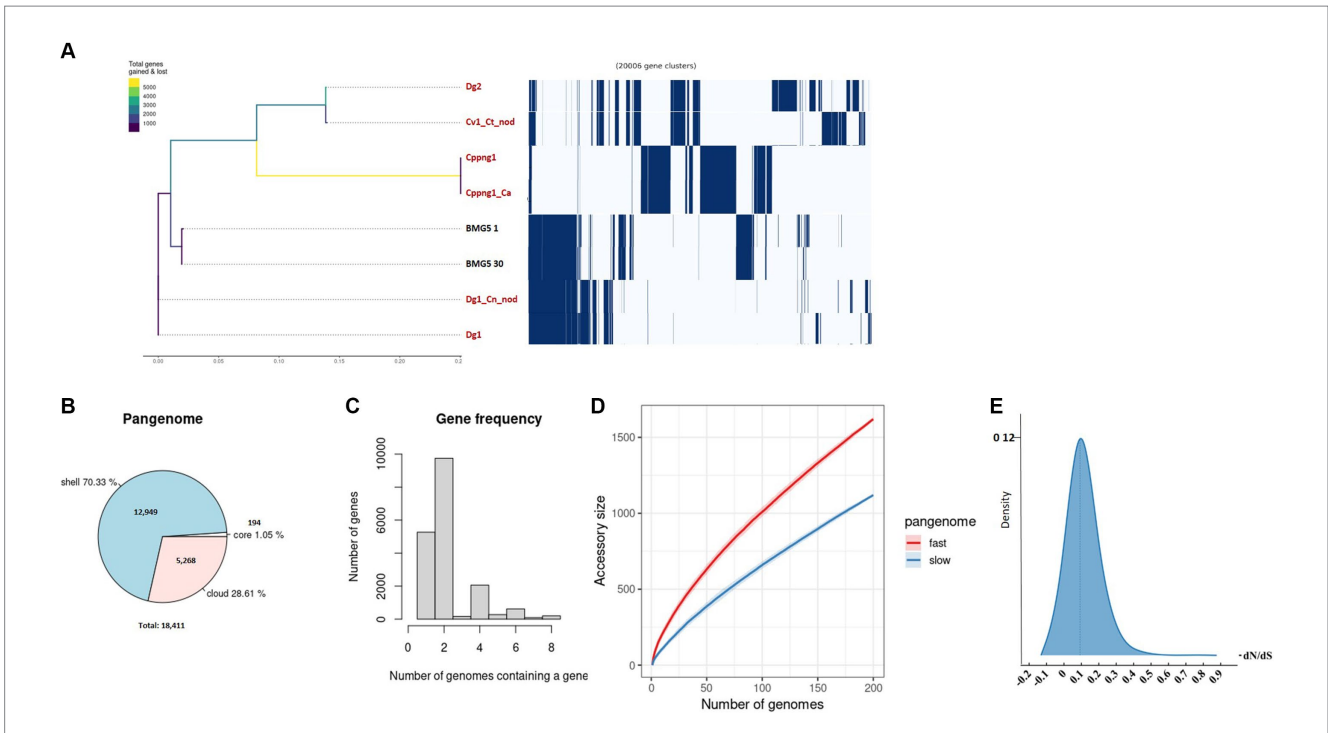


FIGURE 3

An exploration of the pangenome involved the examination of uncultured *Protofrankia* strains and their closely related cultured counterparts. The phylogenetic branching of uncultured *Protofrankia* is portrayed, accompanied by a gene presence/absence matrix resulting from the pangenome analysis, where each row delineates the gene profile of an individual genome (A). Additionally, a pie chart visually represents the proportion of core and accessory genes distributed across the analyzed genomes (B). Gene frequency as the number of occurrences relative to the number of genomes containing a specific gene (C). The dynamics of pangenome accumulation for accessory genes, displaying both rapid and gradual accumulation (D). Density plot of dN/dS values (E). Uncultivable strains are indicated in red.



2015). The difference in core genes (588 vs. 194) between *Frankia* and *Protofrankia* can be attributed to several factors. Sampling bias is a significant factor, as core gene numbers vary with the diversity and quantity of strains analyzed. *Protofrankia* likely shows a smaller core genome due to fewer sequenced strains and closer genetic relationships compared to *Frankia* (Nouioui et al., 2014). Methodological differences also play a crucial role, with core genome estimates significantly impacted by genome incompleteness, fragmentation, and contamination, which are more prevalent in metagenome-assembled genomes (MAGs) compared to single isolate assembled genomes (Li and Yin, 2022). Additionally, genomic diversity affects core genome size and composition; *Frankia* species likely possess a larger core genome due to greater genomic diversification across varied environments, unlike *Protofrankia* species (Gtari, 2022).

Analyses of all three sets—*Frankia*, *Protofrankia*, and *Frankiaceae*—have revealed open pangenomes (Figures 1C, 2C,D, 3C,D), as indicated by the continuous discovery of new gene families with the addition of new genomes (Costa et al., 2020). The configuration of a pangenome, whether open or closed, is intricately tied to the lifestyle of bacterial species, with sympatric species thriving in interactive, communal environments demonstrating open pangenomes characterized by a continuous influx of new genes through horizontal gene transfer (Medini et al., 2005). Conversely, allopatric species, residing in isolated settings, typically exhibit smaller, closed pangenomes (Medini et al., 2005; Georgiades and Raoult, 2011).

Three distinct types of evolutionary selection processes, based on codon substitution, include positive selection ( $dN/dS > 1$ ), purifying selection ( $dN/dS < 1$ ), and neutral selection ( $dN/dS = 1$ ) (Spielman and Wilke, 2015). The average  $dN/dS$  ratios for genes in both *Frankia* and *Protofrankia*, whether cultured or uncultured, were consistently found to be less than 1 (Figures 2E, 3E). This pattern suggests that all analyzed genes are subject to purifying selection, signifying the preservation of functional integrity and evolutionary stability across the analyzed strains (Sarkar and Sen, 2022).

## Homologous genes present uniquely in cultured *Frankia*

The absence of specific genes in uncultivated microsymbionts, as revealed by Orthovenn3 analysis, highlights potential factors contributing to the challenges associated with their cultivability. In cultured *Frankia* strains, 521 unique genes, with functions spanning transcription regulation, fatty acid biosynthesis, and defense responses, suggest enhanced capabilities in cellular regulation, metabolic activities, and defense mechanisms (Tables 2, 3; Supplementary Figure S4). The molecular functions of these genes, including oxidoreductase and ATPase activities, underscore their vital roles in cellular processes. Similarly, *Protofrankia* exhibits 825 exclusive genes, emphasizing its distinct genetic repertoire (Table 3; Supplementary Figure S5). These genes were predominantly associated with transcription regulation and transmembrane transport, indicating potential adaptations to environmental cues and nutrient uptake strategies. The inclusion of functions related to sporulation, phospholipid transport, oxidoreductase, ATPase activities, DNA binding, and metal ion binding in *Protofrankia*'s

exclusive gene set suggests a diverse array of cellular processes. The absence of these genes in uncultivated strains may signify a less ability compared to cultivated strains to regulate gene expression, perform essential metabolic functions, and adapt to diverse environmental conditions, all of which are crucial for successful laboratory cultivation.

## Vital cellular components and central metabolic pathways

### Cell wall and capsule

Bacteria with unconventional cell wall structures can pose challenges for cultivation (Van Teeseling et al., 2015; Markova, 2020; Megrian et al., 2022). Apart from the “Exoenzymes regulatory protein AepA in lipid-linked oligosaccharide synthesis cluster,” absent in certain uncultured *Frankia* strains (AiOr, AvVan, AgTrS, and AgUmASH1), and the “Apolipoprotein N-acyltransferase (EC 2.3.1.-) in lipid-linked oligosaccharide synthesis cluster,” lacking in strain AiOr, all components of the “Cell wall and Capsule pathway” were detected in uncultured strains of *Frankia* and *Protofrankia* compared to their cultivated counterparts (Supplementary Figures S6, S7). This finding suggests that the uncultured *Frankia* and *Protofrankia* strains are capable of generating intact Cell Wall and Capsule structures.

### Dormancy and sporulation

Dynamics of dormancy and sporulation, contributing to our understanding of the cultivability of specific strains (Kell and Young, 2000; Wörmer et al., 2019). Except for the “Spore pigment biosynthetic cluster in Actinomycetes. Polyketide hydroxylase WhiE VIII,” absent in strain Dg2, all components of the “Dormancy and Sporulation, no subcategory” are detectable in uncultured strains of *Frankia* and *Protofrankia* compared to their cultured counterparts. This indicates functional integrity within this pathway across the uncultured strains (Supplementary Figures S6, S7).

### Regulation and signaling

Dearth of regulatory elements in obligate lifestyle stems from the loss of structural genes during reductive evolution is a consequence of adapting to narrow ecological niches (Wilcox et al., 2003). Regulation and Cell Signaling, both cultured and uncultured *Frankia* strains share cAMP signaling components and exhibit consistent expression of essential elements (Supplementary Figure S6). The cAMP signaling, may indicate common regulatory features, and disruptions in these components can influence cultivation success (Vartoukian et al., 2010). The Stringent Response, Programmed Cell Death, and Toxin-Antitoxin Systems are prevalent in all analyzed genomes which contribute to bacterial behavior (Barer and Harwood, 1999; Prozorov and Danilenko, 2010). Uncultured *Protofrankia*, and *Frankia* strains along with cultured related strains share all “Regulatory and cell Signaling” features (Supplementary Figure S7). Common enzymatic activities in the cAMP pathway and consistent presence of regulatory proteins underscore

TABLE 2 GO enrichment of proteins present in cultured and absent in uncultured *Frankia* strains.

| GO term    | Protein count | Description   | Ontology           | p-values                |
|------------|---------------|---|--------------------|-------------------------|
| GO:0006355 | 17            | Regulation of transcription, DNA-templated                  | biological_process | 6.663500585975715 e-64  |
| GO:0006633 | 11            | Fatty acid biosynthetic process                             | biological_process | 7.95951272204721 e-09   |
| GO:0017000 | 10            | Antibiotic biosynthetic process                             | biological_process | 7.927208309074664 e-17  |
| GO:0051607 | 8             | Defense response to virus                                   | biological_process | 0.0002428283280399067   |
| GO:0006099 | 6             | Tricarboxylic acid cycle                                    | biological_process | 2.2038068560022853 e-12 |
| GO:0030435 | 6             | Sporulation resulting in formation of a cellular spore      | biological_process | 8.642292890190224 e-08  |
| GO:0046677 | 4             | Response to antibiotic                                      | biological_process | 5.3922645948956175 e-31 |
| GO:0009405 | 3             | Pathogenesis  | biological_process | 6.955094934645369 e-17  |
| GO:0044873 | 3             | Lipoprotein localization to membrane                        | biological_process | 0.008590721131043375    |
| GO:0009164 | 3             | Nucleoside catabolic process                                | biological_process | 0.008590721131043375    |
| GO:0019439 | 2             | Aromatic compound catabolic process                         | biological_process | 3.6751977821591035 e-08 |
| GO:0006631 | 2             | Fatty acid metabolic process                                | biological_process | 5.746754744693072 e-07  |
| GO:0006352 | 3             | DNA-templated transcription, initiation                     | biological_process | 2.8772858457099777 e-06 |
| GO:0016114 | 2             | Terpenoid biosynthetic process                              | biological_process | 5.3263026401023245 e-06 |
| GO:0055114 | 2             | Oxidation-reduction process                                 | biological_process | 3.014345799313272 e-05  |
| GO:0030497 | 2             | Fatty acid elongation                                       | biological_process | 4.615037074227507 e-05  |
| GO:0045232 | 2             | S-layer organization  | biological_process | 0.016733170871222863    |
| GO:0005886 | 6             | Plasma membrane   | cellular_component | 9.023797064238693 e-26  |
| GO:0016491 | 11            | Oxidoreductase activity                                     | molecular_function | 7.735224554095876 e-42  |
| GO:0046872 | 9             | Metal ion binding   | molecular_function | 1.2421362299588434 e-29 |
| GO:0016887 | 5             | ATPase activity   | molecular_function | 7.142362787033378 e-14  |
| GO:0043565 | 4             | Sequence-specific DNA binding                               | molecular_function | 0.00493561331449674     |
| GO:0008061 | 3             | Chitin binding  | molecular_function | 0.008590721131043375    |
| GO:0003700 | 2             | Sequence-specific DNA binding transcription factor activity | molecular_function | 1.9570856204302962 e-10 |

similarities in regulatory mechanisms across these strains. Shared functionalities in the stringent response, (p)ppGpp metabolism, play a crucial role in bacterial survival under varying conditions (Fagen et al., 2014), and sporulation-related proteins further highlight commonalities in these pathways (Galperin et al., 2012; Browne et al., 2016).

## Central metabolic pathways

The absence or deregulation of specific central metabolic pathways is known to hinder the cultivability of certain bacterial taxa (Pande and Kost, 2017; Bodor et al., 2020; Zhao et al., 2023). Analyzing metabolic pathway data from uncultured *Frankia* and *Protofrankia* strains alongside their closely related cultured counterparts reveals that crucial central metabolic pathways are largely intact across these organisms (Supplementary Figure S8). Key pathways such as glycolysis (Embden-Meyerhof pathway), gluconeogenesis, the citrate cycle (TCA cycle), and the pentose phosphate pathway exhibit high completeness levels. Furthermore, pathways involved in pyruvate oxidation, fatty acid biosynthesis, and various amino acid biosynthesis pathways are also well-represented with substantial completeness (Supplementary Figure S9).

These findings indicate a diverse metabolic repertoire uncultured *Frankia* and *Protofrankia* compared to their cultured counterparts,

enabling efficient utilization of diverse carbon sources and synthesis of essential biomolecules critical for growth and survival. The presence of these pathways underscores the metabolic flexibility of these bacterial strains, likely facilitating their adaptation to different environmental conditions and ecological niches. However, notable exceptions are observed in the uncultured *Protofrankia* Dg2 strain, particularly in pathways such as the Entner-Doudoroff pathway and certain aspects of nucleotide sugar biosynthesis, which display partial completeness. These observations suggest specific metabolic adaptations or limitations in these pathways within the *Protofrankia* Dg2 strain.

## Predictive ecological traits

Conventional genomic analyses often fall short in capturing the ecological intricacies of uncultured microorganisms. The current trend emphasizes organization into a metabolomic blueprint that encapsulates ecological traits (Giovannoni et al., 2014; Lewis et al., 2021; Gtari et al., 2024). This approach, facilitated by advanced bioinformatics tools and algorithms, enables the unveiling of metabolic profiles and the interpretation of cultivability/uncultivability of a given strain. Here we used MicroTrait, an R package (Karaoz and Brodie, 2022) to generate trait profiles for the defined guilds (guilds × traits), with mean (Figures 4A, 5A). Ecological guilds denote groups of species or strains with similar

TABLE 3 GO enrichment of proteins present in cultured and absent in uncultured *Protofrankia* strains.

| GO term    | Protein count | Description   | Ontology           | p-values                |
|------------|---------------|---|--------------------|-------------------------|
| GO:0006355 | 6             | Regulation of transcription, DNA-templated  | biological_process | 5.062598895569772 e-37  |
| GO:0055085 | 5             | Transmembrane transport   | biological_process | 1.0277209374700805 e-26 |
| GO:0030435 | 4             | Sporulation resulting in formation of a cellular spore  | biological_process | 5.046482433806986 e-07  |
| GO:0015914 | 4             | Phospholipid transport  | biological_process | 0.0005561264004201647   |
| GO:0006633 | 4             | Fatty acid biosynthetic process   | biological_process | 1.089172842661041 e-11  |
| GO:0046677 | 2             | Response to antibiotic  | biological_process | 1.1207957183541802 e-21 |
| GO:0006260 | 2             | DNA replication   | biological_process | 1.7912991284821486 e-07 |
| GO:0006777 | 2             | Mo-molybdopterin cofactor biosynthetic process  | biological_process | 0.008059777854697709    |
| GO:0006281 | 2             | DNA repair  | biological_process | 2.3291849989590412 e-05 |
| GO:0006099 | 2             | Tricarboxylic acid cycle  | biological_process | 4.709376529576225 e-10  |
| GO:0006313 | 2             | Transposition, DNA-mediated   | biological_process | 2.163482211983071 e-08  |
| GO:0055114 | 2             | Oxidation-reduction process   | biological_process | 0.00013337251121081238  |
| GO:0005886 | 2             | Plasma membrane   | cellular_component | 2.3387610709332023 e-15 |
| GO:0016491 | 10            | Oxidoreductase activity   | molecular_function | 1.532736552223947 e-16  |
| GO:0016705 | 6             | Oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen | molecular_function | 1.750694992145492 e-10  |
| GO:0016887 | 6             | ATPase activity   | molecular_function | 3.598716601114218 e-08  |
| GO:0003677 | 4             | DNA binding   | molecular_function | 4.2368584396231013 e-16 |
| GO:0046872 | 4             | Metal ion binding   | molecular_function | 7.974237522227637 e-27  |
| GO:0004674 | 2             | Protein serine/threonine kinase activity  | molecular_function | 5.086789909828374 e-14  |
| GO:0043565 | 2             | Sequence-specific DNA binding   | molecular_function | 0.00024753581859025836  |
| GO:0016627 | 2             | Oxidoreductase activity, acting on the CH-CH group of donors  | molecular_function | 0.000391191404846266    |
| GO:0008061 | 2             | Chitin binding  | molecular_function | 0.023850892434054655    |
| GO:0004090 | 2             | Carbonyl reductase (NADPH) activity   | molecular_function | 0.023850892434054655    |

resource-use strategies. AntiSMASH 7.0 (Blin et al., 2023), GapMind (Price et al., 2020, 2021) and PhenDB (Feldbauer et al., 2015), to extract other fitness traits from *Frankia* genome sequences.

## Resource acquisition traits

In contrast to obligate symbionts and pathogens that rely on hosts for nutrients, saprophytic and facultative symbiotic microorganisms can utilize decaying organic matter (Ray et al., 2009; Salje, 2021). Their efficient decomposition involves rich enzymatic sets for breakdown, membrane transporters, and uptake through central metabolic pathways, providing nutrients and energy for cellular functions (Cao et al., 2023). Comparing cultured and uncultured strains in complex carbohydrate depolymerization and resource acquisition pathways reveals shared features correlating with similar nutritional requirements (Figures 4A, 5A). Both groups exhibit proficiency in breaking down complex carbohydrates (cellulose, chitin, heteromannan, mixed linkage glucan, xylan, and xyloglucan) and degrading simple compounds (proteins, EMP pathway carbohydrates, maltose, galactose, mannose, trehalose, and glycerol). For substrate uptake, both groups show similarities in transporting aromatic acids, monosaccharides, polysaccharides, dicarboxylates, monocarboxylates, free amino acids, inorganic ions, metal ions, amides, ammonium, nucleotides, organophosphorus compounds, sulfur compounds,

siderophores, and vitamin B. Both groups display balanced assimilation of C1 compounds via CO<sub>2</sub> fixation through the dicarboxylate hydroxybutyrate pathway, methanotrophy utilizing formaldehyde oxidation or assimilation with serine, assimilatory nitrate reduction, nitrogen fixation, ammonia assimilation to aspartate, glutamate, or glutamine, assimilatory sulfate reduction, and phosphatase-mediated phosphorus compound assimilation. They also exhibit the ability to ferment pyruvate to short-chain fatty acids (SCFAs) and use the oxidative pentose phosphate pathway for aerobic respiration, sharing the presence of the electron transport chain (ETC) complex III. In terms of phototrophy, both cultured and uncultured strains produce the accessory carotenoid pigments neurosporene and lycopene. However, specific pathways like the aerobic respiration nonoxidative pentose phosphate pathway and the glyoxylate cycle are found in various cultured strains but are absent in some uncultured strains. Additionally, the electron transport chain ETC complex IV is present in cultured but missing in all uncultured strains (Figures 4A, 5A). These distinctions highlight the enhanced metabolic diversity and adaptability of cultured strains under controlled conditions.

## Stress tolerance

While host tissues can impose chemical and/or physical stresses, notably through the host immune system's response to

TABLE 4 Main predicted features of uncultured *Frankia* and *Protofrankia* and their closely related counterparts using phendb server.

| Strain        | Gram stain | Aerobe | sporulation | Symbiont | Saccharolytic | Acetic acid production | Ethanol production | motility | <i>nif</i> |
|---------------|------------|--------|-------------|----------|---------------|------------------------|--------------------|----------|------------|
| Dg1           | +          | +      | +           | –        | +             | +                      | +                  | –        | +          |
| BMG5.1        | +          | +      | +           | –        | +             | +                      | +                  | –        | +          |
| BMG5.30       | +          | +      | +           | –        | +             | +                      | +                  | –        | +          |
| Cv_Ct_nod1    | +          | +      | +           | –        | +             | +                      | +                  | –        | +          |
| Dg2           | +          | +      | +           | –        | +             | +                      | +                  | –        | +          |
| Cm1_Cm_nod    | +          | +      | +           | –        | +             | +                      | +                  | –        | +          |
| Cppng1_Ca_nod | +          | +      | n.d.        | –        | +             | +                      | +                  | –        | +          |
| Cppng1        | +          | +      | n.d.        | –        | +             | +                      | +                  | –        | +          |
| ACN14         | +          | +      | +           | –        | +             | n.d.                   | +                  | –        | +          |
| Ag45/Mut15    | +          | +      | +           | –        | +             | +                      | +                  | –        | +          |
| Agnc1-4       | +          | +      | +           | –        | +             | +                      | +                  | –        | +          |
| AgUmASt1      | +          | +      | n.d.        | –        | +             | +                      | +                  | –        | +          |
| AgTrS         | +          | +      | n.d.        | –        | +             | +                      | +                  | –        | +          |
| AgUm          | +          | +      | n.d.        | –        | +             | +                      | +                  | –        | +          |
| AvVan         | +          | +      | n.d.        | –        | +             | +                      | +                  | –        | +          |
| AiOr          | +          | +      | n.d.        | –        | +             | +                      | +                  | –        | +          |

Default settings parameters were used with a balanced accuracy cutoff of 0.75 and a prediction confidence cutoff of 0.6 (Feldbauer et al., 2015). n.d., not determined.

microbe-associated molecular patterns (Schwartzman and Ruby, 2016; Bénard et al., 2020), an obligate lifestyle can be regarded as an adaptive response to narrow ecological niches, providing a buffered environment conducive to maintaining homeostasis. This adaptation involves the regulation of pH, temperature, and osmotic stress, enabling the host to efficiently mitigate these environmental challenges.

The comparison of stress tolerance between cultured and uncultured *Frankia* strains showed extensive shared behaviors (Figure 4A). Both groups excel in scavenging reactive oxygen species (ROS) and maintaining redox balance, indicating strong defenses against oxidative stress (score 1). Cultured strains, however, exhibit enhanced biofilm formation and EPS biosynthesis/export compared to uncultured strains, potentially increasing their ability to form structured communities. Under heat stress, both groups demonstrate robust protein protection and repair mechanisms involving ATP-dependent proteases. Cultured strains also display superior cold-sensing machinery, highlighting their adaptability to temperature fluctuations. In desiccation, osmotic, and salt stress, both groups show similar proficiency in compatible solute synthesis and transport, suggesting shared strategies for osmoregulation. All strains possess redox-sensing systems for pH adaptation, crucial for survival in diverse environmental conditions. Cultured *Protofrankia* strains exhibit higher enzyme counts in ROS scavenging, suggesting distinct oxidative stress response mechanisms compared to uncultured strains (Figure 5A). While biofilm formation varies, both groups demonstrate comparable abilities in protecting, repairing, and degrading denatured proteins (score 4–5), emphasizing their resilience under stress conditions. Temperature-specific, osmotic, salt stress responses, and pH stress responses show variations among strains but consistent patterns within each group. Importantly, all strains share capabilities in ROS scavenging, redox-sensing, and response to oxygen limitation

and envelope stress, including phage resistance, enhancing their resilience in challenging environments.

It is important to note that cultured *Protofrankia* strains have showcased a notable preference for alkalophilic conditions, setting them apart from other cultured members of the *Frankiaceae* family (Gtari et al., 2015). This investigation further delineates these distinctions, revealing that *Frankia* strains generally display a heightened capacity for the consumption of excessive cytoplasmic protons when compared to *Protofrankia*. Significantly, specific enzyme systems implicated in pH stress responses exhibit variations between the two genera. *Frankia* strains, in particular, tend to favor misfolded proteins ATP-dependent proteases, decarboxylase, and antiporter systems, in contrast to *Protofrankia*. These discerning findings underscore the unique adaptive strategies employed by *Frankia* and *Protofrankia* in response to pH stress, potentially elucidating the alkalophilic preference observed in *Protofrankia*.

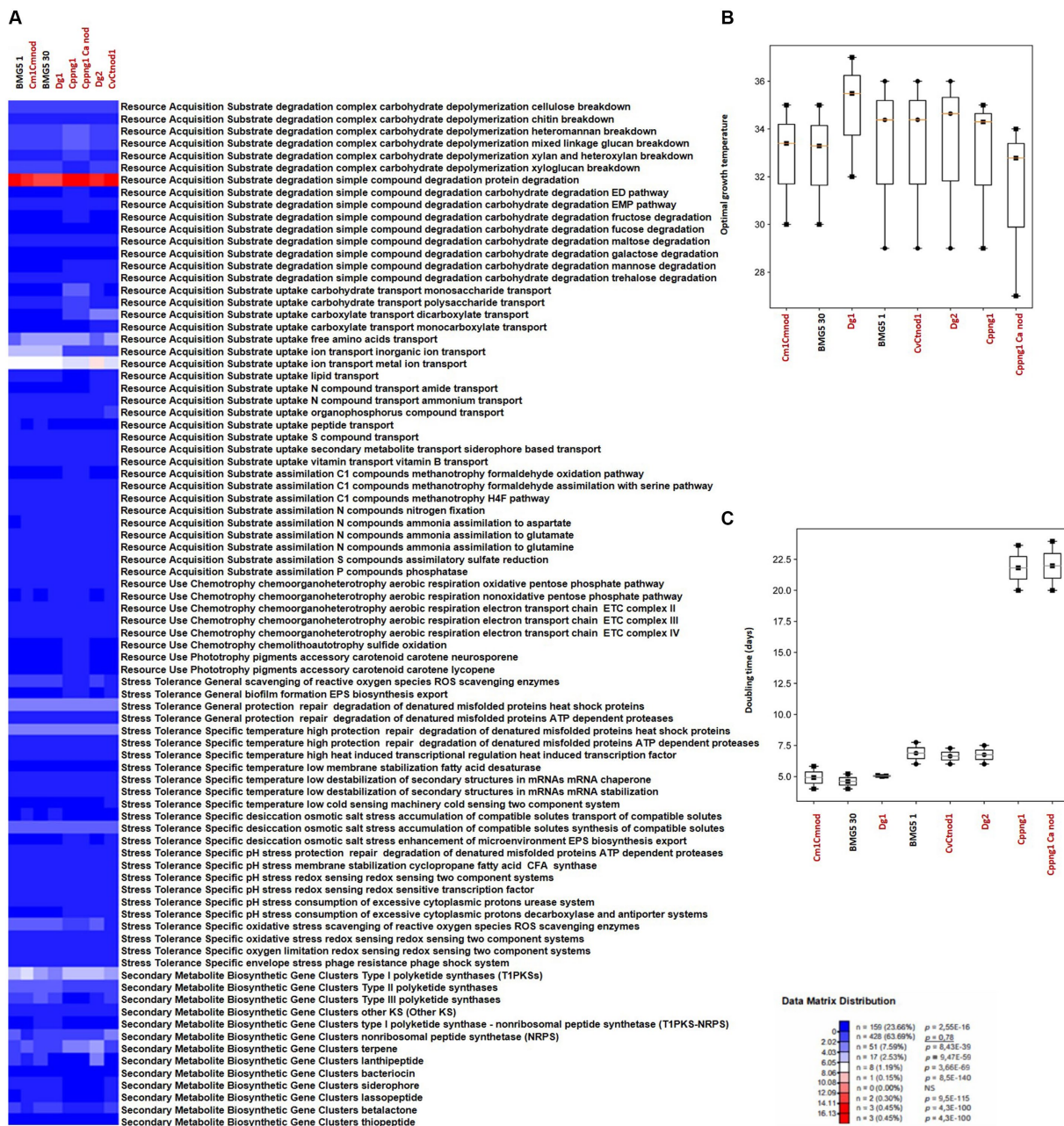
## Secondary metabolite gene clusters

Secondary metabolites, distinct from primary metabolites, are characterized by their low molecular mass and do not play direct roles in the growth, development, or reproduction of the producing organism. However, they are often produced during the late growth phase of the microorganisms and play crucial roles in adaptation, competition, and defense mechanisms, conferring fitness advantages (O'Brien and Wright, 2011). These compounds are versatile in their biological functions, participating in inter-microbial warfare through processes such as antibiotic biosynthesis and resistance (Perry et al., 2022) and bacteriocin production (Riley and Wertz, 2002). Additionally, they can act as mediators of cross-species mutualism (O'Brien and Wright, 2011), or even participate in both processes, as



seen with siderophores (Kramer et al., 2020). It's crucial to note that obligate symbiotic bacteria, such as those residing within root nodules, may produce a distinct set of secondary metabolites compared to facultative symbiotic and cultivable counterparts (Pinto-Carbó et al., 2016). This difference arises from their specialized

lifestyle, which prioritizes mutualistic interactions within host environments rather than competitive survival in broader ecological settings. As a result, these bacteria may focus on producing metabolites that support their symbiotic relationship with the host plant, such as compounds involved in nitrogen fixation or modulation



**FIGURE 5**  
 Overview of microbial traits generated by the MicroTrait pipeline, which extracts fitness traits from microbial genome sequences (Karaoz and Brodie, 2022). This figure showcases ecological trait profiles for specific guilds in uncultured *Protofrankia* (Cm1Cmmod, Dg1, CvCtnod1, Dg2, Cppng1, and Cppng1 Ca nod) and their cultured counterparts, covering resource acquisition, utilization, and stress tolerance. In microTrait, each protein family is a Boolean variable (1 if detected, 0 if not), determined by microTrait-HMMs (Supplementary Table S5), with traits defined by rules using these variables. Secondary metabolite clusters identified using AntiSMASH 7.0 enhance trait visualization (A). Predictions for optimal temperature (Sauer and Wang, 2019) (B) and doubling time (Veira-Silva and Rocha, 2010; Weissman et al., 2021) (C) provide comprehensive insights. The top panel shows the statistical significance of comparing mean trait values across guilds.

of plant physiology, rather than defensive or competitive compounds typically found in free-living bacteria (Gtari et al., 2015).

The analysis of biosynthetic gene clusters in both cultivated and uncultivated *Frankia* strains revealed that both groups exhibit the presence of polyketide synthases (t2PKS, t3PKS, otherKS), non-ribosomal peptide synthetases (NRPS), and genes related to terpene, siderophore, and lassopeptide biosynthesis, suggesting a

common potential for secondary metabolite production (Figure 4A). However, cultured strains show higher counts in specific clusters, including t1PKS1, NRPS, and terpene genes, indicating potential metabolic advantages in these pathways. Conversely, the absence of specific clusters, such as lassopeptide and betalactone, in uncultivated strains suggests potential ecological adaptations and differences in secondary metabolite production. For cultivated and uncultivated

*Protofrankia* strains, shared polyketide synthases (t2PKS, otherKS), NRPS, and genes related to terpene and betalactone biosynthesis are observed (Figure 5A). The presence of siderophore biosynthetic genes suggests a common strategy for iron acquisition. Distinctions, such as lassopeptide genes in cultivated strains and unique distributions of t1PKS1, t3PKS, and lanthipeptide genes in specific uncultivated strains, imply adaptations to specific ecological niches.

## Prediction of lifestyle, nutrition exigence, optimal growth temperature and doubling time

The PhenDB web server analysis conclusively determined that both cultured and uncultured genomes of *Frankia* and *Protofrankia* do not exhibit characteristics typical of obligate intracellular symbionts (Table 4). PhenDB predicted an aerobic nature for all strains, consistent with established knowledge for cultured *Frankia*. The GapMind server analysis identified alanine, aspartate, fumarate, glucose-6-P, L-malate, 2-oxoglutarate, pyruvate, succinate, and propionate as optimal carbon sources for the majority of *Frankia* and *Protofrankia* (Supplementary Table S2), aligning with well-established compositions for routine growing media useful for *Frankia* strains (Blom, 1982; Murry et al., 1984; Lopez et al., 1986; Stowers et al., 1986; Gtari et al., 2015). Notably, Microtrait predicted ammonia as the mineral nitrogen source, excluding nitrite and nitrate, consistent with findings by Blom (1982) and Shipton and Burggraaf (1982). Additionally, GapMind predicted no auxotrophy for any amino acids (Supplementary Table S3).

The Microtrait pipeline predicted distinct temperature preferences and growth rates among cultured and uncultured *Frankia* and *Protofrankia* strains (Figures 4B, 5B). Cultured *Frankia* strains (ACN14a, Ag45/Mut15, Agncl-4) prefer temperatures around 28–29°C. Diverse temperature adaptations are observed in cultured CcI3, CpI1, and QA3, with optimal temperatures of 34.78, 28.75, and 28.85°C, respectively. In contrast, uncultured strains (AgTrS, AiOr, AvVan) exhibit a slightly higher temperature preference at 31–32°C. Uncultured *Frankia* strains UmASH1 and UmAS1, along with most *Protofrankia* strains, lean toward even higher temperatures, around 32.55 and 33.34°C, showcasing variability in temperature preferences.

These optimal temperature values, ranging from 28°C to 35.49°C, generally align with experimentally determined values (Burggraaf and Shipton, 1982; Gtari et al., 2015). For example, the optimal temperature for CpI1, predicted here at 28.75°C, was experimentally determined to be between 30°C and 35°C (Burggraaf and Shipton, 1982). Additionally, strain Agncl-4, predicted here at 28°C, grew optimally between 25°C and 37°C (Nouioui et al., 2023). *Casuarina* isolates, including *Frankia casuarinea* CcI3, which was predicted here at 34.78°C, exhibited maximum growth between 25°C and 37°C (Sayed et al., 1997).

The growth rates among *Frankia* strains vary significantly (Figure 4C). Ag45/Mut15 displays the fastest growth with a doubling time of 3.26 days, followed closely by ACN14a and AgTrS at 4.35 and 3.65 days, respectively. CpI1, UmAS1, UmASH1, and QA3 exhibit moderate growth rates, ranging from 4.75 to 5.48 days. AvVan and CcI3 have slightly longer doubling times at 6.18 and 5.23 days. Notably,

Agncl-4 stands out with the slowest growth, displaying a doubling time of 9.60 days. For *Protofrankia*, BMG5.30 is an efficient grower with a short doubling time of 4.60 days (Figure 5C). Cm1Cmnd and Cv1Ctnod1 also show efficient growth (4.91 and 6.64 days) while Dg1 and Dg2 fall into the category of moderate growers (5.04 and 6.75 days). In contrast, Cppng1 and Cppng1Canod exhibit longer doubling times at 21.82 and 21.98 days. The predicted doubling times in this study are consistent with previously experimentally determined values for *Frankia* strain ArI3 (3.8 days, Ringø et al., 1995) and *Frankia torreyi* CpI1 (4.03 days, Burggraaf and Shipton, 1983). However, the experimentally determined doubling time for *Frankia casuarinea* CcI3, which ranges from 1 to 2 days depending on the medium and incubation conditions (Zhongze et al., 1986; Huang and Benson, 2012), is notably shorter than the predicted values obtained in the present study.

## Conclusion and perspectives

Our in-depth analysis explores the cultivability potential of previously deemed unculturable *Frankia* and *Protofrankia* strains. Despite challenges posed by intracellular lifestyle, genomic intricacies, and ecological adaptations, our findings highlight potential cultivability traits. Genomic comparisons reveal differences, offering insights into the ongoing evolutionary changes and adaptations shaped by environmental pressures and biological interactions. Pangenome analysis unveils substantial diversity, suggesting untapped genetic resources for cultivation. Shared metabolic strategies in cellular components, central metabolic pathways, and resource acquisition traits present promising avenues for future cultivation attempts. Stress tolerance mechanisms demonstrate resilience, indicating adaptability to controlled conditions. Secondary metabolite clusters unveil potential ecological competitiveness with potential predisposition to asymbiotic growth in axenic conditions. This study offers a nuanced perspective, highlighting potential breakthroughs in tailored media formulations and optimal growth conditions. Future efforts aimed at cultivating as-yet-uncultured *Frankia* and *Protofrankia* under axenic conditions should prioritize using a mineral growth medium enriched with simple carbon sources like alanine, aspartate, fumarate, glucose-6-P, L-malate, 2-oxoglutarate, pyruvate, succinate, and propionate. These should be provided in aerobic environments, with atmospheric nitrogen being optionally sufficient to minimize contamination risks, though ammonia can be included as a preferred alternative nitrogen source. Temperature adjustments should align with strain preferences: cultured *Frankia* strains thrive at 28–29°C, while *Protofrankia* strains prefer slightly higher temperatures of 32–35°C, maintaining an alkaline pH range for the latter. Given the potential for extended incubation periods required for growth (with predicted doubling times ranging from 3.26 to 9.60 days, but possibly reaching up to 21.98 days), patience is essential, and meticulous monitoring for contaminants is crucial to optimize cultivation conditions.

Recent advancements in microbial technologies have indeed revolutionized our understanding of the challenges associated with yet uncultured microorganisms. Advanced computational

tools now enable the extraction of mechanistic insights that influence how microorganisms adapt to laboratory conditions. However, despite the substantial benefits offered by genomics, it requires meticulous consideration of potential biases and errors. Issues such as assembly errors, annotation discrepancies, and variations in sequencing depth can distort gene content and complicate the interpretation of data (Stein, 2001; Quince et al., 2017; Tam et al., 2019). These factors are particularly critical in studies utilizing comparative genomics or downstream applications. MAGs encounter additional obstacles, including gaps resulting from low-read depths (Chu et al., 2019), fragmentation due to strain divergence (Chen et al., 2020), and assembly errors like repeat collapse and chimeric reads (Alneberg et al., 2014; Olson et al., 2019; Mineeva et al., 2020). Short scaffolds can lead to binning errors and the presence of unreliable MAGs contaminated by phage or plasmid fragments (Chen et al., 2020). These limitations underscore the critical importance of rigorous quality assessment in the interpretation of MAG data, particularly within the fields of microbiological and environmental genomics.

## Data availability statement

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding author.

## Author contributions

MG: Conceptualization, Investigation, Software, Writing – original draft, Writing – review & editing. RM: Formal analysis, Investigation, Software, Writing – review & editing. FG-G: Investigation, Software, Writing – review & editing. KB: Funding

## References

- Alain, K., and Querellou, J. (2009). Cultivating the uncultured: limits, advances and future challenges. *Extremophiles* 13, 583–594. doi: 10.1007/s00792-009-0261-3
- Alneberg, J., Bjarnason, B. S., De Bruijn, I., Schirmer, M., Quick, J., Jjaz, U. Z., et al. (2014). Binning metagenomic contigs by coverage and composition. *Nat. Methods* 11, 1144–1146. doi: 10.1038/nmeth.3103
- Arnold, B. J., Huang, I. T., and Hanage, W. P. (2022). Horizontal gene transfer and adaptive evolution in bacteria. *Nat. Rev. Microbiol.* 20, 206–218. doi: 10.1038/s41579-021-00650-4
- Barer, M. R., and Harwood, C. R. (1999). Bacterial viability and culturability. *Advances in microbial physiology*, 41, 93–137.
- Becking, J. H. (1970). Frankiaceae fam. Nov.(Actinomycetales) with one new combination and six new species of the genus Frankia Brunchorst 1886, 174. *Int. J. Syst. Evol. Microbiol.* 20, 201–220.
- Bénard, A., Vavre, F., and Kremer, N. (2020). Stress & symbiosis: heads or tails? *Front. Ecol. Evol.* 8:167. doi: 10.3389/fevo.2020.00167
- Berckx, F., Nguyen, T. V., Bandong, C. M., Lin, H. H., Yamanaka, T., Katayama, S., et al. (2022). A tale of two lineages: how the strains of the earliest divergent symbiotic Frankia clade spread over the world. *BMC Genomics* 23:602. doi: 10.1186/s12864-022-08838-5
- Bland, C., Ramsey, T. L., Sabree, F., Lowe, M., Brown, K., Kyrpides, N. C., et al. (2007). CRISPR recognition tool (CRT): a tool for automatic detection of clustered regularly interspaced palindromic repeats. *BMC Bioinformatics* 8, 1–8. doi: 10.1186/1471-2105-8-209
- Blin, K., Shaw, S., Augustijn, H. E., Reitz, Z. L., Biermann, F., Alanjary, M., et al. (2023). antiSMASH 7.0: new and improved predictions for detection, regulation, chemical structures and visualisation. *Nucleic Acids Res.* 51:gakad344. doi: 10.1093/nar/gkad344
- Blom, J. (1982). Carbon and nitrogen source requirements of Frankia strains. *FEMS Microbiol. Lett.* 13, 51–55. doi: 10.1111/j.1574-6968.1982.tb08226.x
- Bodor, A., Bounedjoum, N., Vincze, G. E., Erdeiné Kis, Á., Laczi, K., Bende, G., et al. (2020). Challenges of unculturable bacteria: environmental perspectives. *Rev. Environ. Sci. Biotechnol.* 19, 1–22. doi: 10.1007/s11157-020-09522-4
- Bordenstein, S. R., and Reznikoff, W. S. (2005). Mobile DNA in obligate intracellular bacteria. *Nat. Rev. Microbiol.* 3, 688–699. doi: 10.1038/nrmicro1233
- Bowers, R. M., Kyrpides, N. C., Stepanauskas, R., Harmon-Smith, M., Doud, D., Reddy, T. B. K., et al. (2017). Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat. Biotechnol.* 35, 725–731. doi: 10.1038/nbt.3893
- Brito, I. L. (2021). Examining horizontal gene transfer in microbial communities. *Nat. Rev. Microbiol.* 19, 442–453. doi: 10.1038/s41579-021-00534-7
- Browne, H. P., Forster, S. C., Anonye, B. O., Kumar, N., Neville, B. A., Stares, M. D., et al. (2016). Culturing of ‘unculturable’ human microbiota reveals novel taxa and extensive sporulation. *Nature* 533, 543–546. doi: 10.1038/nature17645
- Burggraaf, A. J. P., and Shipton, W. A. (1982). Estimates of Frankia growth under various pH and temperature regimes. *Plant Soil* 69, 135–147. doi: 10.1007/BF02374509
- Burggraaf, A. J. P., and Shipton, W. A. (1983). Studies on the growth of Frankia isolates in relation to infectivity and nitrogen fixation (acetylene reduction). *Can. J. Bot.* 61, 2774–2782. doi: 10.1139/b83-307
- Callahan, D., Del Tredici, P., and Torrey, J. G. (1978). Isolation and cultivation in vitro of the actinomycete causing root nodulation in Comptonia. *Science* 199, 899–902. doi: 10.1126/science.199.4331.899

acquisition, Writing – review & editing. IS: Investigation, Software, Writing – review & editing.

## Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This contribution is financially supported by the Ministry of Higher Education and Scientific Research of Tunisia for the Unit of Molecular Bacteriology & Genomics of the National Institute of Applied Science and Technology, University of Carthage.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2024.1367490/full#supplementary-material>



- Cao, T., Luo, Y., Shi, M., Tian, X., and Kuzuyakov, Y. (2023). Microbial interactions for nutrient acquisition in soil: miners, scavengers, and carriers. *Soil Biol. Biochem.* 188:109215. doi: 10.1016/j.soilbio.2023.109215
- Casadevall, A. (2008). Evolution of intracellular pathogens. *Ann. Rev. Microbiol.* 62, 19–33. doi: 10.1146/annurev.micro.61.080706.093305
- Chen, L. X., Anantharaman, K., Shaiber, A., Eren, A. M., and Banfield, J. F. (2020). Accurate and complete genomes from metagenomes. *Genome Res.* 30, 315–333. doi: 10.1101/gr.258640.119
- Chen, I. M. A., Chu, K., Palaniappan, K., Ratner, A., Huang, J., Huntemann, M., et al. (2023). The IMG/M data management and analysis system v. 7: content updates and new features. *Nucleic Acids Res.* 51, D723–D732. doi: 10.1093/nar/gkac976
- Chu, C., Li, X., and Wu, Y. (2019). GAPPadder: a sensitive approach for closing gaps on draft genomes with short sequence reads. *BMC Genomics* 20, 1–10. doi: 10.1186/s12864-019-5703-4
- Chun, J., Oren, A., Ventosa, A., Christensen, H., Arahall, D. R., da Costa, M. S., et al. (2018). Proposed minimal standards for the use of genome data for the taxonomy of prokaryotes. *Int. J. Syst. Evol. Microbiol.* 68, 461–466. doi: 10.1099/ijsem.0.002516
- Connon, S. A., and Giovannoni, S. J. (2002). High-throughput methods for culturing microorganisms in very-low-nutrient media yield diverse new marine isolates. *Appl. Environ. Microbiol.* 68, 3878–3885. doi: 10.1128/AEM.68.8.3878-3885.2002
- Costa, S. S., Guimarães, L. C., Silva, A., Soares, S. C., and Baraúna, R. A. (2020). First steps in the analysis of prokaryotic pan-genomes. *Bioinformatics Biol. Insights* 14, 117793222093806–117793222093809. doi: 10.1177/1177932220938064
- Csurös, M. (2010). Count: evolutionary analysis of phylogenetic profiles with parsimony and likelihood. *Bioinformatics* 26, 1910–1912. doi: 10.1093/bioinformatics/btq315
- Fagen, J. R., Leonard, M. T., McCullough, C. M., Edirisinghe, J. N., Henry, C. S., Davis, M. J., et al. (2014). Comparative genomics of cultured and uncultured strains suggests genes essential for free-living growth of *Liberibacter*. *PLoS One*, 9, e84469.
- Feldbauer, R., Schulz, F., Horn, M., and Rattai, T. (2015). Prediction of microbial phenotypes based on comparative genomics. *BMC Bioinformatics* 16, 1–8. doi: 10.1186/1471-2105-16-S14-S1
- Galperin, M. Y., Kristensen, D. M., Makarova, K. S., Wolf, Y. I., and Koonin, E. V. (2019). Microbial genome analysis: the COG approach. *Brief. Bioinform.* 20, 1063–1070. doi: 10.1093/bib/bbx117
- Galperin, M. Y., Mekhedov, S. L., Puigbo, P., Smirnov, S., Wolf, Y. I., and Rigden, D. J. (2012). Genomic determinants of sporulation in Bacilli and Clostridia: towards the minimal set of sporulation-specific genes. *Environ. Microbiol.* 14, 2870–2890. doi: 10.1111/j.1462-2920.2012.02841.x
- Georgiades, K., and Raouf, D. (2011). Defining pathogenic bacterial species in the genomic era. *Front. Microbiol.* 1:151. doi: 10.3389/fmicb.2010.00151
- Ghodhbane-Gtari, F., D'Angelo, T., Gueddou, A., Ghazouani, S., Gtari, M., and Tisa, L. S. (2021). Alone yet not alone: *Frankia* lives under the same roof with other bacteria in actinorhizal nodules. *Front. Microbiol.* 12:749760. doi: 10.3389/fmicb.2021.749760
- Ghodhbane-Gtari, F., Nouioui, I., Chair, M., Boudabous, A., and Gtari, M. (2010). 16S–23S rRNA intergenic spacer region variability in the genus *Frankia*. *Microb. Ecol.* 60, 487–495. doi: 10.1007/s00248-010-9641-6
- Giovannoni, S. J., Cameron Thrash, J., and Temperton, B. (2014). Implications of streamlining theory for microbial ecology. *ISME J.* 8, 1553–1565. doi: 10.1038/ismej.2014.60
- Gtari, M. (2022). Taxogenomic status of phylogenetically distant *Frankia* clusters warrants their elevation to the rank of genus: a description of *Protofrankia* gen. Nov., *Parafrankia* gen. Nov., and *Pseudofrankia* gen. Nov. as three novel genera within the family *Frankiaceae*. *Front. Microbiol.* 13:1041425. doi: 10.3389/fmicb.2022.1041425
- Gtari, M., Beauchemin, N. J., Sarker, I., Sen, A., Ghodhbane-Gtari, F., and Tisa, L. S. (2024). An overview of *Parafrankia* (nod+/fix+) and *Pseudofrankia* (nod+/fix-) interactions through genome mining and experimental modeling in co-culture and co-inoculation of *Elaeagnus angustifolia*. *Appl. Environ. Microbiol.* 90, e00288–e00224. doi: 10.1128/aem.00288-24
- Gtari, M., Ghodhbane-Gtari, F., Nouioui, I., Ktari, A., Hezbri, K., Mimouni, W., et al. (2015). Cultivating the uncultured: growing the recalcitrant cluster-2 *Frankia* strains. *Sci. Rep.* 5:13112. doi: 10.1038/srep13112
- Gu, Z. (2022). Complex heatmap visualization. *Imeta* 1:e43. doi: 10.1002/imt2.43
- Gueddou, A., Swanson, E., Hezbri, K., Nouioui, I., Ktari, A., Simpson, S., et al. (2019). Draft genome sequence of the symbiotic *Frankia* sp. strain BMG5.30 isolated from root nodules of *Coriaria myrtifolia* in Tunisia. *Antonie Van Leeuwenhoek*, 112, 67–74.
- Gutleben, J., Chaib De Mares, M., Van Elsas, J. D., Smidt, H., Overmann, J., and Sipkema, D. (2018). The multi-omics promise in context: from sequence to microbial isolate. *Crit. Rev. Microbiol.* 44, 212–229. doi: 10.1080/1040841X.2017.1332003
- Herrera-Belaroussi, A., Normand, P., Pawlowski, K., Fernandez, M. P., Wibberg, D., Kalinowski, J., et al. (2020). Candidatus *Frankia nodulisporulans* sp. nov., an *Ahnus glutinosa*-infective *Frankia* species unable to grow in pure culture and able to sporulate in-planta. *Syst. Appl. Microbiol.* 43:126134. doi: 10.1016/j.syapm.2020.126134
- Huang, Y., and Benson, D. R. (2012). Growth and development of *Frankia* spp. strain CcI3 at the single-hypha level in liquid culture. *Arch. Microbiol.* 194, 21–28. doi: 10.1007/s00203-011-0734-5
- Huang, L., Zhang, H., Wu, P., Entwistle, S., Li, X., Yohe, T., et al. (2018). dbCAN-seq: a database of carbohydrate-active enzyme (CAZyme) sequence and annotation. *Nucleic Acids Res.* 46, D516–D521. doi: 10.1093/nar/gkx894
- Huss-Danell, K. (1997). Tansley review no. 93. Actinorhizal symbioses and their N<sub>2</sub> fixation. *New Phytol.* 136, 375–405. doi: 10.1046/j.1469-8137.1997.00755.x
- Hyatt, D., Chen, G. L., LoCasio, P. F., Land, M. L., Larimer, F. W., and Hauser, L. J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11, 1–11. doi: 10.1186/1471-2105-11-119
- Kanehisa, M., Goto, S., Kawashima, S., and Nakaya, A. (2002). The KEGG databases at GenomeNet. *Nucleic Acids Res.* 30, 42–46. doi: 10.1093/nar/30.1.42
- Kanehisa, M., Sato, Y., and Morishima, K. (2016). BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J. Mol. Biol.* 428, 726–731. doi: 10.1016/j.jmb.2015.11.006
- Kapinusova, G., Lopez Marin, M. A., and Uhlik, O. (2023). Reaching unreachables: obstacles and successes of microbial cultivation and their reasons. *Front. Microbiol.* 14:1089630. doi: 10.3389/fmicb.2023.1089630
- Karaoz, U., and Brodie, E. L. (2022). microTrait: a toolset for a trait-based representation of microbial genomes. *Front. Bioinform.* 2:918853. doi: 10.3389/fbinf.2022.918853
- Kell, D. B., and Young, M. (2000). Bacterial dormancy and culturability: the role of autocrine growth factors. *Curr. Opin. Microbiol.* 3, 238–243. doi: 10.1016/S1369-5274(00)00082-5
- Kim, Y., Gu, C., Kim, H. U., and Lee, S. Y. (2020). Current status of pan-genome analysis for pathogenic bacteria. *Curr. Opin. Biotechnol.* 63, 54–62. doi: 10.1016/j.copbio.2019.12.001
- Kramer, J., Özkaya, Ö., and Kümmerli, R. (2020). Bacterial siderophores in community and host interactions. *Nat. Rev. Microbiol.* 18, 152–163. doi: 10.1038/s41579-019-0284-4
- Kuo, C. H., Moran, N. A., and Ochman, H. (2009). The consequences of genetic drift for bacterial genome complexity. *Genome Res.* 19, 1450–1454. doi: 10.1101/gr.091785.109
- Kuo, C. H., and Ochman, H. (2010). The extinction dynamics of bacterial pseudogenes. *PLoS Genet.* 6:e1001050. doi: 10.1371/journal.pgen.1001050
- Lagier, J. C., Armougom, F., Million, M., Hugon, P., Pagnier, I., Robert, C., et al. (2012). Microbial culturomics: paradigm shift in the human gut microbiome study. *Clin. Microbiol. Infect.* 18, 1185–1193. doi: 10.1111/1469-0691.12023
- Lajoie, G., and Kembel, S. W. (2019). Making the most of trait-based approaches for microbial ecology. *Trends Microbiol.* 27, 814–823. doi: 10.1016/j.tim.2019.06.003
- Lajoie, G., and Parfrey, L. W. (2022). Beyond specialization: re-examining routes of host influence on symbiont evolution. *Trends Ecol. Evol.* 37, 590–598. doi: 10.1016/j.tree.2022.03.006
- Laugier, J. (2023). The “comfort timing” strategy: a potential pathway for the cultivation of uncultured microorganisms and a possible adaptation for environmental colonisation. *FEMS Microbiol. Ecol.* 99:fiad026. doi: 10.1093/femsec/fiad026
- Lewis, W. H., and Ettema, T. J. (2019). Culturing the uncultured. *Nat. Biotechnol.* 37, 1278–1279. doi: 10.1038/s41587-019-0300-2
- Lewis, W. H., Tahon, G., Geesink, P., Sousa, D. Z., and Ettema, T. J. (2021). Innovations to culturing the uncultured microbial majority. *Nat. Rev. Microbiol.* 19, 225–240. doi: 10.1038/s41579-020-00458-8
- Li, T., and Yin, Y. (2022). Critical assessment of pan-genomic analysis of metagenome-assembled genomes. *Brief. Bioinform.* 23, 1–12. doi: 10.1093/bib/bbac413
- Liu, S., Moon, C. D., Zheng, N., Huws, S., Zhao, S., and Wang, J. (2022). Opportunities and challenges of using metagenomic data to bring uncultured microbes into cultivation. *Microbiome* 10:76. doi: 10.1186/s40168-022-01272-5
- Lo, W. S., Huang, Y. Y., and Kuo, C. H. (2016). Winding paths to simplicity: genome evolution in facultative insect symbionts. *FEMS Microbiol. Rev.* 40, 855–874. doi: 10.1093/femsre/fuw028
- Lopez, M. F., Young, P., and Torrey, J. G. (1986). A comparison of carbon source utilization for growth and nitrogenase activity in two *Frankia* isolates. *Can. J. Microbiol.* 32, 353–358. doi: 10.1139/m86-068
- Markova, N. (2020). Eubiotic vs. dysbiotic human blood microbiota: the phenomenon of cell wall deficiency and disease-trigger potential of bacterial and fungal L-forms. *Discov. Med.* 29, 17–26.
- Medini, D., Donati, C., Tettelin, H., Massignani, V., and Rappuoli, R. (2005). The microbial pan-genome. *Curr. Opin. Genet. Dev.* 15, 589–594. doi: 10.1016/j.gde.2005.09.006
- Megrian, D., Taib, N., Jaffe, A. L., Banfield, J. F., and Gribaldo, S. (2022). Ancient origin and constrained evolution of the division and cell wall gene cluster in Bacteria. *Nat. Microbiol.* 7, 2114–2127. doi: 10.1038/s41564-022-01257-y
- Meier-Kolthoff, J. P., Auch, A. F., Klenk, H. P., and Göker, M. (2013). Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics* 14, 1–14. doi: 10.1186/1471-2105-14-60

- Mineeva, O., Rojas-Carulla, M., Ley, R. E., Schölkopf, B., and Youngblut, N. D. (2020). DeepMAsED: evaluating the quality of metagenomic assemblies. *Bioinformatics* 36, 3011–3017. doi: 10.1093/bioinformatics/btaa124
- Moran, N. A., McCutcheon, J. P., and Nakabachi, A. (2008). Genomics and evolution of heritable bacterial symbionts. *Annu. Rev. Genet.* 42, 165–190. doi: 10.1146/annurev.genet.41.110306.130119
- Murry, M. A., Fontaine, M. S., and Torrey, J. G. (1984). Growth kinetics and nitrogenase induction in Frankia sp. HFPArI 3 grown in batch culture. *Plant Soil* 78, 61–78. doi: 10.1007/BF02277840
- Nakabachi, A., Yamashita, A., Toh, H., Ishikawa, H., Dunbar, H. E., Moran, N. A., et al. (2006). The 160-kilobase genome of the bacterial endosymbiont Carsonella. *Science* 314:267. doi: 10.1126/science.1134196
- Nawrocki, E. P., and Eddy, S. R. (2013). Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* 29, 2933–2935. doi: 10.1093/bioinformatics/btt509
- Nazaret, S., Cournoyer, B., Normand, P., and Simonet, P. (1991). Phylogenetic relationships among Frankia genomic species determined by use of amplified 16S rDNA sequences. *J. Bacteriol.* 173, 4072–4078. doi: 10.1128/jb.173.13.4072-4078.1991
- Nelson, W. C., and Stegen, J. C. (2015). The reduced genomes of Parcubacteria (OD1) contain signatures of a symbiotic lifestyle. *Front. Microbiol.* 6:713. doi: 10.3389/fmicb.2015.00713
- Nguyen, L. T., Schmidt, H. A., Von Haeseler, A., and Minh, B. Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274. doi: 10.1093/molbev/msu300
- Nguyen, T. V., Wibberg, D., Battenberg, K., Blom, J., Vanden Heuvel, B., Berry, A. M., et al. (2016). An assemblage of Frankia cluster II strains from California contains the canonical nod genes and also the sulfotransferase gene nodH. *BMC Genomics* 17, 1–17. doi: 10.1186/s12864-016-3140-1
- Nguyen, T. V., Wibberg, D., Vigil-Stenman, T., Berckx, F., Battenberg, K., Demchenko, K. N., et al. (2019). Frankia-enriched metagenomes from the earliest diverging symbiotic Frankia cluster: they come in teams. *Genome Biol. Evol.* 11, 2273–2291. doi: 10.1093/gbe/evz153
- Normand, P., Lapiere, P., Tisa, L. S., Gogarten, J. P., Alloisio, N., Bagnarol, E., et al. (2007). Genome characteristics of facultatively symbiotic Frankia sp. strains reflect host range and host plant biogeography. *Genome Res.* 17, 7–15. doi: 10.1101/gr.5798407
- Normand, P., Nguyen, T. V., Battenberg, K., Berry, A. M., Heuvel, B. V., Fernandez, M. P., et al. (2017). Proposal of Candidatus Frankia californiensis, the uncultured symbiont in nitrogen-fixing root nodules of a phylogenetically broad group of hosts endemic to western North America. *Int. J. Syst. Evol. Microbiol.* 67, 3706–3715. doi: 10.1099/ijsem.0.002147
- Nouioui, I., Ghodhbane-Gtari, F., Fernandez, M. P., Boudabous, A., Normand, P., and Gtari, M. (2014). Absence of cospeciation between the uncultured Frankia microsymbionts and the disjunct actinorhizal Coriaria species. *Biomed. Res. Int.* 2014:924235. doi: 10.1155/2014/924235
- Nouioui, I., Ghodhbane-Gtari, F., Pötter, G., Klenk, H. P., and Goodfellow, M. (2023). Novel species of Frankia, Frankia gtarii sp. nov. and Frankia tisa sp. nov., isolated from a root nodule of *Alnus glutinosa*. *Syst. Appl. Microbiol.* 46:126377. doi: 10.1016/j.syapm.2022.126377
- Nowrotek, M., Jałowiecki, T., Harnisz, M., and Plaza, G. A. (2019). Culturomics and metagenomics: In understanding of environmental resistome. *Frontiers of Environmental Science & Engineering*, 13, 1–12.
- Nussenzweig, P. M., and Marraffini, L. A. (2020). Molecular mechanisms of CRISPR-Cas immunity in bacteria. *Annu. Rev. Genet.* 54, 93–120. doi: 10.1146/annurev-genet-022120-112523
- O'Brien, J., and Wright, G. D. (2011). An ecological perspective of microbial secondary metabolism. *Curr. Opin. Biotechnol.* 22, 552–558. doi: 10.1016/j.copbio.2011.03.010
- Olson, N. D., Treangen, T. J., Hill, C. M., Cepeda-Espinoza, V., Ghurye, J., Koren, S., et al. (2019). Metagenomic assembly through the lens of validation: recent advances in assessing and improving the quality of genomes assembled from metagenomes. *Briefings in bioinformatics*, 20, 1140–1150.
- Ondov, B. D., Treangen, T. J., Melsted, P., Mallonee, A. B., Bergman, N. H., Koren, S., et al. (2016). Mash: fast genome and metagenome distance estimation using MinHash. *Genome Biol.* 17, 1–14. doi: 10.1186/s13059-016-0997-x
- Overbeek, R., Olson, R., Pusch, G. D., Olsen, G. J., Davis, J. J., Disz, T., et al. (2014). The SEED and the rapid annotation of microbial genomes using subsystems technology (RAST). *Nucleic Acids Res.* 42, D206–D214. doi: 10.1093/nar/gkt1226
- Pande, S., and Kost, C. (2017). Bacterial unculturability and the formation of intercellular metabolic networks. *Trends Microbiol.* 25, 349–361. doi: 10.1016/j.tim.2017.02.015
- Perry, E. K., Meirelles, L. A., and Newman, D. K. (2022). From the soil to the clinic: the impact of microbial secondary metabolites on antibiotic tolerance and resistance. *Nat. Rev. Microbiol.* 20, 129–142. doi: 10.1038/s41579-021-00620-w
- Persson, T., Battenberg, K., Demina, I. V., Vigil-Stenman, T., Vanden Heuvel, B., Pujic, P., et al. (2015). Candidatus Frankia daticae Dg1, the actinobacterial microsymbiont of *Datisca glomerata*, expresses the canonical nod genes nodABC in symbiosis with its host plant. *PLoS One* 10:e0127630. doi: 10.1371/journal.pone.0127630
- Persson, T., Benson, D. R., Normand, P., Vanden Heuvel, B., Pujic, P., Chertkov, O., et al. (2011). Genome sequence of “Candidatus Frankia daticae” Dg1, the uncultured microsymbiont from nitrogen-fixing root nodules of the dicot *Datisca glomerata*. *J. Bacteriol.* 193, 7017–7018. doi: 10.1128/JB.06208-11
- Pham, V. H., and Kim, J. (2016). Improvement for isolation of soil bacteria by using common culture media. *J. Pure Appl. Microbiol.* 10, 49–59.
- Pinto-Carbó, M., Sieber, S., Dessein, S., Wicker, T., Verstraete, B., Gademann, K., et al. (2016). Evidence of horizontal gene transfer between obligate leaf nodule symbionts. *ISME J.* 10, 2092–2105. doi: 10.1038/ismej.2016.27
- Potter, S. C., Luciani, A., Eddy, S. R., Park, Y., Lopez, R., and Finn, R. D. (2018). HMMER web server: 2018 update. *Nucleic Acids Res.* 46, W200–W204. doi: 10.1093/nar/gky448
- Pozzi, A. C. M., Herrera-Belaroussi, A., Schwob, G., Bautista-Guerrero, H. H., Bethencourt, L., Fournier, P., et al. (2020). Proposal of ‘Candidatus Frankia alpina’, the uncultured symbiont of *Alnus alnobetula* and *A. incana* that forms spore-containing nitrogen-fixing root nodules. *Int. J. Syst. Evol. Microbiol.* 70, 5453–5459. doi: 10.1099/ijsem.0.004433
- Price, M. N., Deuschbauer, A. M., and Arkin, A. P. (2020). GapMind: automated annotation of amino acid biosynthesis. *MSystems* 5, 10–1128. doi: 10.1128/mSystems.00291-20
- Price, M. N., Deuschbauer, A. M., and Arkin, A. P. (2021). GapMind for carbon sources: automated annotations of catabolic pathways. *bioRxiv*, 2021-11. doi: 10.1371/journal.pgen.1010156
- Prozorov, A. A., and Danilenko, V. N. (2010). Toxin-antitoxin systems in bacteria: apoptotic tools or metabolic regulators?. *Microbiology*, 79, 129–140.
- Quince, C., Walker, A. W., Simpson, J. T., Loman, N. J., and Segata, N. (2017). Shotgun metagenomics, from sampling to analysis. *Nat. Biotechnol.* 35, 833–844. doi: 10.1038/nbt.3935
- Quinlan, A. R. (2014). BEDTools: the Swiss-army tool for genome feature analysis. *Curr. Protoc. Bioinformatics* 47, 11–12. doi: 10.1002/0471250953.bi1112s47
- Ray, K., Marteyn, B., Sansonetti, P. J., and Tang, C. M. (2009). Life on the inside: the intracellular lifestyle of cytosolic bacteria. *Nat. Rev. Microbiol.* 7, 333–340. doi: 10.1038/nrmicro2112
- Riley, M. A., and Wertz, J. E. (2002). Bacteriocins: evolution, ecology, and application. *Annu. Rev. Microbiol.* 56, 117–137. doi: 10.1146/annurev.micro.56.012302.161024
- Ringø, E., Clausen, E., Løvaas, E., Van Ghelue, M., and Solheim, B. (1995). Effects of extracts of *Alnus glutinosa* seeds on growth of Frankia strain ArI3 under static and fermentor culture conditions. *Plant Soil* 176, 283–288. doi: 10.1007/BF00011792
- Riva, V., Mapelli, F., Bagnasco, A., Mengoni, A., and Borin, S. (2022). A meta-analysis approach to defining the culturable core of plant endophytic bacterial communities. *Appl. Environ. Microbiol.* 88, e02537–e02521. doi: 10.1128/aem.02537-21
- Rodríguez-Gijón, A., Nuy, J. K., Mehrshad, M., Buck, M., Schulz, F., Woyke, T., et al. (2022). A genomic perspective across Earth's microbiomes reveals that genome size in Archaea and Bacteria is linked to ecosystem type and trophic strategy. *Front. Microbiol.* 12:761869. doi: 10.3389/fmicb.2021.761869
- Salje, J. (2021). Cells within cells: Rickettsiales and the obligate intracellular bacterial lifestyle. *Nat. Rev. Microbiol.* 19, 375–390. doi: 10.1038/s41579-020-00507-2
- Sauer, D. B., and Wang, D. N. (2019). Predicting the optimal growth temperatures of prokaryotes using only genome derived features. *Bioinformatics* 35, 3224–3231. doi: 10.1093/bioinformatics/btz059
- Sayed, W. F., Wheeler, C. T., Zahran, H. H., and Shoreit, A. A. M. (1997). Effect of temperature and soil moisture on the survival and symbiotic effectiveness of Frankia spp. *Biol. Fertil. Soils* 25, 349–353. doi: 10.1007/s003740050324
- Schattner, P., Brooks, A. N., and Lowe, T. M. (2005). The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic acids research*, 33, W686–W689.
- Schultz, J., Modolon, F., Peixoto, R. S., and Rosado, A. S. (2023). Shedding light on the composition of extreme microbial dark matter: alternative approaches for culturing extremophiles. *Front. Microbiol.* 14:1167718. doi: 10.3389/fmicb.2023.1167718
- Schwartzman, J. A., and Ruby, E. G. (2016). Stress as a normal cue in the symbiotic environment. *Trends Microbiol.* 24, 414–424. doi: 10.1016/j.tim.2016.02.012
- Schwintzer, C. R. (1990). Spore-positive and spore-negative nodules. In: *The biology of Frankia and Actinorhizal plants* Christa R. Schwintzer and John D. Tjepkema (eds). New York: Academic Press: 177–193.
- Schwob, G., Roy, M., Pozzi, A. C., Herrera-Belaroussi, A., and Fernandez, M. P. (2018). In planta sporulation of Frankia spp. as a determinant of alder-symbiont interactions. *Appl. Environ. Microbiol.* 84, e01737–e01718. doi: 10.1128/AEM.01737-18
- Shade, A., Jones, S. E., Caporaso, J. G., Handelsman, J., Knight, R., Fierer, N., et al. (2014). Conditionally rare taxa disproportionately contribute to temporal changes in microbial diversity. *MBio* 5, 10–1128. doi: 10.1128/mBio.01371-14

- Shipton, W. A., and Burggraaf, A. J. P. (1982). A comparison of the requirements for various carbon and nitrogen sources and vitamins in some Frankia isolates. *Plant Soil* 69, 149–161. doi: 10.1007/BF02374510
- Siguier, P., Filée, J., and Chandler, M. (2006). Insertion sequences in prokaryotic genomes. *Curr. Opin. Microbiol.* 9, 526–531. doi: 10.1016/j.mib.2006.08.005
- Siguier, P., Gourbeyre, E., and Chandler, M. (2014). Bacterial insertion sequences: their genomic impact and diversity. *FEMS Microbiol. Rev.* 38, 865–891. doi: 10.1111/1574-6976.12067
- Simonet, P., Bosco, M., Chapelon, C., Moiroud, A., and Normand, P. (1994). Molecular characterization of Frankia microsymbionts from spore-positive and spore-negative nodules in a natural alder stand. *Appl. Environ. Microbiol.* 60, 1335–1341. doi: 10.1128/aem.60.4.1335-1341.1994
- Sloan, D. B., and Moran, N. A. (2013). The evolution of genomic instability in the obligate endosymbionts of whiteflies. *Genome Biol. Evol.* 5, 783–793. doi: 10.1093/gbe/evt044
- Spielman, S. J., and Wilke, C. O. (2015). The relationship between dN/dS and scaled selection coefficients. *Mol. Biol. Evol.* 32, 1097–1108. doi: 10.1093/molbev/msv003
- Stein, L. (2001). Genome annotation: from sequence to biology. *Nat. Rev. Genet.* 2, 493–503. doi: 10.1038/35080529
- Stewart, E. J. (2012). Growing unculturable bacteria. *J. Bacteriol.* 194, 4151–4160. doi: 10.1128/JB.00345-12
- Stowers, M. D., Kulkarni, R. K., and Steele, D. B. (1986). Intermediary carbon metabolism in Frankia. *Arch. Microbiol.* 143, 319–324. doi: 10.1007/BF00412796
- Sun, J., Lu, F., Luo, Y., Bie, L., Xu, L., and Wang, Y. (2023). OrthoVenn3: an integrated platform for exploring and visualizing orthologous data across genomes. *Nucleic Acids Res.* 51, W397–W403. doi: 10.1093/nar/gkad313
- Syberg-Olsen, M. J., Garber, A. I., Keeling, P. J., McCutcheon, J. P., and Husnik, F. (2022). Pseudofinder: detection of pseudogenes in prokaryotic genomes. *Molecular biology and evolution. Nature Reviews Microbiology*, 19, 225–240.
- Sarkar, I., and Sen, A. (2022). Niche adaptation of Frankia do not drastically influence their metabolic profiling. *Journal of Forest Research*, 27, 100–105.
- Tam, V., Patel, N., Turcotte, M., Bossé, Y., Paré, G., and Meyre, D. (2019). Benefits and limitations of genome-wide association studies. *Nat. Rev. Genet.* 20, 467–484. doi: 10.1038/s41576-019-0127-1
- Tettelin, H., Massignani, V., Cieslewicz, M. J., Donati, C., Medini, D., Ward, N. L., et al. (2005). Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial “pan-genome”. *Proc. Natl. Acad. Sci.* 102, 13950–13955. doi: 10.1073/pnas.0506758102
- Tisa, L. S., Oshone, R., Sarkar, I., Ktari, A., Sen, A., and Gtari, M. (2016). Genomic approaches toward understanding the actinorhizal symbiosis: an update on the status of the Frankia genomes. *Symbiosis* 70, 5–16. doi: 10.1007/s13199-016-0390-2
- Tonkin-Hill, G., Gladstone, R. A., Pöntinen, A. K., Arredondo-Alonso, S., Bentley, S. D., and Corander, J. (2023). Robust analysis of prokaryotic pangenome gene gain and loss rates with Panstripe. *Genome Res.* 33, 129–140. doi: 10.1101/gr.277340.122
- Tonkin-Hill, G., MacAlasdair, N., Ruis, C., Weimann, A., Horesh, G., Lees, J. A., et al. (2020). Producing polished prokaryotic pangenomes with the Panaroo pipeline. *Genome Biol.* 21, 1–21. doi: 10.1186/s13059-020-02090-4
- Touchon, M., and Rocha, E. P. (2007). Causes of insertion sequences abundance in prokaryotic genomes. *Mol. Biol. Evol.* 24, 969–981. doi: 10.1093/molbev/msm014
- Van Teeseling, M. C., Mesman, R. J., Kuru, E., Espallat, A., Cava, F., Brun, Y. V., et al. (2015). Anammox Planctomycetes have a peptidoglycan cell wall. *Nat. Commun.* 6:6878. doi: 10.1038/ncomms7878
- Vieira-Silva, S., and Rocha, E. P. (2010). The systemic imprint of growth and its uses in ecological (Meta)genomics. *PLoS Genet.* 6:e1000808. doi: 10.1371/journal.pgen.1000808
- Vartoukian, S. R., Palmer, R. M., and Wade, W. G. (2010). Strategies for culture of ‘unculturable’ bacteria. *FEMS microbiology letters*, 309, 1–7.
- Weissman, J. L., Hou, S., and Fuhrman, J. A. (2021). Estimating maximal microbial growth rates from cultures, metagenomes, and single cells via codon usage patterns. *Proc. Natl. Acad. Sci. USA* 118:e2016810118. doi: 10.1073/pnas.2016810118
- Wernegreen, J. J. (2015). Endosymbiont evolution: predictions from theory and surprises from genomes. *Ann. N. Y. Acad. Sci.* 1360, 16–35. doi: 10.1111/nyas.12740
- Wheeler, C. T., Akkermans, A. D. L., and Berry, A. M. (2008). “Frankia and actinorhizal plants: a historical perspective” in Nitrogen-fixing actinorhizal symbioses (Dordrecht: Springer Netherlands), 1–24.
- Wilcox, J. L., Dunbar, H. E., Wolfinger, R. D., and Moran, N. A. (2003). Consequences of reductive evolution for gene expression in an obligate endosymbiont. *Mol. Microbiol.* 48, 1491–1500. doi: 10.1046/j.1365-2958.2003.03522.x
- Wörmer, L., Hoshino, T., Bowles, M. W., Viehweger, B., Adhikari, R. R., Xiao, N., et al. (2019). Microbial dormancy in the marine subsurface: global endospore abundance and response to burial. *Science. Advances* 5:eaav1024. doi: 10.1126/sciadv.aav1024
- Xie, J., Chen, Y., Cai, G., Cai, R., Hu, Z., and Wang, H. (2023). Tree visualization by one Table (tvBOT): a web application for visualizing, modifying and annotating phylogenetic trees. *Nucleic Acids Res.* 51, W587–W592. doi: 10.1093/nar/gkad359
- Xie, B. B., Li, M., Anantharaman, K., and Ravin, N. V. (2021). The uncultured microorganisms: novel technologies and applications. *Front. Microbiol.* 12:756287. doi: 10.3389/fmicb.2021.756287
- Yan, C., Owen, J. S., Seo, E. Y., Jung, D., and He, S. (2023). Microbial interaction is among the key factors for isolation of previous uncultured microbes. *J. Microbiol.* 61, 655–662. doi: 10.1007/s12275-023-00063-3
- Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* 24, 1586–1591. doi: 10.1093/molbev/msm088
- Yoon, S. H., Ha, S. M., Lim, J., Kwon, S., and Chun, J. (2017). A large-scale evaluation of algorithms to calculate average nucleotide identity. *Antonie Van Leeuwenhoek* 110, 1281–1286. doi: 10.1007/s10482-017-0844-4
- Zhang, Y., and Sievert, S. M. (2014). Pan-genome analyses identify lineage- and nichespecific markers of evolution and adaptation in Epsilonproteobacteria. *Front. Microbiol.* 5:110. doi: 10.3389/fmicb.2014.00110
- Zhao, Y., Feng, Y., Zhou, J., Zhang, K., Sun, J., Wang, L., et al. (2023). Potential bacterial isolation by dosing metabolites in cross-feedings. *Water Res.* 231:119589. doi: 10.1016/j.watres.2023.119589
- Zhongze, Z., Murry, M. A., and Torrey, J. G. (1986). Culture conditions influencing growth and nitrogen fixation in Frankia sp. HFPC cI3 isolated from Casuarina. *Plant Soil* 91, 3–15. doi: 10.1007/BF02181814