



OPEN ACCESS

EDITED BY

Muneer Ahmad Malla,
German Centre for Integrative Biodiversity
Research (iDiv), Germany

REVIEWED BY

Satheesh Sathianeson,
King Abdulaziz University, Saudi Arabia
Liu Dong,
Anqing Vocational and Technical College,
China

*CORRESPONDENCE

Natalia Rutkowska
✉ natalia.rutkowska@dokt.p.lodz.pl

RECEIVED 16 April 2025

ACCEPTED 03 June 2025

PUBLISHED 30 June 2025

CITATION

Rutkowska N, Daroch M and
Marchut-Mikołajczyk O (2025) Exploring
the diversity and genomics of cultivable
Bacillus-related endophytic bacteria from
the medicinal plant *Galium aparine* L.
Front. Microbiol. 16:1612860.
doi: 10.3389/fmicb.2025.1612860

COPYRIGHT

© 2025 Rutkowska, Daroch and
Marchut-Mikołajczyk. This is an open-access
article distributed under the terms of the
[Creative Commons Attribution License](#)
(CC BY). 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.

Exploring the diversity and genomics of cultivable *Bacillus*-related endophytic bacteria from the medicinal plant *Galium aparine* L.

Natalia Rutkowska^{1*}, Maurycy Daroch² and
Olga Marchut-Mikołajczyk¹

¹Institute of Molecular and Industrial Biotechnology, Lodz University of Technology, Łódź, Poland,

²School of Environment and Energy, Peking University Shenzhen Graduate School, Shenzhen, China

Introduction: Endophytes are crucial partners that contribute to the plants' health and overall wellbeing. Apart from the elucidation of the relationship between plants and their microbiota, the metabolic potential of endophytes is also of a special interest. Therefore, it is crucial to isolate and taxonomically identify endophytes, as well as to investigate their genomic potential to determine their significance in plant health and potential as bioactive metabolite producers for industrial application.

Methods: In this study, we isolated ten endophytic bacterial strains from different tissues of medicinal plant *Galium aparine* L. and performed *de novo* assembly of their genomes using short and long reads. Comparative genomic analysis was conducted to assess the accurate taxonomic identification of the strains. The investigation also focused on the presence of mobile genetic elements and their significance concerning endophytic lifestyles. We performed functional annotation of coding sequences, particularly targeted genes that encode carbohydrate enzymes and secondary metabolites within gene clusters.

Results: Through sequencing using two complementary methods, we obtained 10 bacterial genomes, ranging in size, coding density and number of mobile genetic elements. Our findings provide a first insight into the cultivable bacterial community of the medicinal plant *Galium aparine* L., their genome biology, and potential for producing valuable bioactive metabolites. Obtained whole genome sequences allowed for complete phylogenetic analysis, which revealed crucial insights into the taxonomic status of bacteria and resulted in the discovery of two putatively novel bacterial species from the *Bacillus* and *Priestia* genera, suggesting that plants are hiding a reservoir of novel species with potentially useful properties and unknown mechanisms related to their relationship with plant host.

KEYWORDS

bacterial endophytes, *Galium aparine* L., phylogenetics, genome mining, mobile genetic elements, secondary metabolites, plant microbiome

Introduction

Plants have developed complex and diverse relationships with microorganisms in their neighborhood as they co-exist in the same ecological niche, constantly subjected to selection pressures. Endophytic microorganisms are an exceptional group of plant-associated microorganisms, as they thrive inside their tissues without causing any detrimental effect or disease to the host (Hardoim et al., 2015). The relationship between plants and their endophytes is mutually beneficial. The plant provides a safe niche less disturbed by fluctuating conditions (e.g., temperature amplitude) and a constant supply of nutrients, while the endophytes provide positive effects on plant health and fitness, therefore ensuring their own survival (Afzal et al., 2019). Among the beneficial traits exhibited by endophytes, the most common ones are nitrogen fixation and phosphorus solubilization, which enhance the uptake of these crucial nutrients by a plant. Additionally, endophytes produce various bioactive substances such as plant hormones (indole-3-acetic acid, gibberellins, abscisic acid, etc.) siderophores, antibiotics, and insecticides (Afzal et al., 2019; Egamberdieva et al., 2017; Eid et al., 2021). Medicinal plants are particularly intriguing for the identification of their endophytes, especially bacteria, due to the abundant production of various chemicals with therapeutic qualities. However, the investigation of these bacterial endophytes has been limited (Sharma et al., 2023). Existing evidence proves that endophytes are capable of producing *ex planta* bioactive compounds that are identical or very similar to those of their host plants; therefore, this concept may 1 day be applied as technology of their large-scale production (Kusari and Spiteller, 2011; Rutkowska et al., 2023).

Abbreviations: AA, auxiliary activity; AMR, antimicrobial resistance; ANI, Average Nucleotide Identity; antiSMASH, antibiotics and secondary metabolite analysis shell; ARGs, antibiotic resistance genes; ARO, antibiotic resistance ontology; BGCs, biosynthetic gene clusters; BUSCO, Benchmarking Universal Single-Copy Orthologs; CARD, Comprehensive Antibiotic Resistance Database; Cas, CRISPR-associated; CAZymes, carbohydrate-active enzymes; CBM, carbohydrate-binding module; CDPS, tRNA-dependent cyclodipeptide synthase; CDS – coding sequence; CE, carbohydrate esterase; CRISPR, clustered regularly interspaced short palindromic repeats; CSIs, conserved signature indels; dDDH, digital DNA:DNA hybridization; GH, glycoside hydrolase; GlcNAc, N-acetyl-D-glucosamine; GT, glycosyltransferase; HGT, horizontal gene transfer; HMMs, Hidden Markov Models; IAA, indole-3-acetic acid; IAald, indole-3-acetaldehyde; ICEs, integrative and conjugative elements; IE, integration/excision; IPA, indole pyruvate; LAP, linear azol(in)e-containing peptides; LPSN, List of Prokaryotic Names with Standing in Nomenclature; MGEs, mobile genetic elements; MIBiG, Minimum Information about a Biosynthetic Gene Cluster; MLST, multi-locus sequence typing; mobileOG-db, mobile orthologous groups database; MRSA, methicillin-resistant *Staphylococcus aureus*; MS, mass spectroscopy; NCBI, National Center for Biotechnology Information; ncRNA, non-coding RNA; NMR, nuclear magnetic resonance spectroscopy; NRPS, non-ribosomal peptide synthetases; ONT, Oxford Nanopore Technologies; OSMAC, One Strain Many Compounds; P, phage; PGAP, Prokaryotic Genome Annotation Pipeline; PGPB, plant growth-promoting bacteria; PHASTER, PHAge Search Enhanced Release; PKS, polyketide synthases; PL, polysaccharide lyase; QUASt, Quality ASsessment Tool for Genome ASsemblies; RAST, The Rapid Annotation using Subsystem Technology; REALPHY, the Reference sequence Alignment based Phylogeny builder; RGI, Resistance Gene Identifier; RiPPs, ribosomally synthesized and post-translationally modified peptides; rRNA, ribosomal RNA; RRR, replication/recombination/repair; SMR, small multidrug resistance; SNPs, Single Nucleotide Polymorphisms; STD, stability/transfer/defense; T, transfer; TFBS, transcription factor binding sequences; tRNA, transfer RNA; TYGS, Type (Strain) Genome Server; WHO, World Health Organization.

A broad group of endophytes, classified as “facultative,” enters the plant via root or stem cracks, previously being rhizospheric or epiphytic microorganisms, respectively. Another type, known as “obligate,” maintains an intrinsic bond with its host throughout its entire life cycle and typically transmits itself vertically through seeds (Rosenblueth and Martínez-Romero, 2006). Seed microbiota is transmitted by plants across generations, being the starting point for community assembly in the new seedling (Semenzato et al., 2022a; Truyens et al., 2015). Obligate endophytes are recalcitrant or even impossible to grow outside plants in laboratory conditions; thus, cultivable endophytes are considered to be mostly facultative (Hardoim et al., 2008; Papik et al., 2020). Members of the *Bacillaceae* family, especially *Bacillus* species, are among the most commonly encountered cultivable endophytic bacteria. Their plant growth-promoting activities and versatility for industrial use have led to extensive research in recent decades (Jasim et al., 2016; Lopes et al., 2018; Marchut-Mikolajczyk et al., 2018, 2020; Marchut-Mikolajczyk et al., 2023; Zhou et al., 2021).

Apart from *ex planta* studies of isolated endophytes, modern omics approaches conjoining different aspects of single endophyte interaction with the host plant and other inhabitants are crucial to deciphering not only their role in host plant ecology, but also the impact of it on endophytes' functioning. As the plant itself represents a peculiar niche that requires specific adaptations from its inhabitants, endophytes influence the plant environment and contribute to its secondary metabolism as well. Kaul et al. (2016) broadly discussed methodologies essential for resolving some of the issues, including genomics, transcriptomics, and proteomics. Comparative genomics provides insights into mechanisms steering the endophytic lifestyle, such as colonization patterns, plant growth promotion, induced resistance, interactions with other associated microorganisms, as well as the secondary metabolism of an endophyte (Kaul et al., 2016; Pinski et al., 2019). The conventional approach to screening for novel compounds relies on biological assays based on their putative activity; however, this approach is time- and cost-consuming and often results in rediscovery of known compounds (Kenshole et al., 2021). Therefore, in the last years, there has been a shift toward a genome mining approach, which takes advantage of recent advancements in DNA sequencing technologies and bioinformatics. This reveals the hidden potential of cultivable endophytes to produce specific compounds (enzymes, non-ribosomal peptides, polyketides), which are often not expressed under standard laboratory conditions. This knowledge could guide future cultivation methods and approaches to activate genes responsible for their synthesis (Kenshole et al., 2021). All of those endophytic bioactive compounds can be successfully used in pharmaceutical and other industrial applications, as detergents (Marchut-Mikolajczyk et al., 2018), biofuels (Zhang et al., 2014), biopesticides (Rong et al., 2020) or pigments (Hagaggi and Abdul-Raouf, 2023), as well as in agriculture and crop production (Eid et al., 2021; Lopes et al., 2018).

To our knowledge, this is the first report on endophytic bacteria isolated from *Galium aparine* L., a medicinal herb widely growing in Europe, North America and Asia; known for its beneficial effects on kidneys, skin disorders, wounds, high blood pressure, and insomnia (Al-Snafi, 2018; Ilina et al., 2019). While there is no prior published record on endophytes, both bacterial and fungal, a research by Bemmam and Kulle (1988) reported the presence

TABLE 1 Genome assembly statistics and annotation features of endophytic bacteria from *Galium aparine* L.

Feature	Strain name									
	GR1	GR2	GR3	GR4	GS1	GS2	GS3	GL1	GL2	GL3
Size (Mbp)	5.66	5.63	5.97	5.69	5.94	4.17	5.70	6.10	5.70	5.99
GC content (%)	35.23	40.48	34.87	37.70	34.86	37.69	35.05	35.03	35.12	35.11
Contigs	18	1	10	45	58	9	26	19	21	8
N50 (Mbp)	4.99	5.63	3.13	0.91	0.61	3.77	2.92	2.13	3.78	5.29
L50	1	1	1	3	4	1	1	2	1	1
Genes (all)	5,840	5,392	6,085	5,937	6,061	4,285	5,803	6,307	5,808	6,132
CDSs*	5,716	5,261	5,941	5,842	5,986	4,157	5,676	6,173	5,659	5,982
rRNAs*	19	42	34	11	10	27	28	27	38	39
tRNAs*	100	83	105	76	60	96	94	102	106	106
ncRNAs*	5	6	5	8	5	5	5	5	5	5
Completeness (%)	99.41	99.35	99.11	99.27	99.27	99.35	99.22	98.24	99.22	98.82
Contamination (%)	0.00	0.32	0.06	0.65	0.97	0.00	0.43	0.10	0.43	0.25
Complete BUSCOs (%)	99.8	99.8	99.5	100.0	99.8	99.6	99.8	99.3	99.8	99.8

*CDS, coding sequence; ncRNA, non-coding RNA; rRNA, ribosomal RNA, tRNA, transfer RNA.

of rhizospheric fungi belonging to the genera *Aspergillus*, *Mucor*, *Penicillium*, and *Rhizopus* (Bemmann and Kulle, 1988). All plants developed close associations with surrounding microorganisms, as their support is crucial for plants as sessile organisms to adapt effectively to changing environmental conditions. However, many plant have yet to be studied in terms of their microbiome (Rutkowska et al., 2023). The study aimed to (1) isolate cultivable endophytic bacteria from the tissues of *Galium aparine* L., (2) obtain their genomic sequences and identify them taxonomically, (3) study the presence of mobile genetic elements as drivers of evolution and (4) mine their genomes for encoded bioactive compounds with beneficial uses in agriculture and other industries.

Results and discussion

Isolation of endophytic bacteria

From the surface-sterilized fragments of leaves, stems and roots of *Galium aparine* L., a total of 10 bacterial isolates were isolated and purified by subsequent streak plating. The effectiveness of the surface sterilization method was confirmed by the lack of microbial growth on the control plates containing water from the plant's last rinsing after 7 days of incubation. Obtained strains were named as G followed by R, S, or L for roots, stem and leaves, respectively.

Genome sequencing and assembly

Whole genome sequencing (WGS) was performed, followed by *de novo* hybrid assembly and annotation (Table 1). The assembly size of 10 isolated strains ranges from 4.17 (GS2) to 6.10 Mbp (GL1). The average genome-wide GC content varies between 34.9% and 40.5%. The number of assembled contigs varied from sample to sample, but of all the datasets, only the GR2 genome is assembled to

the complete chromosome level. The strains differ in the number of genes. The obtained results ensure good quality assembly without contamination.

The genome sequences are deposited in the National Center for Biotechnology Information (NCBI, United States) database under the BioProject number PRJNA1068863.

Taxonomic identification

The initial attempt at identification was conducted by manually extracting 16S rDNA and *gyrA* sequences from annotated genomes, followed by their comparison against the NCBI database using BLASTN (data not shown). However, for most of the strains, this approach was highly inconclusive as the percentage of identity in gene sequences was the same for several species, a common occurrence within the *Bacillus* genus (Liu et al., 2022). What is more, many members of *Bacillus cereus* group members are identified solely based on their 16S rDNA sequences, leading to further misidentifications in the NCBI database. As a consequence, taxonomic identification of isolated strains was based on whole genome sequence analysis using Type (Strain) Genome Server (TYGS), which calculates digital DNA:DNA hybridization (dDDH) parameter values for *in silico* species delineation, considering only verified type strains from the List of Prokaryotic names with Standing in Nomenclature (LPSN) (Meier-Kolthoff et al., 2022; Tindall et al., 2010). So-called "reference" or "representative" strains are not reliable and their use in identification may lead to further taxonomic misidentification (Meier-Kolthoff et al., 2022). Also, the Average Nucleotide Identity (ANI) values, the second parameter for species delineation, were calculated. The proposed identification of the isolates with ANI and dDDH is presented in Table 2.

The majority of the isolates surpassed thresholds recognized as the cut-offs for species delineation, $\geq 95\%$ – 96% for ANI and $\geq 70\%$ for dDDH (Auch et al., 2010; Richter and Rosselló-Móra, 2009). As stated in Table 2, most of the isolated species belong to the *Bacillus*

TABLE 2 Whole genome sequencing (WGS)-based identification of isolated strains.

Strain	Proposed identification (NCBI GenBank Assembly Accession no.)	Closest type-strain match (NCBI GenBank Assembly Accession no.)	ANI (%)*	dDDH (%)*
GR1	<i>Bacillus pretiosus</i> GR1 (GCA_036408975.1)	<i>Bacillus pretiosus</i> SAIUCEU11T(GCA_025916425.1)	97.14	78.20
GR2	<i>Peribacillus frigiditolerans</i> GR2 (GCA_036352075.1)	<i>Peribacillus frigiditolerans</i> DSM 8801 (GCA_024169475.1)	97.24	80.30
GR3	<i>Bacillus cereus</i> GR3 (GCA_036408805.1)	<i>Bacillus cereus</i> ATCC 14579 (GCA_006094295.1)	97.16	74.10
GR4	<i>Priestia megaterium</i> GR4 (GCA_036350305.1)	<i>Priestia megaterium</i> ATCC 14581 (GCA_017086525.1)	96.91	72.80
GS1	<i>Bacillus thuringiensis</i> GS1 (GCA_036350635.1)	<i>Bacillus thuringiensis</i> ATCC 10792 (GCA_000161615.1)	96.52	69.10
GS2	<i>Priestia</i> sp. GS2 (GCA_036409025.1)	<i>Priestia flexa</i> NBRC 15715 (GCA_001591565.1)	92.40	47.40
GS3	<i>Bacillus cereus</i> GS3 (GCA_036408815.1)	<i>Bacillus cereus</i> ATCC 14579 (GCA_006094295.1)	98.30	83.40
GL1	<i>Bacillus</i> sp. GL1 (GCA_039680825.1)	<i>Bacillus wiedmannii</i> FSL W8-0169 (GCA_001583695.1)	95.06	59.80
GL2	<i>Bacillus cereus</i> GL2 (GCA_036408835.1)	<i>Bacillus cereus</i> ATCC 14579 (GCA_006094295.1)	98.30	83.50
GL3	<i>Bacillus wiedmannii</i> GL3 (GCA_036350675.1)	<i>Bacillus wiedmannii</i> FSL W8-0169 (GCA_001583695.1)	96.63	70.00

*ANI, Average Nucleotide Identity; dDDH, digital DNA-DNA hybridization.

genus or are closely related. Strains can be divided into three groups (Figure 1): (1) *Priestia* spp. – *Priestia* sp. GS2, *Priestia megaterium* GR4, (2) *Peribacillus frigiditolerans* GR2 and (3) *Bacillus cereus* group – *Bacillus thuringiensis* GS1, *Bacillus cereus* GS3, GL2, and GR3, *Bacillus wiedmannii* GL3, *Bacillus pretiosus* GR1, and *Bacillus* sp. GL1.

Among the strains isolated from *Galium aparine* L., two strains, GS2 and GL1, showed the possibility of being novel strains, as ANI and dDDH values are too low to classify them into known species (Table 2). Their taxonomic identification was determined to be *Priestia* sp. and *Bacillus* sp., respectively. Surprisingly, based on manual comparisons of the two strains, *Priestia* sp. GS2 displayed high similarity to *Bacillus* sp. 1708r2a1 (NCBI GenBank assembly no. GCA_024134725.1), which was isolated from the clean room of NASA Center with ANI and dDDH values reaching 99.48% and 95.70%, respectively. In contrast, these parameters are below threshold established for the closest type strain *Priestia flexa* NBRC 15715 (NCBI GenBank assembly no. GCA_001591565.1).

Firmicutes, with *Bacillus* as the most representative genus, is one of the most prevalent phyla among culturable endophytic bacteria. *Bacillus* spp. endophytes are known for their plant growth potential and often demonstrate antibacterial and antifungal activities (Lopes et al., 2018). They are very abundant in soil; thus, they probably colonize the plant via roots attracted by their exudates. *Bacillus* genus was known for its heterogeneity, which was a result of loose classification criteria in the past, by which the capability of forming spores in the presence of oxygen was sufficient (Logan et al., 2009). A total of 2020 was a significant year for *Bacillus* genus taxonomy (Gupta et al., 2020; Patel and Gupta, 2020). First, six new genera of *Bacillus* species were proposed after a broad analysis of 1,172 core *Bacillaceae* proteins and the identification of conserved signature indels (CSIs) for each – *Peribacillus*, *Cytobacillus*, *Mesobacillus*, *Neobacillus*, *Metabacillus*, and *Alkalihalobacillus* (Patel and Gupta, 2020). To *Peribacillus* spp., species such as *Bacillus muralis*, *butanolivorans*, and *simplex*, have been transferred due to the three CSIs (HAMP domain-containing protein, phosphor-N-acetylmuramoyl pentapeptide-transferase and stage II sporulation protein E). *Peribacillus frigiditolerans* has come an even longer way to proper identification, as it was first classified as *Brevibacterium*

frigiditolerans, then reclassified as *Bacillus frigiditolerans* (Liu et al., 2020; Montecillo and Bae, 2022a). Later, another 17 distinct *Bacillus* species clades were distinguished based on CSIs, which included the separation of *Priestia* genus (*megaterium* clade), to which *Bacillus abyssalis*, *aryabhattai*, *endophyticus*, *filamentosus*, *flexus*, *koreensis*, and *megaterium* were transferred as they all contain two CSIs in the protein oligoribonuclease NrnB (DHH superfamily) (Gupta et al., 2020).

Although the *Bacillus subtilis* and *Bacillus cereus* clades are not phylogenetically related, the *Bacillus* genus is restricted only to members of these two clades. The *subtilis* clade is actually *Bacillus sensu stricto*, as it contains the type strain *Bacillus subtilis*. On the other hand, however, the *cereus* clade cannot be named a separate genus as it contains multiple human pathogens according to Rule 56a of the International Code of Nomenclature of Prokaryotes (Gupta et al., 2020; Oren et al., 2023). *Bacillus cereus* group members often contaminate food and feedstock, as their spores can survive dehydration and pasteurization processes. Distinguishing markers between three pathogenic *Bacillus cereus* group members (*Bacillus cereus*, *Bacillus thuringiensis*, *Bacillus anthracis*) are often encoded by genes contained in plasmids. *Bacillus cereus* is an opportunistic pathogen that causes food poisoning to which chromosomally encoded Nhe (non-hemolytic enterotoxin), Hbl (hemolytic enterotoxin), and CytK (cytotoxin K) protein toxins contribute, as well as cereulide toxin on its megaplasmid pCER270 (Ehling-Schulz et al., 2019). *Bacillus thuringiensis* is known for its insecticidal activity due to the presence of *cry*, *cyt* and *vip* genes (*Bt* genes) on its large plasmids; however, biopesticides based on *B. thuringiensis* are recognized as safe for humans (Ehling-Schulz et al., 2019; Raymond and Federici, 2017). *Bacillus anthracis* holds historical significance in microbiology as it was the first bacterium discovered to be a pathogen and the first evidence for the germ theory of diseases (Blevins and Bronze, 2010). Its pathogenicity depends on two plasmids, pXO1 and pXO2, containing genes encoding for the anthrax toxin (*pag*, *lef*, *cya*) and polyglutamate capsule (*capABCDE*) (Blevins and Bronze, 2010; Koehler, 2002).

To validate the taxonomic identification of *Bacillus cereus* group members, the BTyp3 tool was used (Carroll et al., 2020). The identification of the strains presented in Table 2 was confirmed, and they were assigned to two different phylogenetic groups based

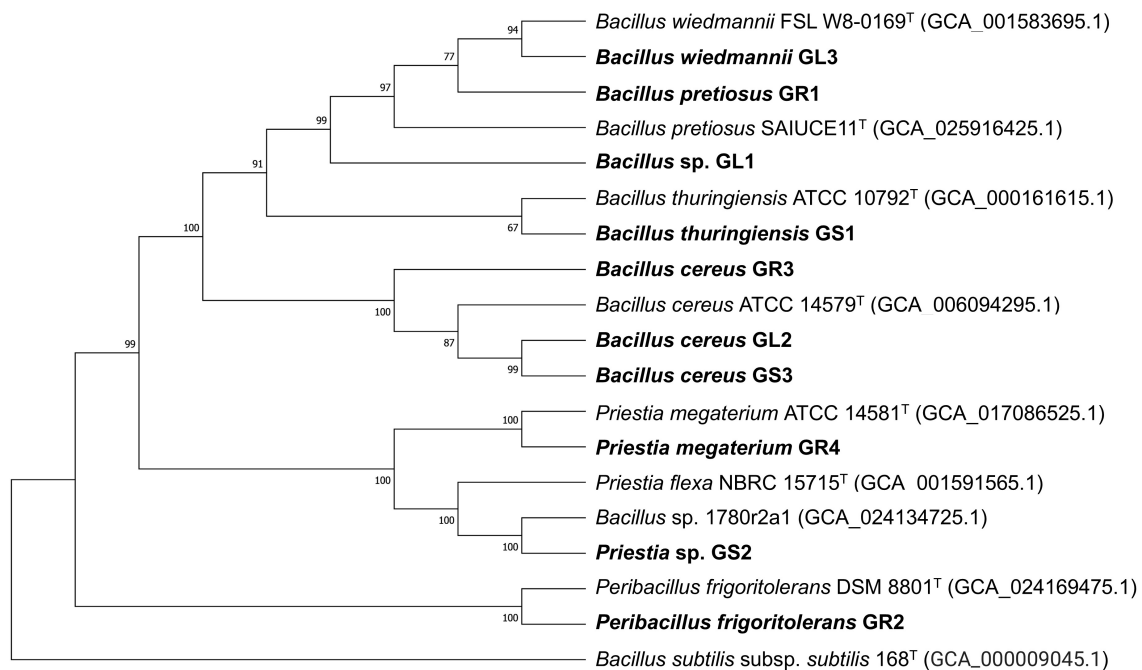


FIGURE 1

Phylogenetic tree of *Galium aparine* L's endophytes and their closest type-strain matches based on whole genome alignments. Single nucleotide polymorphisms (SNPs) and indels within the multiple sequence alignments were constructed by the reference sequence Alignment based Phylogeny builder (REALPHY) and extracted for subsequent phylogeny reconstruction using MEGA v11.0.9 by the Neighbor-joining method, with a bootstrap of 1,000 replications (Bertels et al., 2014; Tamura et al., 2021). Bootstrap confidence levels are indicated at the internodes.

on *panC* sequence (pantoate-beta-alanine ligase gene). GR3, GL2, GS1, and GS3 were assigned to group IV (*Bacillus cereus* sensu stricto), while GR1, GL1, and GL3 strains were assigned to group II (*Bacillus mosaicus/luti*). Neither anthrax nor emetic toxin cereulide encoding gene was found in any genome; all of them contain *nhe* and *hbl* gene clusters and *sph* (sphingomyelinase C) gene. The *cytK* gene was identified only in group IV members (*Bacillus cereus* sensu stricto). No genes encoding *Bt* proteins, including the GS1 genome, were detected. Still, strain GS1 is identified as *Bacillus thuringiensis* due to higher % of ANI and dDDH values between *Bacillus thuringiensis* ATCC 10,792 than *Bacillus cereus* ATCC 14,579. No putative circular plasmid sequence has been detected in GS1 genome; however, there are known cases of *B. thuringiensis* without its characteristic plasmids (Bianco et al., 2021). Strains GR3, GL2, and GS3 were classified as *Bacillus cereus*. Comparison between strains (Figure 2) suggests that GS3 and GL2 may be the same strain, which putatively translocated through the stem to the leaf, with 100% and 99.99% as ANI and DDH values, respectively. Strains GL3 and GR1 were classified as *Bacillus wiedmannii* and *Bacillus pretiosus*, respectively, and GL1 as *Bacillus* sp. with close relation to them. Interestingly, strain GR1 is only the second example of *Bacillus pretiosus*; the first one being isolated from the rhizosphere of *Medicago sativa* in 2022 (Robas Mora et al., 2023).

Researchers have investigated few medicinal plants for the presence of beneficial endophytes, most of which contain *Bacillus*-related species with multiple plant growth-promoting activities. For instance, 65% of bacteria isolated from licorice (*Glycyrrhiza uralensis* F.) were found to belong to the *Bacillus* spp. based on 16S rDNA sequences, and most of them were able to produce indole-3-acetic acid (IAA) and siderophores, fix nitrogen, and solubilize

phosphate (Li et al., 2018). Marchut-Mikołajczyk et al. (2023) isolated two *Bacillus cereus* strains and one *Bacillus mycoides* with the capability of producing polyphenols present in their host plant, *Urtica dioica*. The medicinal plant *Alectra sessiliflora* comprises five endophytic strains from *Lysinibacillus*, *Peribacillus*, and *Bacillus* genera that exhibit notable antibacterial and antitumor activities (Maela et al., 2022). Nisa et al. (2022) managed to obtain all three most iconic strains of *Bacillus cereus* group – *Bacillus cereus*, *Bacillus thuringiensis* and *Bacillus anthracis* – from the medicinal plant *Berberis lycium*; however, identification was based solely on 16S rDNA sequence. Four strains inhabiting *Origanum vulgare* L. were found capable of growing on diesel fuel, probably because it resembled the hydrocarbon components of its essential oil, and all of them belong to *Bacillus* genera based on their whole genome sequences – two *Priestia megaterium*, *Metabacillus dongyingensis*, and *Paenibacillus xylanexedens* (Semenzato et al., 2022b).

Also, *Bacillus*-related bacterial species are often identified as the main endophytic species enhancing plant stress tolerance in extremophilic conditions, which emphasizes the importance of the plant microbiome for its survival and fitness. Bokhari et al. (2019) examined the potential of *Bacillus* strains isolated from five desert medicinal plants. Only under salt stress conditions, *Bacillus cereus*, *B. subtilis*, and *B. circulans* isolates significantly enhanced growth of non-host plants (Bokhari et al., 2019). *Priestia* spp. are also known as very advantageous endophytes. Its prime example, *Priestia megaterium*, has been found to control pear fire blight disease caused by plant pathogen *Erwinia amylovora* (Cui et al., 2023). Similarly, researchers have recognized *Priestia aryabhatai* from wheat as a multi-stress reducer (Shahid et al., 2022). The potential for biofertilization with *Bacillus*

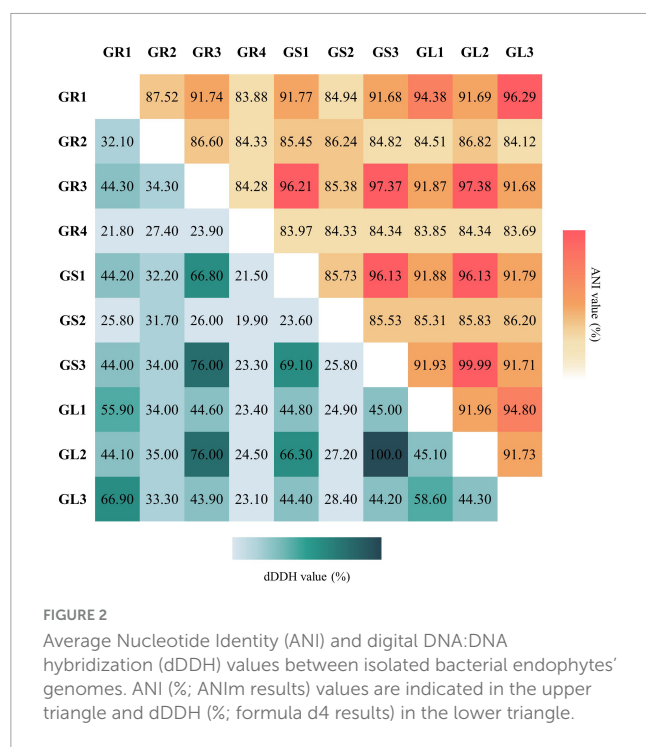


FIGURE 2

Average Nucleotide Identity (ANI) and digital DNA:DNA hybridization (dDDH) values between isolated bacterial endophytes' genomes. ANI (%; ANIm results) values are indicated in the upper triangle and dDDH (%; formula d4 results) in the lower triangle.

endophytes is also well-known (Etesami et al., 2023). Recently, co-inoculation of rice, an important crop in the food industry, with endophytic *B. siamensis* and *P. megaterium*, was found to significantly improve its growth (Rios-Ruiz et al., 2023). In addition, pre-sowing treatment of wheat seeds with endophytic *B. subtilis* and salicylic acid reduced drastically the development of *Fusarium culmorum* instigated root disease in mature plants while simultaneously increasing yields under normal and drought conditions (Lastochkina et al., 2020).

Numerous novel *Bacillus*-related strains are of endophytic origin, including for example *Bacillus endophyticus* (2002), *Bacillus graminis* (2011), *Bacillus endoradicis* (2012), *Bacillus lycopersici* (2015), *Bacillus cabrialesi* (2019), *Bacillus taxi* (2020), *Bacillus mexicanus* (2023), *Bacillus dicomae* (2023) (Bibi et al., 2011; de los Santos Villalobos et al., 2019, 2023; Lin et al., 2015; Makuwa et al., 2023; Reva et al., 2002; Tuo et al., 2020; Zhang et al., 2012), all published validly in *International Journal of Systematic and Evolutionary Microbiology* by the Microbiology Society. However, still not so many whole genomic sequences of endophytic bacteria have been deposited in the NCBI database, leaving open doors for insightful analyses of the features responsible for the wide distribution of *Bacillus*-related bacteria as endophytes.

Functional annotation

Genes associated with plant growth promotion

A substantial amount of genes is significant to the formation and maintenance of an endophytic lifestyle, as well as indirectly for plant growth promotion, adaptation and protection (Figure 3).

Endophytic bacteria enhance the metabolism of essential nutrients in their plant hosts, such as phosphorus, nitrogen, and sulfur, mostly by increasing the bioavailability of these elements, which later serve for amino acids and protein synthesis (Chaudhary et al., 2022). *Galium aparine* L.'s bacteria possess genes encoding various phosphatases (EC 3.1.3.1, EC 3.6.1.11, EC 3.1.3.18, etc.) which catalyze solubilization of insoluble phosphorus forms so the plant can absorb it or mobilize its uptake by binding proteins (e.g., PhnK, PhnL) facilitating the transport onto the roots (Walia et al., 2017). In regards to nitrogen, few acquisition mechanisms are possible, however, in the studied strains, there is a high predominance of enzymes associated with ammonia assimilation (glutamate synthase – EC 1.4.1.13, glutamine synthetase – EC 6.3.1.2), while at the same time, all genomes lack nitrogen fixation enzymes. Additionally, these endophytes can also mineralize organic sulfur compounds in the soil and secrete enzymes such as alkanesulfonate monooxygenase (EC 1.14.14.5), thioredoxin reductases (EC 1.8.1.9) and peroxidases (EC 1.11.1.15) to convert sulfates into sulfites that can be assimilated by plants. Iron acquisition and metabolism category comprises mostly genes related to siderophores synthesis and iron transport. Iron is a crucial micronutrient for plants as it is involved in chlorophyll synthesis, and further maintenance of chloroplast structure, not to mention that it is required by many enzymes for proper work (Rout and Sahoo, 2015). Quorum sensing and biofilm genes do not vary across strains, they contain gene for S-ribosylhomocysteine lyase (EC 4.4.1.21) and S-adenosylmethionine synthetase (EC 2.5.1.6), which take part in the formation of an autoinducer (AI-2) in the quorum-sensing mechanism. Metabolism of aromatic compounds is important to environmental strains as they often have to provide themselves with nitrogen and carbon sources from recalcitrant waste substrates and/or toxic aromatic compounds. Except for *Peribacillus frigiditolerans* GR2, studied endophytes exhibit similar repertoires of enzymes involved in n-phenylalkanoic acid, phenylpropanoid, quinate, gentisate degradation, indicative of their shared environmental origin.

A pivotal plant growth-promoting trait of many bacterial endophytes is their ability to synthesize indole-3-acetic acid (IAA), primarily by tryptophan-dependent pathway (Tang et al., 2023). As a member of the auxin family of indole derivatives, IAA regulates almost every aspect of plant development, including cell division, roots and stem elongation, as well as responses to environmental cues (Duca and Glick, 2020; Tang et al., 2023). Strains *Bacillus pretiosus* GR1, *Peribacillus frigiditolerans* GR2, *Priestia megaterium* GR4, and *Priestia* sp. GS2 possess genes involved in tryptophan biosynthesis, including anthranilate phosphoribosyltransferase (EC 2.4.2.18), phosphoribosylanthranilate isomerase (EC 5.3.1.24), tryptophan synthase alpha chain (EC 4.2.1.20) and tryptophan synthase beta chain (EC 4.2.1.20). The rest of the strains (all belonging to *Bacillus* genus) additionally have indole-3-pyruvate decarboxylase (EC 4.1.1.74), which facilitates the conversion of indole pyruvate (IPA) to indole-3-acetaldehyde (IAAld), suggesting that most likely the IAA production occurs according to the IPA pathway, which was first discovered in plants, now one of the major pathways for microbial IAA biosynthesis (Tang et al., 2023).

While a single endophytic species alone is not able to drastically improve plant fitness and life conditions, a whole group of them can significantly contribute. Notably, certain species demonstrate “specialization” in specific plant growth-promoting traits, as

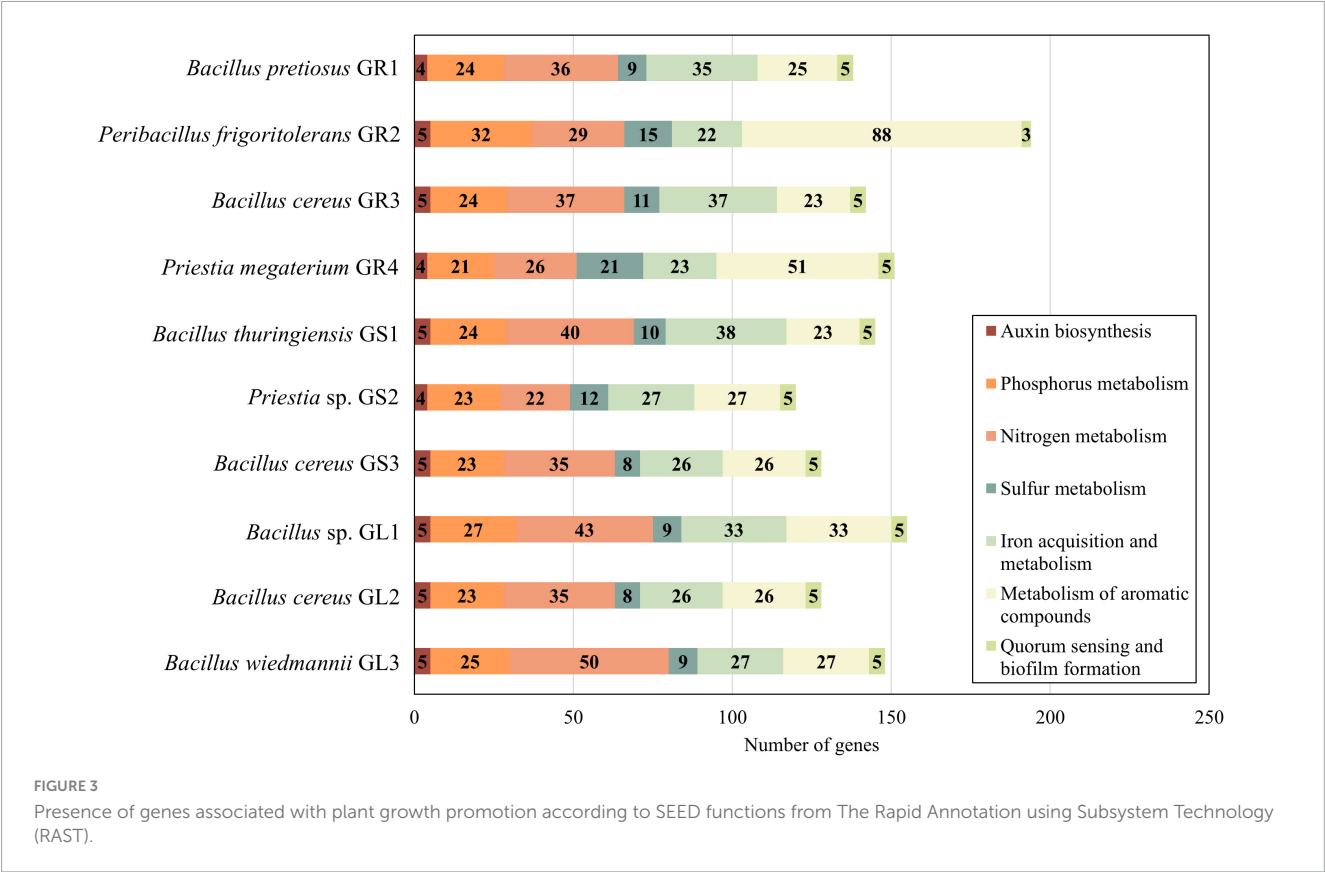


TABLE 3 Mobile genetic elements (MGEs) assigned to major mobileOG categories.

MGEs category	GR1	GR2	GR3	GR4	GS1	GS2	GS3	GL1	GL2	GL3
Integration/excision (IE)*	51	40	70	24	68	14	61	73	61	86
Replication/recombination/repair (RRR)*	52	44	47	57	50	42	46	52	46	50
Phage (P)*	44	16	32	23	48	18	39	33	38	29
Stability/transfer/defense (STD)*	9	4	7	6	6	6	13	6	12	9
Transfer (T)*	27	16	35	29	30	14	28	34	28	32
All	183	120	191	139	202	94	187	198	185	206

*IE, integration/excision; RRR, replication/recombination/repair; P, phage; STD, stability/transfer/defense; T, transfer.

exemplified by *Peribacillus frigiditolerans* GR2, which possesses almost two times more enzymes connected to degradation (especially n-phenylalkanoic acids) than *Bacillus wiedmannii* GL3 with its heightened amount of reductatases for organic sulfur mineralization and assimilation. Given their well-studied plant growth-enhancing properties, *Bacillus* spp. stand out as exemplars for future research on the cruciality of endophytes for plant metabolism and the development of sustainable agriculture agents.

Mobile genetic elements (MGEs)

As major drivers of horizontal gene transfer (HGT), bacterial mobile genetic elements (MGEs), also known as the “bacterial mobilome,” can move within the genome or transfer between bacterial species, ensuring their better adaptation to the occupied ecological niche. That broad term is represented not only by

plasmids and phages but also by transposable elements, insertion, restriction and modification systems, integrative and conjugative elements (ICEs) among others (Khedkar et al., 2022).

Five different categories of MGEs provided by mobileOG-db database include integration and excision (IE; recombinases, transposases, etc.) replication, recombination or nucleic acid repair (RRR; repair and recombination systems, plasmid and phage replication initiators, etc.), phage-related biological processes (P; lysis and lysogeny-associated machinery, etc.) stability, transfer, or defense (STD; CRISPR proteins, etc.) and inter-organism transfer (T; conjugation machinery, etc.); each one corresponding to another key molecular machinery (Brown et al., 2022). Table 3 shows that *Bacillus* spp. strains (GR1, GR3, GS1, GS3, GL1, GL2, GL3) have approximately 180–200 mobile genetic elements (MGEs), whereas *Peribacillus frigiditolerans* GR2, *Priestia megaterium* GR4, and *Priestia* sp. GS2 has slightly fewer, with *Priestia* sp. GS2 has just 94 MGEs.

These data suggest that the bacteria in question are not an exception in nature, as the movement of genetic material acts as a driving force in their adaptation and fitness by facilitating gene loss or gain, in further perspective contributing to evolution. The study of HGT in the context of endophytes and plants has been limited. However, accurate detection of its presence is essential for understanding the complex relationship between them and their mutual influences. This knowledge can directly help to enhance the application of bacteria in sustainable agriculture (Rajabal et al., 2024; Tiwari and Bae, 2020).

In an evolutionary context, a prime example of MGEs is bacteria's acquisition of antibiotic resistance. The rise of antibiotic resistance is one of the most pressing problems in the modern world. According to the World Health Organization (WHO), 700,000 people die each year as a result of this phenomenon and that number is projected to increase to 10,000,000 by 2050 if no effective solutions are developed (Mancuso et al., 2021). The extensive use of antibiotics in clinical settings, hospitals, farm animals, agriculture, and aquacultures has created selective pressure in bacteria toward the acquisition and further dissemination of antibiotic-resistance genes among bacteria (Zhuang et al., 2021). Research conducted by Xu et al. (2021) showed that endophytic bacteria can easily spread antibiotic resistance genes (ARGs) carried on a plasmid.

The resistomes of isolated strains were analyzed to obtain information related to antibiotic resistance genes, which resulted in identification of five antimicrobial resistance (AMR) gene families, each displaying four resistance mechanisms (Table 4).

Strains belonging to *Bacillus* spp. species (GR1, GR3, GS1, GS3, GL1, GL2, GL3) exhibit a comparable number and distribution of AMR genes, suggesting similar resistance to antibiotics among them. Conversely, strains from *Priestia* (GS2, GR4) and *Peribacillus* (GR2) genera have significantly lower numbers of AMR genes. The major family in all bacterial genomes encodes the glycopeptide antibiotic resistance proteins (*vanW*, *vanY*, *vanR* and *vanT* genotypes), which function through the antibiotic inactivation resistance mechanism. Glycopeptide antibiotics, such as vancomycin and teicoplanin, are commonly used to treat life-threatening community-acquired infections caused by methicillin-resistant *Staphylococcus aureus* (MRSA) (Binda et al., 2014). In addition, *Bacillus cereus* Bc β -lactamase gene family was identified, which is implicated in carbapenems resistance (e.g., cephalosporin and β -lactams) by inactivation mechanism. Carbapenem-resistant bacterial strains are being detected in farm animals, due to increased use of antibiotics, and they are mostly plasmid-borne; thus, easily spread (Zhuang et al., 2021). The same resistance mechanism is adapted by fosfomycin thiol transferase encoded by *FosB* gene family acting on phosphonic acid antibiotics, such as fosfomycin (acute cystitis treatment for pregnant women) and fosmidomycin (anti-malarial) (Chekan et al., 2016). Another AMR gene family acting via antibiotic efflux (impacting disinfecting agents and antiseptics) is primarily represented by *qacJ* genotype; however, *qacG* is also additionally present in *Priestia megaterium* GR4, which confers resistance to quaternary ammonium compounds, like acriflavine, ethidium bromide, benzalkonium chloride, cetrimide, dequalinium (Jaglic and Cervinkova, 2012). Of all the strains, only *Peribacillus frigoritolerans* GR2 and *Priestia* sp. GS2 lack the *tetB* gene encoding for tetracycline antibiotic resistance. Tetracycline

TABLE 4 Antimicrobial resistance (AMR) gene families predicted in *Galium aparine* L's bacterial endophytes.

AMR* gene family	ARO* term	Drug class	Antibiotic resistance mechanism	Number of hits									
				GR1	GR2	GR3	GR4	GS1	GS2	GS3	GL1	GL2	GL3
Glycopeptide resistance gene cluster	<i>vanR/vanT/vanW/vanY</i>	Glycopeptide antibiotic	Target alteration	8	3	11	6	10	4	9	8	9	9
<i>Bacillus cereus</i> Bc beta-lactamase	BcI/BcII/BcIII	Cephalosporin, penem	Inactivation	4	1	4	1	4	1	2	5	2	5
Fosfomycin thiol transferase	<i>FosB</i>	Phosphonic acid antibiotic	Inactivation	1	1	1	1	1	1	1	1	2	1
SMR* antibiotic efflux pump	<i>qacJ/qacG</i>	Disinfecting agents and antiseptics	Efflux	1	2	1	3	1	1	1	1	1	1
Tetracycline-resistant ribosomal protection protein	<i>tetB</i>	Tetracycline antibiotics	Target protection	1	0	1	1	1	0	1	1	1	1
All				15	7	18	12	18	7	14	17	14	18

*ARO, antibiotic resistance ontology; AMR, antimicrobial resistance; SMR, small multidrug resistance.

TABLE 5 Phages identified in endophytic genomes.

Strain	Identified phage (GenBank Accession no.)	Region length (kbp)	Start	End	No. of total proteins	GC (%)
<i>Bacillus pretiosus</i> GR1	<i>Bacillus phage</i> phiS3501 (NC_019502)	58.5	3,173,596	3,232,186	68	36.78
	<i>Bacillus phage</i> IEBH (NC_011167)	50.8	779	51,609	90	36.40
<i>Bacillus cereus</i> GR3	<i>Bacillus phage</i> phi4J1 (NC_029008)	29.5	2,859,647	2,889,205	42	36.47
	<i>Brevibacillus phage</i> Jenst (NC_028805)	65.0	565730	630,753	59	33.45
<i>Priestia megaterium</i> GR4	<i>Bacillus phage</i> PM1 (NC_020833)	71.5	786,539	858,040	88	37.05
<i>Bacillus thuringiensis</i> GS1	<i>Geobacillus phage</i> GBSV1 (NC_008376)	58.3	671,230	725,088	50	34.83
	<i>Bacillus phage</i> phiS3501 (NC_019502)	41.0	1	41,007	61	35.17
<i>Bacillus cereus</i> GS3	<i>Bacillus phage</i> Vb_BhaS_171 (NC_03090)	19.9	4,150	24,136	32	35.86

antibiotics, including tetracycline and oxytetracycline, are widely employed in veterinary medicine (Zhuang et al., 2021).

One of the other key factors shaping the ecological balance and evolution of bacteria are phages, the most abundant biological entities on Earth. As predators, they kill bacterial hosts by lysis, causing selective pressure for bacteria resistant to phage infection, which further results in the emergence of different adaptation systems, such as the CRISPR-Cas immune system, leading to co-evolution of phages and bacteria (Casas and Maloy, 2018; Grose et al., 2014). However, phages serve additional functions beyond predation. Phages have the ability to package fragments of any bacterial genome region into their capsids, which they then transfer to other bacterial hosts for integration into the genome through recombination (Chiang et al., 2019). This process, known as transduction, may confer additional advantages for occupied niches as AMR, virulence or novel metabolic properties, which can be fixed in the population. Also, phages can integrate into the bacterial genome during the lysogenic cycle, replicating within bacteria and, when induced, shifting to the lytic phase (Chiang et al., 2019). During dormancy, bacteria are often immune to related phages, yet too high energetic costs associated with maintaining the phage genes may force them to reduce their number. The persistence of their remnants within the genome raises inquiries regarding the potential fitness advantages they may confer; however, the precise nature of these benefits remains unclear (Casas and Maloy, 2018; Nedialkova et al., 2016). After identification of the genes and/or remnants of phages by MGEs analysis, whole phage regions were detected (Table 5).

As a result, the strains *Bacillus pretiosus* GR1, *Bacillus cereus* GR3 and *Bacillus thuringiensis* GS1 were found to harbor two copies of complete phage regions across their genomes, while *Priestia megaterium* GR4 and *Bacillus cereus* GS3 each contained one. *Bacillus* spp. are highly ubiquitous in many ecological niches, making them ideal subjects for studying host-phage interactions (Grose et al., 2014). The majority of the bacteriophage genes are associated with transcriptional regulation and phage machinery, such as capsid, tail, portal, terminase, integrase and replication proteins. While bacteriophages have been identified in some of the endophytic bacteria, their role in the adaptation to endophytic lifestyle and phytobeneficial traits has yet to be elucidated (Li et al., 2021; Wang, 2024). As mentioned before, clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated (Cas) is a bacterial immunity mechanism based on the bacterial ability to

incorporate a fragment of phage genome into their own as a spacer sequence between repeated sequences in CRISPR. When genetic material complementary to the spacer is detected, indicating phage infection, the system initiates degradation, protecting the bacterium (Watson et al., 2021). We identified CRISPR-Cas system elements in genomes of all endophytic bacteria isolates from *Galium aparine* L. (Supplementary Table 1).

Lastly, we searched for putative plasmid sequences within our genomic data, as plasmids represent the largest and most commonly encountered examples of MGEs. First, we used PlasmidSPAdes to find plasmid sequences from short-reads (Antipov et al., 2016). Then we compared these results with circular contigs that we got from Unicycler hybrid assemblies. These sequences were then subjected to a BLASTN search against the NCBI Nucleotide database, and those showing hits to plasmids were further analyzed using the PLSDb database (Schmartz et al., 2022). Results are shown in Table 6. We subjected the identified putative plasmid sequences to Prokka annotation, searched for MGEs, ARGs, and phages, and then visualized as circular map (Supplementary Figure 1).

Half of the isolated strains contain putative plasmid sequences, some even more than one. Except for two hypothetical plasmids (*Bacillus pretiosus* GR1, contig 5; *Bacillus wiedmannii* GL3, contig 4), all of them were matched to plasmids present in the PLSDb database; additionally, all even to those from taxonomically matching strains' genera. Putative plasmid sequences have lower GC content than whole genomes (Table 1), and some of them are especially large (up to 10% of the whole genome as with *Bacillus cereus* GR3), which makes them so-called megaplasmids. Thresholds for minimum megaplasmid size vary as the overall genome size should be considered: however, ≥ 350 kbp proposed by DiCenzo and Finan (2017) based on 10% of the median bacterial genome size is generally accepted cut-off. Here, we report three examples of such putative plasmids (Table 6), all encoding a plethora of enzymes useful for the occupied niche, e.g., esterases, peptidoglycan-N-acetylglucosamine deacetylases, bacillolysin, and biosynthetic gene clusters, which were manually found using BLASTN in various other *Bacillus* plasmids. Thanks to the growing availability of short- and long-read sequencing, which enables hybrid assembly and high-quality genome, it is possible to detect not only plasmids and megaplasmids, but also other more peculiar replicons, such as secondary chromosomes and chromids. The main difference is that chromids (originating from megaplasmids) and secondary chromosomes (from a split of an

TABLE 6 Putative plasmid sequences in *Galium aparine* L's bacterial endophytes.

Strain	Contig no.	Length (kbp)	GC (%)	CDSs*	MGEs*	ARGs*	CRISPR/Cas*	PLSDB (NCBI accession no.)		
								Best match	Cover (%)	Identity (%)
<i>Bacillus pretiosus</i> GR1	2	383.181	32.59	365	43	1	0	<i>Bacillus wiedmannii</i> bv. <i>thuringiensis</i> strain FCC41 plasmid pFCC41-1-490K (CP024685.1)	62	99.48
	5	54.381	36.3	94	15	0	0	–	–	–
<i>Bacillus cereus</i> GR3	3	637.249	32.03	579	71	2	0	<i>Bacillus thuringiensis</i> strain B13 plasmid pBt13367-1 (NZ_CP074713.1)	41	98.93
<i>Priestia megaterium</i> GR4	22	7.738	35.38	8	3	0	0	<i>Priestia megaterium</i> QM B1551 plasmid pBM200 (NC_010009.2)	79	92.51
<i>Bacillus</i> sp. GL1	4	213.927	32.83	188	26	1	1	<i>Bacillus tropicus</i> strain EMB20 plasmid pBEMB20-2 (NZ_CP078083.1)	47	96.56
	6	72.347	31.66	96	11	0	0	<i>Bacillus mycoides</i> strain BPN43/2 plasmid p75 (NZ_CP036011.1)	74	90.88
	9	3.965	34.91	4	1	0	0	<i>Bacillus mycoides</i> plasmid pBMY1 (NC_005703.1)	55	86.41
<i>Bacillus wiedmannii</i> GL3	2	464.727	32.40	409	41	2	0	<i>Bacillus wiedmannii</i> strain EPS29 plasmid ppl557 (NZ_CP133558.1)	70	99.49
	3	226.263	32.72	231	30	0	0	<i>Bacillus thuringiensis</i> HD-771 plasmid p02 (NC_018501.1)	44	89.89
	4	6.402	30.55	7	2	0	0	–	–	–

*CDSs, coding sequences; MGEs, mobile genetic elements; ARGs, antibiotic resistance genes; CRISPR/Cas, clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated (Cas).

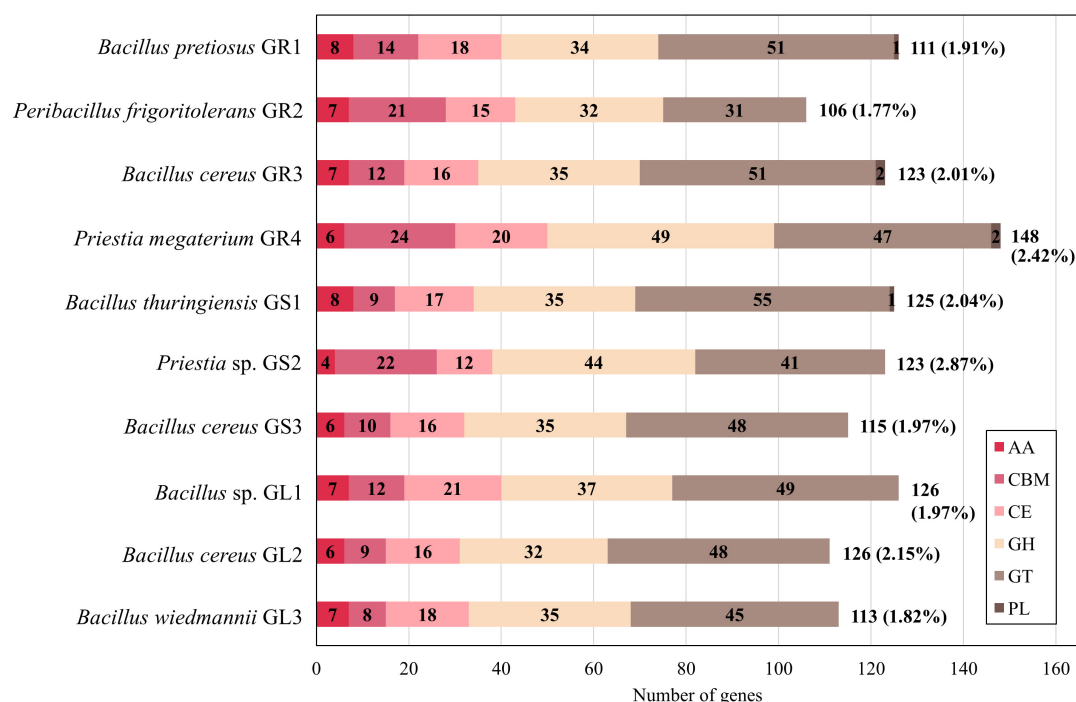


FIGURE 4

Comparison of the distribution of carbohydrate-active enzyme (CAZyme) classes identified in *Galium aparine* L.'s bacterial endophytes. AA, auxiliary activity; CBM, carbohydrate-binding module; CE, carbohydrate esterase; GH, glycoside hydrolase; GT, glycosyltransferase; PL, polysaccharide lyase.

ancestral chromosome) carry essential housekeeping genes found on the chromosome in other species, which is often confirmed experimentally (Hall et al., 2022).

None of the identified plasmids were similar to those encoding toxins characteristic of *Bacillus anthracis* or *thuringiensis* from *Bacillus cereus* group, which further confirms that they are not members of those genera. However, there is a high possibility that strain GS1 is a member of *Bacillus thuringiensis*, but has lost its characteristic plasmid, making it difficult to differentiate from *Bacillus cereus*. Conversely, strain GR3 classified as *Bacillus cereus* member based on ANI and dDDH parameters surpassing thresholds in comparison to the type strain, does possess *Bacillus thuringiensis* plasmid, although it lacks *cry* gene encoding insecticidal crystal protein. In addition, the best match according to the PLSDb database is *Bacillus thuringiensis* strain B13 plasmid pBt13367-1, which also does not possess that gene. Interestingly, *cry* gene was found in the chromosome sequence within a genomic island, the same as in *Bacillus thuringiensis* HER1410 genome (Lechuga et al., 2020; Shwed et al., 2022). These findings highlight the significance and complexity of horizontal gene transfer events in bacteria, showcasing the permanent integration of plasmid-encoded traits into the chromosome structure.

Carbohydrate-active enzymes (CAZymes)

Carbohydrate-active enzymes are families of enzymes intricately associated with the synthesis, degradation and modification of carbohydrates, divided into five classes based

on their function within the CAZy database¹. These classes comprise glycoside hydrolases (GHs), carbohydrate esterases (CEs), glycosyl transferases (GTs), polysaccharide lyases (PLs), and now also non-catalytic carbohydrate-binding modules (CBMs) and auxiliary activity enzymes (AAs) acting in collaboration with other CAZymes to enhance their activity (Davies et al., 2018).

As a result of CAZymes identification analysis, between 106 (*Peribacillus frigoritolerans* GR2) and 148 (*Priestia megaterium* GR4) genes were mapped to the CAZymes family. Figure 4 illustrates the distribution of predicted CAZymes' families across the genomes. The number behind the bar indicates the total number of CAZymes, while the number in brackets represents the percentage of CAZymes to the total number of genes predicted by RAST.

The proportions of CAZymes families differ slightly depending on the species represented. It was the glycosyl transferases (GTs) that were most common in *Bacillus* spp. strains, followed by glycoside hydrolases (GHs) and then by carbohydrate esterases (CEs) and carbohydrate-binding modules (CBMs). In comparison, in strains belonging to *Priestia* spp. and *Peribacillus* spp., the amount of GTs and GHs is similar and CBMs are more abundant than in *Bacillus* spp. CEs and AAs levels are similar in all strains. Comprehensive analysis showed the highest abundance of CBM50 (LysM domains binding to the N-acetylglucosamine residues in bacterial peptidoglycan and chitin), CE4 (esterases catalyzing the de-acylation of polysaccharides), GH13 (hydrolases acting on substrates with α -glucoside linkages), GT2 and GT4 (catalyzing glycoside synthesis) families (CAZypedia, 2024; Davies et al., 2018

¹ www.cazy.org

and [Supplementary Figure 2](#)). Interestingly, only *Bacillus cereus* GR3 and *Priestia megaterium* GR4 contain enzymes from PL12 family (heparin lyases activity) ([Cazy, 2024](#)). PL9 family is more frequent across strains (GR1, GR3, GR4, GS1); its activity is connected to pectin cleaving, a major plant cell wall polysaccharide ([Cazy, 2024](#)).

We have used the number, type, and proportions of CAZymes an organism carries as a marker to assess its adaptation to a specific environment and gain insights into its lifestyle. Polysaccharides consist of diverse glycosyl units, often branched. Microorganisms can use these polysaccharides as sources of carbon and energy, acting as an adaptation mechanism to deal with temporary periods of starvation. The glucose units have to be released by specific enzymes – the key ones belonging to GHs ([Wang et al., 2023](#)). GHs and GTs are the most abundant across the genomes of many plant growth-promoting bacteria (PGPB) ([Wang et al., 2023](#)). GTs synthesize extracellular polysaccharides, which are crucial also for biofilm formation, resistance to environmental pressures, and other significant activities for endophytic bacteria ([Wang et al., 2023](#)). In *Bacillus* spp. strains there is no significant difference in the number and distribution of CAZymes between soil-, leaf- and other-associated PGPB, which is contrary to other common PGPB species (e.g., *Pseudomonas*, *Burkholderia*); thus, makes *Bacillus* spp. remarkably stable genus ([Wang et al., 2023](#)). Also, endophytic isolates from *Galium aparine* L. displayed a proportion of CAZymes to all predicted gene sequences ranging from 1.77%

in *Peribacillus frigiditolerans* GR2 to 2.87% in *Priestia* sp. GS2 ([Figure 4](#)). Free-living organisms typically exhibit a CAZyme repertoire ranging from 1% to 5% of all predicted sequences. A significant reduction suggests a strict intracellular parasitic lifestyle ([Lombard et al., 2014](#)). Endophytic bacteria can not only colonize plants through natural openings and wounds, but some of them can do it actively using cell wall-degrading enzymes (cellulases, xylanases, etc.). *In planta*, the production of such hydrolases can help the plants establish systemic resistance against various pathogenic attacks, particularly chitinases and cellulases, as their activities correlate with the biocontrol of fungal plant pathogens ([Xu et al., 2020](#)).

Bacterial CAZymes are successfully applied for multiple biotechnological (e.g., food processing, detergent additives), medical (e.g., synthesis of pharmaceutical intermediates) and industrial (e.g., xenobiotics degradation, dye production) purposes as they are often more sustainable, cheaper and time efficient solution. Genome mining for CAZymes is a very useful tool for picking the right bacteria and planning more experiments to break down tough substrates like cellulose, starch, lignin, and others. Because those substrates possess unique complex structures, enzymes have to work in conjunction for that purpose. For instance, for effective and complete degradation of chitin, the polymer of (1→4)-β-linked N-acetyl-D-glucosamine (GlcNAc), and chitosan, its deacetylated form, not only chitinases and chitosanases are needed, but also deacetylases, aminidases and lytic monooxygenases ([Kaczmarek et al., 2019](#)). Endophytic bacteria

Degraded substrate	Enzyme [EC number]	Total numbers									
		GR1	GR2	GR3	GR4	GS1	GS2	GS3	GL1	GL2	GL3
Chitin /chitosan	chitinase [3.2.1.14]	2	0	2	0	2	0	2	2	2	2
	chitosanase [3.2.1.132]	2	1	2	0	2	0	2	1	2	2
	lytic chitin monooxygenase [1.14.99.53]	4	0	3	0	4	1	2	3	2	3
	peptidoglycan-N-acetylglucosamine deacetylase [3.5.1.104]	3	2	3	2	3	1	3	2	3	3
	β-N-acetylhexosaminidase [3.2.1.52]	0	1	0	0	0	1	0	0	0	0
Cellulose	cellulase [3.2.1.4]	1	1	1	0	1	0	1	0	1	1
	β-glucosidase [3.2.1.21]	3	1	1	4	0	3	0	2	0	1
	glucan 1,4-β-glucosidase [3.2.1.74]	0	0	0	0	0	1	0	0	0	0
Starch	α-amylase [3.2.1.1]	1	4	2	6	2	5	2	2	2	2
	β-amylase [3.2.1.2]	1	1	0	2	0	2	0	0	0	0
	pullulanase [3.2.1.41]	3	1	3	1	3	3	3	3	3	3
	neopullulanase [3.2.1.135]	1	1	1	0	1	1	1	1	1	1
	α-glucosidase [3.2.1.20]	2	2	2	2	1	2	2	2	2	2
	cyclomaltodextrinase [3.2.1.54]	1	1	1	1	1	1	1	1	1	1
	glucan 1,6-α-glucosidase [3.2.1.70]	2	2	2	2	1	2	2	2	2	2
	glucan 1,4-α-maltohydrolase [3.2.1.133]	1	1	1	0	1	1	1	1	1	1
	galacturonan 1,4-α-galacturonidase [3.2.1.67]	0	0	0	1	0	0	0	0	0	0
	glucan 1,4-α-maltohexaosidase [3.2.1.98]	1	0	1	0	1	1	1	1	1	1
	endo-1,4-β-xylanase [3.2.1.8]	0	1	0	0	0	0	0	0	0	0
Xylan	xylan 1,4-β-xylosidase [3.2.1.37]	2	0	0	1	0	1	0	1	0	1
	acetylxyylan esterase [3.1.1.72]	0	0	0	0	0	1	0	0	0	0
	feruloyl esterase [3.1.1.73]	1	0	0	0	0	0	0	0	0	1
	pectin lyase [4.2.2.10]	0	0	0	0	0	0	0	0	0	0
Pectin	endo-polygalacturonase [3.2.1.15]	0	0	0	1	0	0	0	0	0	0
	α-L-rhamnosidase [3.2.1.40]	1	0	0	0	0	0	0	1	0	1
Lignin	laccase [1.10.3.2]	2	0	1	0	0	0	1	1	1	1

FIGURE 5
Genes associated with polysaccharide degradation.

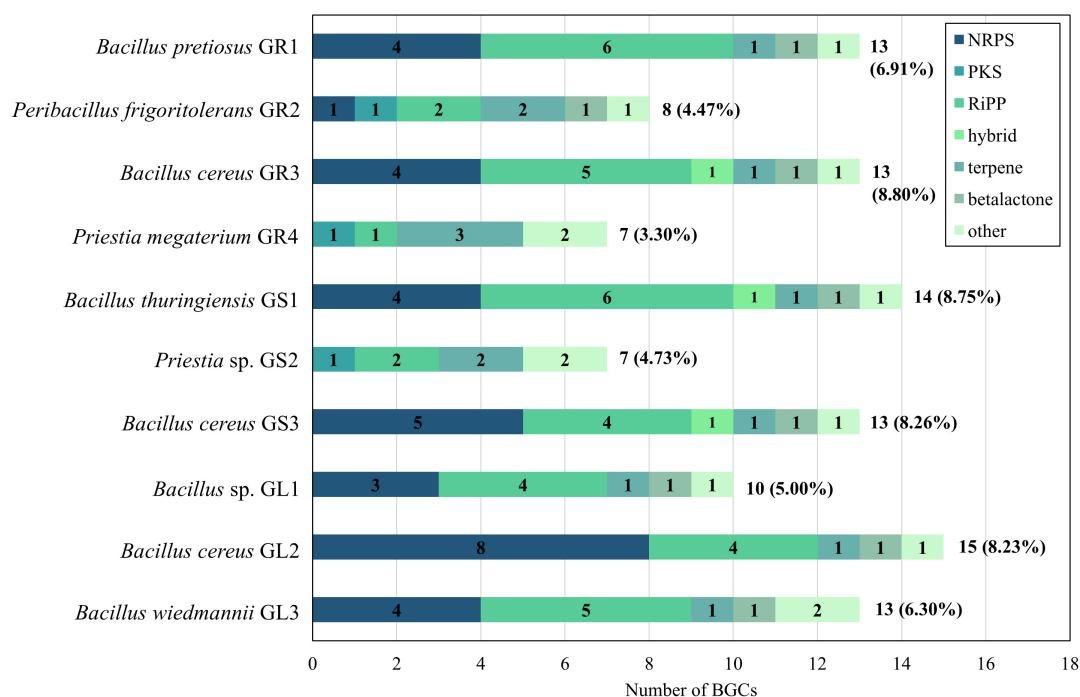


FIGURE 6

Comparison of the distribution of biosynthetic gene clusters (BGCs) types. NRPS, non-ribosomal peptide synthetase; PKS, polyketide synthase; RiPP, ribosomally synthesized and post-translationally modified peptides.

are valuable sources of biotechnologically important enzymes; however, still little information is available on their abundance and potential.

The CAZyme repertoires from *Galium aparine* L. endophytes were manually screened for the presence of specific enzymes associated with the degradation of selected polysaccharides (Figure 5).

All isolated strains display genomic potential for the degradation of polysaccharides, especially chitin, chitosan, cellulose and starch. Strains belonging to *Bacillus* spp. demonstrate a higher number of genes encoding enzymes associated with chitin and chitosan degradation, while strains from *Priestia* spp. those associated with starch hydrolysis, e.g., six copies of an α -amylase gene in *Priestia megaterium* GR4. Only a few genes associated with xylan, pectin and lignin degradation were found across the genomes of all isolates. However, the enzymes presented in Table 4 are only known enzymes with assigned EC number, other enzymes involved in the degradation process may be “hidden” on the family level.

Biosynthetic gene clusters (BGCs)

Bacteria are capable of synthesizing diverse biologically active compounds during their secondary metabolism. Most of the secondary metabolites serve as traits bringing fitness advantages to the occupied ecological niche and microbial community, including cell-cell signaling or nutrient scavenging. However, their production also brings costs to the individual cell as it uses resources from primary metabolism (Santamaria et al., 2022). Apart from their ecological role, secondary metabolites have significant

potential as therapeutics and agricultural agents. Traditional approaches for screening for their production by bacteria involved bioactivity-based assays of their culture supernatants, their extraction and fractionation with solvents and attempts at characterization by such techniques as mass spectroscopy (MS) and nuclear magnetic resonance spectroscopy (NMR). These processes are both expensive and labor-intensive and often lead to the rediscovery of known compounds. As a result of the greater affordability of DNA sequencing, genome mining approach has emerged. Through recognizing specific genetic sequences encoding highly conserved enzymes, biosynthetic gene clusters (BGCs) responsible for producing these metabolites can be identified in bacterial genomes (Russell and Truman, 2020; Yvens et al., 2022). BGCs are categorized based on the presence of such highly conserved enzymes; two main BGCs types are non-ribosomal peptide synthetases (NRPS) and polyketide synthases (PKS), which are produced solely by encoded enzymes, not by ribosomes (Yvens et al., 2022). Ribosomal products are the result of another BGCs type, coding for ribosomally synthesized and post-translationally modified peptides (RiPPs) (Russell and Truman, 2020).

Genomes of *Galium aparine* L.'s bacterial endophytes were analyzed using the antibiotics and secondary metabolite analysis shell (antiSMASH) 7.0 tool to explore their secondary metabolites BGCs (Blin et al., 2023; Figure 6).

Bacteria dedicate approximately 4%–8% of their genomes to the production of secondary metabolites, with *Bacillus cereus* (GR3, GS3, GL2) and *Bacillus thuringiensis* (GS1) species reaching the upper limit of this range. Strains belonging to *Bacillus cereus* species, both *Bacillus cereus* and *thuringiensis* and *Bacillus wiedmannii* (GL1, GL3, GR1) subgroups, overall exhibit a higher

number of clusters (10–15) compared to those in *Priestia* spp. and *Peribacillus frigoritolerans* GR2 (7–8).

The diversity of BGCs types also demonstrates a genus-specific distribution pattern. NRPS and RiPP clusters are more prevalent in *Bacillus* spp. strains (Grubbs et al., 2017; Yin et al., 2023 and Supplementary Table 2). Only one NRPS cluster in all *Bacillus* spp. strains display significant similarity to genes encoding known compounds in Minimum Information about a Biosynthetic Gene Cluster (MIBiG). That cluster is similar in 85% to bacillibactin from *Bacillus subtilis* subsp. *subtilis* str. 168 (MIBiG no.: 0000309), which is a well-described *Bacillus*-specific siderophore (Lalitha and Nithyapriya, 2021). A broad range of RiPPs, including lanthipeptides, sactipeptides and linear azol(in)e-containing peptides (LAP) was found, although only a few clusters showed $\geq 50\%$ similarity to clusters encoding known compounds. An example is the sactipeptide cluster from *Bacillus cereus* GR3, which shares 70% gene similarity to thuricin H from *Bacillus thuringiensis* SF361 (MIBiG no. BGC0000600). Notably, these strains lack any PKS cluster. However, *Bacillus cereus* GR3 and *Bacillus thuringiensis* GS1 each possess one RiPP-NRPS-T1PKS hybrid cluster with 100% of genes similar to genes responsible for biosynthesis of zwittermicin A in *Bacillus cereus* UW85 (MIBiG no.: BGC 0001059). The presence of zwittermicin A is particularly intriguing in terms of plant-endophyte interaction as it exhibits various plant beneficial traits, such as antifungal activities, the ability to suppress plant disease caused by protists and the enhancement of the insecticidal activity of the toxins from *Bacillus thuringiensis* (He et al., 1994; Kevany et al., 2009). *Bacillus cereus* GS3 and *Bacillus cereus* GL2 both have NRPS-transAT-PKS hybrid cluster, though without significant gene similarity to known compounds. All of *Bacillus* spp. members possess one terpene and one beta-lactone gene cluster, as well as one in the “other” category, which includes NRPS-independent siderophore gene cluster similar to petrobactin from *Bacillus anthracis* str. Ames (100%, MIBiG no.: 0000942). In addition to the “other” category, *Bacillus wiedmannii* GL3 encodes a cluster on its putative plasmid sequence featuring a tRNA-dependent cyclodipeptide synthase (CDPS) as the core enzyme with 66% similarity to the pulcherriminic acid encoding cluster from *Bacillus subtilis* subsp. *subtilis* str. 168 (MIBiG no. BGC0002103).

Strains of *Priestia* spp., *Priestia* sp. GS2, *Priestia megaterium* GR4, and *Peribacillus frigoritolerans* GR2 contain fewer BGCs in comparison to *Bacillus* spp. members. They lack NRPS clusters but have a greater number of terpene gene clusters. Both *Priestia* spp. strains harbor a carotenoid gene cluster (terpene; 50% with *Halobacillus halophilus* DSM 2266; MIBiG no. BGC0000645) and type III PKS cluster. Interestingly, *Priestia* sp. GS2 possesses gene cluster encoding bacillopaline (other; 100% similarity with *Paenibacillus mucilaginosus* KNP414; MIBiG no. BGC0002488), metallophore known for its zinc-binding ability (Morey and Kehl-Fie, 2020). *Peribacillus frigoritolerans* GR2 contains an NRPS gene cluster with 87% similarity to the one encoding koranimine from *Bacillus* sp. NK2003 (MIBiG no. BGC0000377), which is also found in other endophytes of that genus (Maela and Serepa-Dlamini, 2023; Montecillo and Bae, 2022b).

As stated before, secondary metabolites are not essential for bacterial growth, but enhance their chances of survival in the occupied environment; a principle that applies also to endophytes. Many compounds, whose encoding clusters were found in *Galium*

aparine L.'s endophytes, exhibit antibacterial and/or antifungal activities or metal-binding properties, directly influencing their ability to thrive inside the plant host. Although current knowledge about endophytic lifestyle is extensive, detailed comparative studies of BGCs in endophytic and non-endophytic bacterial genomes are needed to properly understand the importance of secondary metabolism in the establishment and maintenance of endophytism. Many of these clusters, especially those without any similarity to known compound genes remain silent, meaning they are not expressed under standard cultivation conditions (Zarins-Tutt et al., 2016). Therefore, finding a way to activate their synthesis, both by simply testing various cultivation conditions by One Strain Many Compounds (OSMAC) strategy or molecular biology techniques, is crucial for a comprehensive understanding of bacterial metabolism.

Conclusion

To our knowledge, this is the first research of the endophytic community of the medicinal plant *Galium aparine* L., and one of the few that thoroughly analyzes genomes of endophytic bacteria from the same plant host. We isolated ten strains of bacterial strains from different tissues of *Galium aparine* L. and performed high-quality *de novo* assembly of their genomes using both short and long reads. Taxonomic identification based on whole genome sequences showed that all of them are members of the Bacillaceae family with *Priestia* sp. GS2 and *Bacillus* sp. GL1 potentially representing new species. Additionally, *Bacillus pretiosus* GR1 was identified as the second member of its genus. We examined the distribution of mobile genetic elements, including phages and plasmids, as well as their repertoires of carbohydrate-active enzymes and secondary metabolites. Our findings show the rich biosynthetic capacity among the endophytes, which may not only play an important role in adaptation to the endophytic lifestyle but, moreover, offer potential for diverse biotechnological applications. Although *Bacillus*-related strains are well-known as plant growth promoters, their full metabolic capacities are yet to be explored. Further detailed genome-guided studies are needed as they can lead to the discovery of novel enzymes and metabolites. Medicinal plants are a particularly rich and valuable source of bioactive compounds, yet their bacterial endophytes are still highly. Therefore, research focused on isolating culturable endophytes and exploring their biotechnological potential is of high importance.

Materials and methods

Isolation of endophytic bacteria

Endophytic bacteria were isolated from healthy *Galium aparine* L. herb collected in spring 2022 from a neighboring area of a military airport in Łask, Poland (51°34'02.8"N 19°11'04.6"E). The plant was dug up and quickly transported to the laboratory. After cleaning with tap water to remove soil particles, healthy parts were subjected to surface sterilization (90% ethanol - 3 min, 6.25% sodium hypochlorite - 5 min, 90% ethanol - 30 s), followed by rinsing five times in sterile water (Marchut-Mikolajczyk et al.,

2018). Sterilized plant fragments, as well as 200 µl of water from last rinsing as sterilization control, were placed on LB agar medium (bacto-peptone, 10 g; sodium chloride, 10 g; yeast extract, 5 g; agar, 25 g; dH₂O, 1 L; pH 7.0 ± 0.2) and incubated for 7 days at 30°C. Obtained bacterial colonies were purified, Gram-stained and maintained in −80°C as 25% (v/v) glycerol stocks (Coico, 2005).

DNA extraction, library preparation and sequencing

Genomic DNA of all strains was extracted from overnight cultures (LB medium, 30°C, 120 rpm) using both Gram Plus and Yeast Genomic DNA Purification Kit (Eurx, Poland), following the manufacturer protocol, as well as using phenol-chloroform method for Gram-positive bacteria (Wilson, 2001). Obtained genomic DNA was quantified with the Qubit dsDNA BR Assay Kit (Life Technologies), and its purity and integrity were assessed by spectrophotometric absorbance measurement and by agarose gel electrophoresis, respectively. For strains GS2, GS3, GL1, GL2, GR1, GR2, GR3 paired-end libraries (2 × 150 bp) were prepared from kit-extracted gDNA using MGIEasy FS PCR-Free DNA Library Prep Set and sequencing was performed at the BGI-TECH (Wuhan, China) on MGISEQ-2000 Sequencer (MGI, Shenzhen, China); for strains GS1, GL3 and GR4 TruSeq DNA PCR-Free Kit was used and sequencing was performed by MacroGen Europe on Illumina system sequencer. Phenol-chloroform extracted gDNA was subjected to library preparation by Native Barcoding Kit 24 V14 (Oxford Nanopore Technology) and NEBNext Companion Module (New England Biolabs) for subsequent sequencing on MinION Mk1b Nanopore sequencer (Oxford Nanopore Technology).

Genome *de novo* assembly

Short reads underwent quality filtering and trimming using Trimmomatic v0.36 (Bolger et al., 2014). Nanopore-filtered reads were obtained after base calling, demultiplexing and debarcoding using Dorado v7.2.13 from Oxford Nanopore Technologies (ONT). FastQC v0.12.1 was used to examine the quality of reads (Andrews, 2010). *De novo* assembly of genomes was accomplished using Unicycler v0.5.1 using bold mode with an additional Pilon polishing step (Wick et al., 2017). PlasmidSPAdes v3.15.3 was used to assemble putative plasmid sequences from short-read data (Antipov et al., 2016). The assembly metrics and quality of obtained contigs was evaluated using QUAST v5.2.0 (Gurevich et al., 2013), while completeness and contamination with CheckM v1.2.2 (Parks et al., 2015) and BUSCO v5.5.0 (Manni et al., 2021).

Taxonomic identification

The Type Strain Genome Server (TYGS) was used for initial whole genome-based taxonomic analysis against type-strain species present in their database and calculation of dDDH values (Meier-Kolthoff et al., 2022). The genomic data of the closely related strains were downloaded from the NCBI database, and further analyzed by

JSpeciesWS tool to compute ANI values based on BLAST + (ANiB) and on MUMmer (ANIm) (Richter et al., 2016). Additionally, BType3 version 3.4.0 was used for further identification of *Bacillus cereus* group members, using multi-locus sequence typing (MLST), *panC* group assignment and virulence gene detection (Carroll et al., 2020). The Reference sequence Alignment based Phylogeny builder (REALPHY) tool was used for genome-wide comparisons of the closest related species, then MEGA v11.0.9 for the construction of a phylogenetic tree by the Neighbor-joining method with bootstrap values of 1,000 replications (Bertels et al., 2014; Tamura et al., 2021).

Annotation and functional analysis

Genomes were annotated using NCBI PGAP (Tatusova et al., 2016) and RASTtk v1.073 (Aziz et al., 2008). Prediction of CRISPR and Cas proteins was done by the CRISPRCasFinder v4.2.20 (Couvin et al., 2018). The CARD v3.2.8 with RGI v6.0.3 was used to identify antibiotic resistance genes (perfect and strict hits-only mode) (Alcock et al., 2023). Phage regions were identified using the PHASTER tool (Arndt et al., 2016). The presence of bacterial MGEs was investigated by the program mobileOG-db 1.6v1 (Brown et al., 2022). Putative plasmid sequences were subjected to PLSDB database using mash dist mode, and their features were studied using the programs described above, and visualized using Proksee (Grant et al., 2023; Schmartz et al., 2022). CAZyme gene analyses were performed using the dbCAN3 server; searches were run against Pfam Hidden Markov Models (HMMs) database for sub- and family annotations (e-value < 1e-15, coverage > 0.35) with DIAMOND (e-value < 1e-102) for improved prediction accuracy (Zheng, 2023). Hits predicted by at least two tools were selected for analysis. Secondary metabolite gene clusters were found using bacterial antiSMASH v7.0 in relaxed detection strictness mode with all extra features on, including KnownClusterBlast, MIBiG cluster comparison, and transcription factor binding sequences (TFBS) analysis (Blin et al., 2023).

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: <https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA1068863>, PRJNA1068863.

Author contributions

NR: Software, Investigation, Conceptualization, Writing – original draft, Writing – review and editing, Data curation, Visualization, Formal Analysis, Methodology. MD: Software, Resources, Data curation, Writing – review and editing, Investigation, Methodology. OM-M: Conceptualization, Resources, Writing – review and editing, Supervision.

Funding

The author(s) declare that no financial support was received for the research and/or publication of this article.

Acknowledgments

We thank Bartosz Sekuła and Maciej Nielipiński for their help with MinION software and computing resources, and Małgorzata Ryngajło for contacting with Maurycy Daroch. This article has been completed while the first author was the Doctoral Candidate in the Interdisciplinary Doctoral School at the Lodz University of Technology, Poland.

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.

References

- Afzal, I., Shinwari, Z. K., Sikandar, S., and Shahzad, S. (2019). Plant beneficial endophytic bacteria: Mechanisms, diversity, host range and genetic determinants. *Microbiol. Res.* 221, 36–49. doi: 10.1016/j.micres.2019.02.001
- Alcock, B. P., Huynh, W., Chalil, R., Smith, K. W., Raphenya, A. R., Włodarski, M. A., et al. (2023). CARD 2023: Expanded curation, support for machine learning, and resistome prediction at the comprehensive antibiotic resistance database. *Nucleic Acids Res.* 51, D690–D699. doi: 10.1093/nar/gkac920
- Al-Snafi, A. E. (2018). Chemical constituents and medical importance of galium aparine - a review. *Indo Am. J. Pharmaceut. Sci.* 05, 1739–1744. doi: 10.5281/zenodo.1210517
- Andrews, S. (2010). *FastQC: A quality control tool for high throughput sequence data*. Cambridge: Babraham Bioinformatics.
- Antipov, D., Hartwick, N., Shen, M., Raiko, M., Lapidus, A., and Pevzner, P. A. (2016). PlasmidSPAdes: Assembling plasmids from whole genome sequencing data. *Bioinformatics* 32, 3380–3387. doi: 10.1093/bioinformatics/btw493
- Arndt, D., Grant, J. R., Marcu, A., Sajed, T., Pon, A., Liang, Y., et al. (2016). PHASTER: A better, faster version of the PHAST phage search tool. *Nucleic Acids Res.* 44, W16–W21. doi: 10.1093/nar/gkw387
- Auch, A. F., von Jan, M., Klenk, H. P., and Göker, M. (2010). Digital DNA-DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison. *Stand. Genom. Sci.* 2, 117–134. doi: 10.4056/sigs.531120
- Aziz, R. K., Bartels, D., Best, A., DeJongh, M., Disz, T., Edwards, R. A., et al. (2008). The RAST Server: Rapid annotations using subsystems technology. *BMC Genom.* 9:75. doi: 10.1186/1471-2164-9-75
- Bemmann, W., and Kulle, P. (1988). Fungi in the rhizosphere of *Galium aparine* L. *Zentralblatt Mikrobiol.* 143, 389–396. doi: 10.1016/S0232-4393(88)80033-7
- Bertels, F., Silander, O. K., Pachkov, M., Rainey, P. B., and Van Nimwegen, E. (2014). Automated reconstruction of whole-genome phylogenies from short-sequence reads. *Mol. Biol. Evol.* 31, 1077–1088. doi: 10.1093/molbev/msu088
- Bianco, A., Capozzi, L., Monno, M. R., Del Sambio, L., Manzulli, V., Pesole, G., et al. (2021). Characterization of *Bacillus cereus* group isolates from human bacteremia by whole-genome sequencing. *Front. Microbiol.* 11:599524. doi: 10.3389/fmicb.2020.599524
- Bibi, F., Chung, E. J., Jeon, C. O., and Chung, Y. R. (2011). *Bacillus graminis* sp. nov., an endophyte isolated from a coastal dune plant. *Int. J. Syst. Evol. Microbiol.* 61, 1567–1571. doi: 10.1099/ij.s.0.023820-0
- Binda, E., Marinelli, F., and Marcone, G. L. (2014). Old and new glycopeptide antibiotics: Action and resistance. *Antibiotics* 3, 572–594. doi: 10.3390/antibiotics3040572
- Blevins, S. M., and Bronze, M. S. (2010). Robert Koch and the “golden age” of bacteriology. *Int. J. Infect. Dis.* 14, 744–751. doi: 10.1016/j.ijid.2009.12.003
- 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, W46–W50. doi: 10.1093/nar/gkad344
- Bokhari, A., Essack, M., Lafi, F. F., Andres-Barrao, C., Jalal, R., Alamoudi, S., et al. (2019). Bioprospecting desert plant *Bacillus* endophytic strains for their potential to enhance plant stress tolerance. *Sci. Rep.* 9:18154. doi: 10.1038/s41598-019-54685-y
- Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170
- Brown, C. L., Mullet, J., Hindi, F., Stoll, J. E., Gupta, S., Choi, M., et al. (2022). mobileOG-db: A manually curated database of protein families mediating the life cycle of bacterial mobile genetic elements. *Appl. Environ. Microbiol.* 88:e00991. doi: 10.1128/aem.00991-22
- Carroll, L. M., Cheng, R. A., and Kovac, J. (2020). No assembly required: Using BType3 to assess the congruency of a proposed taxonomic framework for the bacillus cereus group with historical typing methods. *Front. Microbiol.* 11:580691. doi: 10.3389/fmicb.2020.580691
- Casas, V., and Maloy, S. (2018). “The role of phage in the adaptation of bacteria to new environmental niches,” in *Molecular mechanisms of microbial evolution. grand challenges in biology and biotechnology*, ed. P. Rampelotto (Cham: Springer), 267–306. doi: 10.1007/978-3-319-69078-0_11
- Cazy (2024). *The carbohydrate-active enZYmes database*. Available online at: <https://www.cazy.org/> (accessed 22 May, 2024)
- CAZypedia (2024). *CAZypedia database*. Available online at: <https://www.cazypedia.org> (accessed 22 May, 2024)
- Chaudhary, P., Agri, U., Chaudhary, A., Kumar, A., and Kumar, G. (2022). Endophytes and their potential in biotic stress management and crop production. *Front. Microbiol.* 13:933017. doi: 10.3389/fmicb.2022.933017
- Chekan, J. R., Cogan, D. P., and Nair, S. K. (2016). Molecular basis for resistance against phosphonate antibiotics and herbicides. *MedChemComm* 7, 28–36. doi: 10.1039/c5md00351b
- Chiang, Y. N., Penadés, J. R., and Chen, J. (2019). Genetic transduction by phages and chromosomal islands: The new and noncanonical. *PLoS Pathog.* 15:e1007878. doi: 10.1371/journal.ppat.1007878
- Coico, R. (2005). Gram staining basic protocol commonly used techniques. *Curr. Protocols Microbiol.* Appendix 3:Appendix 3C. doi: 10.1002/9780471729259.mca03cs00.

Generative AI statement

The authors declare that no Generative AI was used in the creation of this manuscript.

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.2025.1612860/full#supplementary-material>

- Couvin, D., Bernheim, A., Toffano-Nioche, C., Touchon, M., Michalik, J., Néron, B., et al. (2018). CRISPR-CasFinder, an update of CRISPRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. *Nucleic Acids Res.* 46, W246–W251. doi: 10.1093/nar/gky425
- Cui, Z., Hu, L., Zeng, L., Meng, W., Guo, D., and Sun, L. (2023). Isolation and characterization of *Priestia megaterium* KD7 for the biological control of pear fire blight. *Front. Microbiol.* 14:1099664. doi: 10.3389/fmicb.2023.1099664
- Davies, G., Gilbert, H., Henrissat, B., Svensson, B., Vocadlo, D., and Williams, S. (2018). Ten years of CAZypedia: A living encyclopedia of carbohydrate-active enzymes. *Glycobiology* 28, 3–8. doi: 10.1093/glycob/cwx089
- de los Santos Villalobos, S., Robles, R. I., Parra Cota, F. I., Larsen, J., Lozano, P., and Tiedje, J. M. (2019). *Bacillus cabrialesii* sp. nov., an endophytic plant growth promoting bacterium isolated from wheat (triticum turgidum subsp. durum) in the yaqui valley, Mexico. *Int. J. Syst. Evol. Microbiol.* 69, 3939–3945. doi: 10.1099/ijsem.0.003711
- de los Santos Villalobos, S., Félix Pablos, C. M., Ruiz, V. V., and Parra Cota, F. I. (2023). *Bacillus mexicanus* sp. nov., a biological control bacterium isolated from the common bean (*Phaseolus vulgaris* L.) crop in Sinaloa, Mexico. *Int. J. Syst. Evol. Microbiol.* 73:006110. doi: 10.1099/ijsem.0.006110
- DiCenzo, G. C., and Finan, T. M. (2017). The divided bacterial genome. *Microbiol. Mol. Biol. Rev.* 81:e00019-17. doi: 10.1128/MMBR.00019-17
- Duca, D. R., and Glick, B. R. (2020). Indole-3-acetic acid biosynthesis and its regulation in plant-associated bacteria. *Appl. Microbiol. Biotechnol.* 104, 8607–8619. doi: 10.1007/s00253-020-10869-5
- Egamberdieva, D., Wirth, S. J., Alqarawi, A. A., Abd-Allah, E. F., and Hashem, A. (2017). Phytohormones and beneficial microbes: Essential components for plants to balance stress and fitness. *Front. Microbiol.* 8:2104. doi: 10.3389/fmicb.2017.02104
- Ehling-Schulz, M., Lereclus, D., and Koehler, T. M. (2019). The *Bacillus cereus* group: *Bacillus* species with pathogenic potential. *Microbiol. Spect.* 7:10.1128/microbiolspec.gpp3-0032-2018. doi: 10.1128/microbiolspec.gpp3-0032-2018
- Eid, A. M., Fouda, A., Abdel-Rahman, M. A., Salem, S. S., Elsaied, A., Oelmüller, R., et al. (2021). Harnessing bacterial endophytes for promotion of plant growth and biotechnological applications: An overview. *Plants* 10:935. doi: 10.3390/plants10050935
- Etesami, H., Jeong, B. R., and Glick, B. R. (2023). Potential use of *Bacillus* spp. as an effective biostimulant against abiotic stresses in crops—a review. *Curr. Res. Biotechnol.* 5:100128. doi: 10.1016/j.crbiot.2023.100128
- Grant, J. R., Enns, E., Marinier, E., Mandal, A., Herman, E. K., Chen, C. Y., et al. (2023). Proksee: In-depth characterization and visualization of bacterial genomes. *Nucleic Acids Res.* 51, W484–W492. doi: 10.1093/nar/gkad326
- Grose, J. H., Jensen, G. L., Burnett, S. H., and Breakwell, D. P. (2014). Genomic comparison of 93 *Bacillus* phages reveals 12 clusters, 14 singletons and remarkable diversity. *BMC Genom.* 15:1184. doi: 10.1186/1471-2164-15-1184
- Grubbs, K. J., Bleich, R. M., Santa Maria, K. C., Allen, S. E., Farag, S., AgBiome Team, et al. (2017). Large-scale bioinformatics analysis of *Bacillus* genomes uncovers conserved roles of natural products in bacterial physiology. *mSystems* 2:e00040-17. doi: 10.1128/mSystems.00040-17
- Gupta, R. S., Patel, S., Saini, N., and Chen, S. (2020). Robust demarcation of 17 distinct *Bacillus* species clades, proposed as novel *Bacillaceae* genera, by phylogenomics and comparative genomic analyses: Description of *Robertmurraya kyonggiensis* sp. nov. and proposal for an emended genus *Bacillus* limiting it. *Int. J. Syst. Evol. Microbiol.* 70, 5753–5798. doi: 10.1099/ijsem.0.004475
- Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. (2013). QUAST: Quality assessment tool for genome assemblies. *Bioinformatics* 29, 1072–1075. doi: 10.1093/bioinformatics/btt086
- Hagaggi, N. S. A., and Abdul-Raouf, U. M. (2023). Production of bioactive β -carotene by the endophytic bacterium *Citricoccus parietis* AUCs with multiple in vitro biological potentials. *Microb. Cell Fact.* 22:90. doi: 10.1186/s12934-023-02108-z
- Hall, J. P. J., Botelho, J., Cazares, A., and Baltrus, D. A. (2022). What makes a megaplasmid? *Philos. Trans. R. Soc. B: Biol. Sci.* 377:20200472. doi: 10.1098/rstb.2020.0472
- Hardoim, P. R., van Overbeek, L. S., Berg, G., Pirttilä, A. M., Compant, S., Campisano, A., et al. (2015). The hidden world within plants: Ecological and evolutionary considerations for defining functioning of microbial endophytes. *Microbiol. Mol. Biol. Rev.* 79, 293–320. doi: 10.1128/mmbr.00050-14
- Hardoim, P. R., van Overbeek, L. S., Elsas, J. D., and van. (2008). Properties of bacterial endophytes and their proposed role in plant growth. *Trends Microbiol.* 16, 463–471. doi: 10.1016/j.tim.2008.07.008
- He, H., Silo-Suh, L. A., Handelsman, J., and Clardy, J. (1994). Zwittermixin A, an antifungal and plant protection agent from *Bacillus cereus*. *Tetrahedron Lett.* 35, 2499–2502. doi: 10.1016/S0040-4039(00)77154-1
- Ilin, T., Kashpur, N., Granica, S., and Bazylko, A. (2019). Phytochemical profiles and in vitro immunomodulatory activity of ethanolic extracts from *Galium aparine* L. *Plants* 8:541. doi: 10.3389/plants8120541
- Jaglic, Z., and Cervinkova, D. (2012). Genetic basis of resistance to quaternary ammonium compounds - The qac genes and their role: A review. *Vet. Med.* 57, 275–281. doi: 10.17221/6013-VETMED
- Jasim, B., Sreelakshmi, K. S., Mathew, J., and Radhakrishnan, E. K. (2016). Surfactin, iturin, and fengycin biosynthesis by endophytic *Bacillus* sp. from *Bacopa monnieri*. *Microb. Ecol.* 72, 106–119. doi: 10.1007/s00248-016-0753-5
- Kaczmarek, M. B., Struszczyk-Swita, K., Li, X., Szczesna-Antczak, M., and Daroch, M. (2019). Enzymatic modifications of chitin, chitosan, and chitooligosaccharides. *Front. Bioeng. Biotechnol.* 7:243. doi: 10.3389/fbioe.2019.00243
- Kaul, S., Sharma, T., and Dhar, M. K. (2016). “Omics” tools for better understanding the plant–endophyte interactions. *Front. Plant Sci.* 7:955. doi: 10.3389/fpls.2016.00955
- Kenshole, E., Herisse, M., Michael, M., and Pidot, S. J. (2021). Natural product discovery through microbial genome mining. *Curr. Opin. Chem. Biol.* 60, 47–54. doi: 10.1016/j.cbpa.2020.07.010
- Kevany, B. M., Rasko, D. A., and Thomas, M. G. (2009). Characterization of the complete zwittermixin biosynthesis gene cluster from *Bacillus cereus*. *Appl. Environ. Microbiol.* 75, 1144–1155. doi: 10.1128/AEM.02518-08
- Khedkar, S., Smyshlyayev, G., Letunic, I., Maistrenko, O. M., Coelho, L. P., Orakov, A., et al. (2022). Landscape of mobile genetic elements and their antibiotic resistance cargo in prokaryotic genomes. *Nucleic Acids Res.* 50, 3155–3168. doi: 10.1093/nar/gkac163
- Koehler, T. M. (2002). *Bacillus anthracis* genetics and virulence gene regulation. *Curr. Top. Microbiol. Immunol.* 271, 143–164. doi: 10.1007/978-3-662-05767-4_7
- Kusari, S., and Spiteller, M. (2011). Are we ready for industrial production of bioactive plant secondary metabolites utilizing endophytes? *Nat. Prod. Rep.* 28, 1203–1207. doi: 10.1039/c1np00030f
- Lalitha, S., and Nithyapriya, S. (2021). Production of bacillibactin siderophore from soil bacteria, *Bacillus subtilis*: A bioinoculant enhances plant growth in *Arachis hypogaea* L. through elevated uptake of nutrients. *Proc. Int. Sem. Promot. Local Resources Sustainable Agricult. Dev.* 13, 71–82. doi: 10.2991/absr.k.210609.013
- Lastochkina, O., Garshina, D., Allagulova, C., Fedorova, K., Koryakov, I., and Vladimirova, A. (2020). Application of endophytic *Bacillus subtilis* and salicylic acid to improve wheat growth and tolerance under combined drought and fusarium root rot stresses. *Agronomy* 10:1343. doi: 10.3390/agronomy10091343
- Lechuga, A., Lood, C., Salas, M., van Noort, V., Lavigne, R., and Redrejo-Rodríguez, M. (2020). Completed genomic sequence of *Bacillus thuringiensis* HER1410 reveals a cry-containing chromosome, two megaplasmids, and an integrative plasmidial prophage. *G3: Genes Genomes Genet.* 10, 2927–2939. doi: 10.1534/g3.120.401361
- Li, C., Cheng, P., Zheng, L., Li, Y., Chen, Y., Wen, S., et al. (2021). Comparative genomics analysis of two banana *Fusarium* wilt biocontrol endophytes *Bacillus subtilis* R31 and TR21 provides insights into their differences on phyto-beneficial trait. *Genomics* 113, 900–909. doi: 10.1016/j.ygeno.2021.02.006
- Li, L., Mohamad, O. A. A., Ma, J., Friel, A. D., Su, Y., Wang, Y., et al. (2018). Synergistic plant–microbe interactions between endophytic bacterial communities and the medicinal plant *Glycyrrhiza uralensis* F. *Antonie van Leeuwenhoek* 111, 1735–1748. doi: 10.1007/s10482-018-1062-4
- Lin, S. Y., Hameed, A., Liu, Y. C., Wen, C. Z., Lai, W. A., Hsu, Y. H., et al. (2015). *Bacillus lycopersici* sp. nov., isolated from a tomato plant (*Solanum lycopersicum* L.). *Int. J. Syst. Evol. Microbiol.* 65, 2085–2090. doi: 10.1099/ijms.0.000223
- Liu, G. H., Liu, B., Wang, J. P., Che, J. M., and Li, P. F. (2020). Reclassification of *Brevibacterium frigoritolerans* DSM 8801T as *Bacillus frigoritolerans* comb. nov. based on genome analysis. *Curr. Microbiol.* 77, 1916–1923. doi: 10.1007/s00284-020-01964-x
- Liu, Y., Štefanič, P., Miao, Y., Xue, Y., Xun, W., Zhang, N., et al. (2022). Housekeeping gene gyrA, a potential molecular marker for *Bacillus* ecology study. *AMB Exp.* 12:133. doi: 10.1186/s13568-022-01477-9
- Logan, N. A., Berge, O., Bishop, A. H., Busse, H. J., De Vos, P., Fritze, D., et al. (2009). Proposed minimal standards for describing new taxa of aerobic, endospore-forming bacteria. *Int. J. Syst. Evol. Microbiol.* 59, 2114–2121. doi: 10.1099/ijms.0.013649-0
- Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M., and Henrissat, B. (2014). The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* 42, 490–495. doi: 10.1093/nar/gkt1178
- Lopes, R., Tsui, S., Gonçalves, P. J. R. O., and de Queiroz, M. V. (2018). A look into a multifunctional toolbox: Endophytic *Bacillus* species provide broad and underexploited benefits for plants. *World J. Microbiol. Biotechnol.* 34:94. doi: 10.1007/s11274-018-2479-7
- Maela, M. P., and Serepa-Dlamini, M. H. (2023). Genome sequence and characterisation of *Peribacillus* sp. strain AS_2, a bacterial endophyte isolated from *Alectra sessiliflora*. *Microbiol. Res.* 15, 50–65. doi: 10.3390/microbiolres15010004
- Maela, M. P., van der Walt, H., and Serepa-Dlamini, M. H. (2022). The antibacterial, antitumor activities, and bioactive constituents' identification of *Alectra sessiliflora* bacterial endophytes. *Front. Microbiol.* 13:870821. doi: 10.3389/fmicb.2022.870821

- Makuwa, S. C., Motadi, L. R., Choene, M., Liu, Y., and Serepa-Dlamini, M. H. (2023). *Bacillus dicomae* sp. nov., a new member of the *Bacillus cereus* group isolated from medicinal plant *Dicoma anomala*. *Int. J. Syst. Evol. Microbiol.* 73:006112. doi: 10.1099/ijsem.0.006112
- Mancuso, G., Midiri, A., Gerace, E., and Biondo, C. (2021). Bacterial antibiotic resistance: The most critical pathogens. *Pathogens* 10:1310. doi: 10.3390/pathogens10101310
- Manni, M., Berkeley, M. R., Seppey, M., Simão, F. A., and Zdobnov, E. M. (2021). BUSCO update: Novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol. Biol. Evol.* 38, 4647–4654. doi: 10.1093/molbev/msab199
- Marchut-Mikolajczyk, O., Chlebicz, M., Kawecka, M., Michalak, A., Prucnal, F., Nielipinski, M., et al. (2023). Endophytic bacteria isolated from *Urtica dioica* L.: preliminary screening for enzyme and polyphenols production. *Microb. Cell Fact.* 22:169. doi: 10.1186/s12934-023-02167-2
- Marchut-Mikolajczyk, O., Drozdzyński, P., Pietrzyk, D., and Antczak, T. (2018). Biosurfactant production and hydrocarbon degradation activity of endophytic bacteria isolated from *Chelidonium majus* L. *Microb. Cell Fact.* 17:171. doi: 10.1186/s12934-018-1017-5
- Marchut-Mikolajczyk, O., Drozdzyński, P., and Struszczyk-Świta, K. (2020). Biodegradation of slop oil by endophytic *Bacillus cereus* EN18 coupled with lipase from *Rhizomucor miehei* (Palatase®). *Chemosphere* 250:126203. doi: 10.1016/j.chemosphere.2020.126203
- Meier-Kolthoff, J. P., Carbasse, J. S., Peinado-Olarte, R. L., and Göker, M. (2022). TYGS and LPSN: A database tandem for fast and reliable genome-based classification and nomenclature of prokaryotes. *Nucleic Acids Res.* 50, D801–D807. doi: 10.1093/nar/gkab902
- Montecillo, J. A., and Bae, H. (2022a). Reclassification of *Brevibacterium frigoritolerans* as *Peribacillus frigoritolerans* comb. nov. based on phylogenomics and multiple molecular synapomorphies. *Int. J. Syst. Evol. Microbiol.* 72:005389. doi: 10.1099/ijsem.0.005389
- Montecillo, J. A. V., and Bae, H. (2022b). In silico analysis of koranimine, a cyclic imine compound from *Peribacillus frigoritolerans* reveals potential nematocidal activity. *Sci. Rep.* 12:18883. doi: 10.1038/s41598-022-20461-8
- Morey, J. R., and Kehl-Fie, T. E. (2020). Bioinformatic mapping of opine-like zincophore biosynthesis in bacteria. *mSystems* 5:e00554-20. doi: 10.1128/mSystems.00554-20
- Nedialkova, L. P., Sidstedt, M., Koepfel, M. B., Spriewald, S., Ring, D., Gerlach, R. G., et al. (2016). Temperate phages promote colicin-dependent fitness of *Salmonella enterica* serovar Typhimurium. *Environ. Microbiol.* 18, 1591–1603. doi: 10.1111/1462-2920.13077
- Nisa, S., Shoukat, M., Bibi, Y., Al Ayoubi, S., Shah, W., Masood, S., et al. (2022). Therapeutic prospects of endophytic *Bacillus* species from *Berberis lycium* against oxidative stress and microbial pathogens. *Saudi J. Biol. Sci.* 29, 287–295. doi: 10.1016/j.sjbs.2021.08.099
- Oren, A., Arahall, D. R., Göker, M., Moore, E. R. B., Rossello-Mora, R., and Sutcliffe, I. C. (2023). International code of nomenclature of prokaryotes. prokaryotic code (2022 Revision). *Int. J. Syst. Evol. Microbiol.* 73:005585. doi: 10.1099/ijsem.0.005585
- Papik, J., Folkmanova, M., Polivkova-Majorova, M., Suman, J., and Uhlik, O. (2020). The invisible life inside plants: Deciphering the riddles of endophytic bacterial diversity. *Biotechnol. Adv.* 44:107614. doi: 10.1016/j.biotechadv.2020.107614
- Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., and Tyson, G. W. (2015). CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* 25, 1043–1055. doi: 10.1101/gr.186072.114
- Patel, S., and Gupta, R. S. (2020). A phylogenomic and comparative genomic framework for resolving the polyphyly of the genus *Bacillus*: Proposal for six new genera of *Bacillus* species, *Peribacillus* gen. nov., *Cytobacillus* gen. nov., *Mesobacillus* gen. nov., *Neobacillus* gen. nov., *Metabacillus* gen. nov. *Int. J. Syst. Evol. Microbiol.* 70, 406–438. doi: 10.1099/ijsem.0.003775
- Pinski, A., Betekhtin, A., Hupert-Kocurek, K., Mur, L. A. J., and Hasterok, R. (2019). Defining the genetic basis of plant–endophytic bacteria interactions. *Int. J. Mol. Sci.* 20:1947. doi: 10.3390/ijms20081947
- Rajabal, V., Ghaly, T. M., Egidi, E., Ke, M., Penesyan, A., Qi, Q., et al. (2024). Exploring the role of mobile genetic elements in shaping plant–bacterial interactions for sustainable agriculture and ecosystem health. *Plants People Planet* 6, 408–420. doi: 10.1002/ppp3.10448
- Raymond, B., and Federici, B. A. (2017). In defense of *Bacillus thuringiensis*, the safest and most successful microbial insecticide available to humanity - a response to EFSA. *FEMS Microbiol. Ecol.* 93:fx084. doi: 10.1093/femsec/fix084
- Reva, O. N., Smirnov, V. V., Pettersson, B., and Priest, F. G. (2002). *Bacillus endophyticus* sp. nov., isolated from the inner tissues of cotton plants (*Gossypium* sp.). *Int. J. Syst. Evol. Microbiol.* 52, 101–107. doi: 10.1099/00207713-52-1-101
- Richter, M., and Rosselló-Móra, R. (2009). Shifting the genomic gold standard for the prokaryotic species definition. *Proc. Natl. Acad. Sci. U. S. A.* 106, 19126–19131. doi: 10.1073/pnas.0906412106
- Richter, M., Rosselló-Móra, R., Oliver Glöckner, F., and Peplies, J. (2016). JSpeciesWS: A web server for prokaryotic species circumscription based on pairwise genome comparison. *Bioinformatics* 32, 929–931. doi: 10.1093/bioinformatics/btv681
- Rios-Ruiz, W. F., Tuanama-Reátegui, C., Huamán-Córdova, G., and Valdez-Núñez, R. A. (2023). Co-Inoculation of endophytes *Bacillus siamensis* TUR07-02b and *Priestia megaterium* SMBH14-02 promotes growth in rice with low doses of nitrogen fertilizer. *Plants* 12:524. doi: 10.3390/plants12030524
- Robas Mora, M., Fernández Pastrana, V. M., Oliva, L. L. G., Lobo, A. P., and Jiménez Gómez, P. A. (2023). Plant growth promotion of the forage plant *Lupinus albus* var. orden dorado using *Pseudomonas agronomica* sp. nov. and *Bacillus pretiosus* sp. nov. added over a valorized agricultural biowaste. *Front. Microbiol.* 13:1046201. doi: 10.3389/fmicb.2022.1046201
- Rong, S., Xu, H., Li, L., Chen, R., Gao, X., and Xu, Z. (2020). Antifungal activity of endophytic *Bacillus safensis* B21 and its potential application as a biopesticide to control rice blast. *Pesticide Biochem. Physiol.* 162, 69–77. doi: 10.1016/j.pestbp.2019.09.003
- Rosenblueth, M., and Martínez-Romero, E. (2006). Bacterial endophytes and their interactions with hosts. *Mol. Plant-Microbe Interact.* 19, 827–837. doi: 10.1094/MPMI-19-0827
- Rout, G. R., and Sahoo, S. (2015). Role of Iron in Plant Growth and Metabolism. *Rev. Agricultural Sci.* 3, 1–2. doi: 10.7831/ras.3.1
- Russell, A. H., and Truman, A. W. (2020). Genome mining strategies for ribosomally synthesised and post-translationally modified peptides. *Comp. Struct. Biotechnol. J.* 18, 1838–1851. doi: 10.1016/j.csbj.2020.06.032
- Rutkowska, N., Drozdzyński, P., Rynagajlo, M., and Marchut-Mikolajczyk, O. (2023). Plants as the extended phenotype of endophytes — the actual source of bioactive compounds. *Int. J. Mol. Sci.* 24:10096. doi: 10.3390/ijms241210096
- Santamaria, G., Liao, C., Lindberg, C., Chen, Y., Wang, Z., Rhee, K., et al. (2022). Evolution and regulation of microbial secondary metabolism. *eLife* 11:e76119. doi: 10.7554/eLife.76119
- Schmartz, G. P., Hartung, A., Hirsch, P., Kern, F., Fehlmann, T., Müller, R., et al. (2022). PLSDB: Advancing a comprehensive database of bacterial plasmids. *Nucleic Acids Res.* 50, D273–D278. doi: 10.1093/nar/gkab1111
- Semenzato, G., Alonso-Vázquez, T., Duca, S., Del Vassallo, A., Riccardi, C., et al. (2022a). Genomic analysis of endophytic bacillus-related strains isolated from the medicinal plant *Origanum vulgare* L. revealed the presence of metabolic pathways involved in the biosynthesis of bioactive compounds. *Microorganisms* 10:919. doi: 10.3390/microorganisms10050919
- Semenzato, G., Faddetta, T., Falsini, S., Del Duca, S., Esposito, A., Padula, A., et al. (2022b). Endophytic bacteria associated with *Origanum heracleoticum* L. (Lamiaceae) seeds. *Microorganisms* 10:2086. doi: 10.3390/microorganisms10102086
- Shahid, M., Zeyad, M. T., Syed, A., Singh, U. B., Mohamed, A., Bahkali, A. H., et al. (2022). Stress-Tolerant endophytic isolate *Priestia arylabhattai* BPR-9 modulates physio-biochemical mechanisms in wheat (*Triticum aestivum* L.) for enhanced salt tolerance. *Int. J. Environ. Res. Public Health* 19:10883. doi: 10.3390/ijerph191710883
- Sharma, G., Agarwal, S., Verma, K., Bhardwaj, R., and Mathur, V. (2023). Therapeutic compounds from medicinal plant endophytes: Molecular and metabolic adaptations. *J. Appl. Microbiol.* 134:lxad074. doi: 10.1093/jambio/lxad074
- Shwed, P. S., Crosthwait, J., Weedmark, K., and Leveque, G. (2022). The complete genome sequence of *Bacillus thuringiensis* strain ATCC 13367 features a cry-containing chromosome. *Microbiol. Resource Announc.* 11:e0122721. doi: 10.1128/mra.01227-21
- Tamura, K., Stecher, G., and Kumar, S. (2021). MEGA11: Molecular evolutionary genetics analysis Version 11. *Mol. Biol. Evol.* 38, 3022–3027. doi: 10.1093/molbev/msab120
- Tang, J., Li, Y., Zhang, L., Mu, J., Jiang, Y., Fu, H., et al. (2023). Biosynthetic pathways and functions of Indole-3-Acetic acid in microorganisms. *Microorganisms* 11:2077. doi: 10.3390/microorganisms11082077
- Tatsova, T., Dicuccio, M., Badretidin, A., Chetvernin, V., Nawrocki, E. P., Zaslavsky, L., et al. (2016). NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res.* 44, 6614–6624. doi: 10.1093/nar/gkw569
- Tindall, B. J., Rosselló-Móra, R., Busse, H. J., Ludwig, W., and Kämpfer, P. (2010). Notes on the characterization of prokaryote strains for taxonomic purposes. *Int. J. Syst. Evol. Microbiol.* 60, 249–266. doi: 10.1099/ijms.0.016949-0
- Tiwari, P., and Bae, H. (2020). Horizontal gene transfer and endophytes: An implication for the acquisition of novel traits. *Plants* 9:305. doi: 10.3390/plants9030305
- Truyens, S., Weyens, N., Cuypers, A., and Vangronsveld, J. (2015). Bacterial seed endophytes: Genera, vertical transmission and interaction with plants. *Environ. Microbiol. Rep.* 7, 40–50. doi: 10.1111/1758-2229.12181
- Tuo, L., Liu, F., Yan, X. R., and Liu, Y. (2020). *Bacillus taxi* sp. nov., a novel endophytic bacterium isolated from root of *Taxus chinensis* (Pilger) Rehd. *Int. J. Syst. Evol. Microbiol.* 70, 481–486. doi: 10.1099/ijsem.0.003776
- Walia, A., Guleria, S., Chauhan, A., and Mehta, P. (2017). “Endophytic bacteria: Role in phosphate solubilization,” in *Endophytes: Crop productivity and protection*.

sustainable development and biodiversity, eds D. Maheshwari and K. Annapurna (Cham: Springer), doi: 10.1007/978-3-319-66544-3_4

Wang, Y. (2024). Whole-genome analysis revealed the growth-promoting and biological control mechanism of the endophytic bacterial strain *Bacillus halotolerans* Q2H2, with strong antagonistic activity in potato plants. *Front. Microbiol.* 14:1287921. doi: 10.3389/fmicb.2023.1287921

Wang, Z., Lu, K., Liu, X., Zhu, Y., and Liu, C. (2023). Comparative functional genome analysis reveals the habitat adaptation and biocontrol characteristics of plant growth-promoting bacteria in NCBI databases. *Microbiol. Spect.* 11:e0500722. doi: 10.1128/spectrum.05007-22

Watson, B. N. J., Steens, J. A., Staals, R. H. J., Westra, E. R., and van Houte, S. (2021). Coevolution between bacterial CRISPR-Cas systems and their bacteriophages. *Cell Host Microbe* 29, 715–725. doi: 10.1016/j.chom.2021.03.018

Wick, R. R., Judd, L. M., Gorrie, C. L., and Holt, K. E. (2017). Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comp. Biol.* 13:e1005595. doi: 10.1371/journal.pcbi.1005595

Wilson, K. (2001). Preparation of genomic DNA from bacteria. *Curr. Protocols Mol. Biol.* Chapter 2:Unit 2.4. doi: 10.1002/0471142727.mb0204s56

Xu, H., Chen, Z., Huang, R., Cui, Y., Li, Q., Zhao, Y., et al. (2021). Antibiotic resistance gene-carrying plasmid spreads into the plant endophytic bacteria using soil bacteria as carriers. *Environ. Sci. Technol.* 55, 10462–10470. doi: 10.1021/acs.est.1c01615

Xu, W., Zhang, L., Goodwin, P. H., Xia, M., Zhang, J., Wang, Q., et al. (2020). Isolation, identification, and complete genome assembly of an endophytic *Bacillus velezensis* YB-130, potential biocontrol agent against *Fusarium graminearum*. *Front. Microbiol.* 11:598285. doi: 10.3389/fmicb.2020.598285

Yin, Q. J., Ying, T. T., Zhou, Z. Y., Hu, G. A., Yang, C. L., Hua, Y., et al. (2023). Species-specificity of the secondary biosynthetic potential in *Bacillus*. *Front. Microbiol.* 14:1271418. doi: 10.3389/fmicb.2023.1271418

Yvens, A. J., Aida, M., and De, M. (2022). Genome mining for bioprospecting of biosynthetic genes clusters for bacterial metabolites potentially useful in agroecological production. *Agro Productividad* 9, 77–87. doi: 10.32854/agrop.v15i9.2226

Zarins-Tutt, J. S., Barberi, T. T., Gao, H., Mearns-Spragg, A., Zhang, L., Newman, D. J., et al. (2016). Prospecting for new bacterial metabolites: A glossary of approaches for inducing, activating and upregulating the biosynthesis of bacterial cryptic or silent natural products. *Nat. Product Rep.* 33, 54–72. doi: 10.1039/c5np00111k

Zhang, Q., Li, Y., and Xia, L. (2014). An oleaginous endophyte *Bacillus subtilis* HB1310 isolated from thin-shelled walnut and its utilization of cotton stalk hydrolysate for lipid production. *Biotechnol. Biofuels* 7:152. doi: 10.1186/s13068-014-0152-4

Zhang, Y. Z., Chen, W. F., Li, M., Sui, X. H., Liu, H. C., Zhang, X. X., et al. (2012). *Bacillus endoradicis* sp. nov., an endophytic bacterium isolated from soybean root. *Int. J. Syst. Evol. Microbiol.* 62, 359–363. doi: 10.1099/ijs.0.028936-0

Zheng, J. (2023). dbCAN3: Automated carbohydrate-active enzyme and substrate annotation. *Nucleic Acids Res.* 51, W115–W121. doi: 10.1093/nar/gkad328

Zhou, H., Ren, Z. H., Zu, X., Yu, X. Y., Zhu, H. J., Li, X. J., et al. (2021). Efficacy of plant growth-promoting bacteria *Bacillus cereus* YN917 for biocontrol of rice blast. *Front. Microbiol.* 12:684888. doi: 10.3389/fmicb.2021.684888

Zhuang, M., Achmon, Y., Cao, Y., Liang, X., Chen, L., Wang, H., et al. (2021). Distribution of antibiotic resistance genes in the environment. *Environ. Pollut.* 285:117402. doi: 10.1016/j.envpol.2021.117402