

Unravelling the genomic and functional arsenal of *Bacilli* endophytes from plants with different lifestyles and their antimicrobial potential

Nikolaos P. Arapitsas^{1,2,†}, Christos A. Christakis^{1,2,†}, Savvas Paragkamian^{1,†}, Stefanos Soultatos^{1,3}, Franziska Reden^{1,4}, Chrysianna Psarologaki¹, Emmanouil Avramakis⁵, Alexandros Stamatakis^{4,6,7}, Emmanouil A. Markakis³ and Panagiotis F. Sarris^{1,2,8,*}

Abstract

Endophytic microbiomes of crop wild relatives (CWRs) adapted to extreme environments, such as halophytes, are promising sources of plant-beneficial bacteria and secondary metabolites for sustainable food production. Here, we analysed 25 *Bacilli* isolates obtained from CWRs, halophytes and other plant species in Crete, Greece. Using a hybrid Illumina–PacBio sequencing approach, we generated high-quality genomes and performed comparative genomics, phylogenetic and pangenome analyses, complemented by *in vitro* assays. We identified 312 biosynthetic gene clusters (BGCs), nearly 60% of which showed no similarity to known clusters, revealing extensive unexplored biosynthetic potential. These unique BGCs may constitute an adaptive feature enabling endophytic *Bacilli* to colonize and interact with host plants. The isolates spanned diverse genera (*Bacillus*, *Paenibacillus*, *Peribacillus*, *Neobacillus*, *Cytobacillus* and *Rosellomorea*), including three novel species. Phenotypic assays of our isolates demonstrated high salinity tolerance (up to 17.5% wt/v NaCl) and strong antagonism against major bacterial and fungal phytopathogens. Genome mining further revealed a broad array of putatively plant-beneficial traits related to growth promotion, stress adaptation, host interaction and inhibition of pathogens. Together, these findings show that *Bacilli* endophytes from wild and halophytic plants possess exceptional phylogenetic novelty, functional diversity and biosynthetic capacity, providing new genomic and ecological insights into *Bacilli* associated with plants inhabiting extreme environments.

Impact Statement

Discovery of novel diversity: Isolation and genomic characterization of 25 *Bacilli* endophytes from crop wild relatives and halophytes, including three putative new species. **Extreme stress tolerance and biocontrol:** Identification of strains with remarkable salinity tolerance (up to 17.5% NaCl) and strong antagonism against bacterial and fungal phytopathogens. **Unexplored biosynthetic potential:** Prediction of 312 biosynthetic gene clusters, with nearly 60% showing no similarity to known clusters, highlighting *Bacilli* endophytes as a rich reservoir of novel secondary metabolites. **Integrated genomic–functional framework:** Combination of hybrid genome sequencing, comparative genomics, pangenomics and *in vitro* assays reveals the ecological and functional novelty of *Bacilli* endophytes and their relevance for sustainable agriculture.

Received 19 February 2026; Accepted 07 July 2026; Published 24 September 2026

Author affiliations: ¹Department of Biology, University of Crete, Heraklion, Crete 714 09, Greece; ²Institute of Molecular Biology and Biotechnology, Foundation for Research and Technology–Hellas, Heraklion, Crete 70013, Greece; ³Department of Agriculture, School of Agricultural Sciences, Hellenic Mediterranean University, Estavromenos, 71410, Heraklion, Greece; ⁴Institute of Computer Science, Foundation for Research and Technology–Hellas (FORTH), Heraklion, Greece; ⁵Natural History Museum of Crete, University of Crete, Heraklion, Crete, Greece; ⁶Computational Molecular Evolution Group, Heidelberg Institute for Theoretical Studies, Heidelberg, Germany; ⁷Institute for Theoretical Informatics, Karlsruhe Institute of Technology, Karlsruhe, Germany; ⁸Department of Biosciences, College of Life and Environmental Sciences, University of Exeter, Exeter, UK.

*Correspondence: Panagiotis F. Sarris, p.sarris@imbb.forth.gr

Keywords: *Bacilli*; Crete; crop wild relatives; endophytes; halophytes; phytopathogens; plant microbiome.

Abbreviations: BGC, biosynthetic gene cluster; CLP, cyclic lipopeptide; CWR, crop wild relative; GTDB, Genome Taxonomy Database; NA, Nutrient Agar; NRPS, nonribosomal peptide synthetase; PGP, plant growth promotion; PKS, polyketide synthase; RiPP, ribosomally synthesized and post-translationally modified peptide; VOC, volatile organic compound.

†These authors contributed equally to this work

Twelve supplementary figures and three supplementary tables are available with the online version of this article.

001801 © 2026 The Authors



This is an open-access article distributed under the terms of the Creative Commons Attribution License.

DATA SUMMARY

Sequences and assemblies are registered at ENA with the PRJEB93985 accession number and the related accession numbers (Sample ID and Assembly ID) of each isolate are provided in Table 1. Sample metadata are provided at SM 4 (see below) and all accession numbers (Sample ID and Assembly ID) are also provided at SM 5.

Genome and statistics analysis repository: https://github.com/nik-arapitsas/Bacillus_project

Phylogenetic analysis repository: https://github.com/FranziskaReden/bacillus_project

SM: Available at <https://doi.org/10.5281/zenodo.20731929>.

- SM 1=Supplementary_Material_1_Methods_Figures_Tables.pdf
- SM 2=Supplementary_Material_2_in_vitro_Antifungal_Data.xlsx
- SM 3=Supplementary_Material_3_Sequencing_Data.xlsx
- SM 4=Supplementary_Material_4_Checklist_GSC_MIxS.xlsx
- SM 5=Supplementary_Material_5_Collective_Isolates_Data.xlsx
- SM 6=Supplementary_Material_6_Identity_Matrix.xlsx
- SM 7=Supplementary_Material_7_Gene_Lists_query_terms_with_refs.xlsx
- SM 8=Supplementary_Material_8_OrthoFinder_statistics_per_isolate.xlsx
- SM 9=Supplementary_Material_9_BGC_Data.xlsx
- SM 10=Supplementary_Material_10_Genome_Mining_Data.xlsx.

INTRODUCTION

Climate change and unsustainable agricultural practices pose growing threats to food security by driving drought, soil degradation and reduced crop yields [1–4]. Plant-associated microbiomes, particularly endophytes residing within plant tissues, play key roles in host adaptation, stress tolerance and interspecies interactions, yet their diversity and functional capacities remain underexplored in extreme habitats [5–8]. Crop wild relatives (CWRs) and halophytes, which naturally thrive under saline or arid conditions, represent important ecological systems for studying microbial adaptation and co-evolution with plant hosts [9–11].

Members of the class *Bacilli* are ubiquitous in plant microbiomes and have evolved diverse ecological strategies to survive in fluctuating environments, including spore formation, antimicrobial production and competitive interactions with other microbes producing diverse bioactive compounds [12–16]. Taxonomic frameworks for *Bacilli* have been repeatedly revised through 16S rRNA gene sequencing and comparative genomics, revealing previously unrecognized phylogenetic lineages and novel genera [17–23]. Despite this progress, significant gaps remain regarding their functional potential, particularly the repertoire of biosynthetic gene clusters (BGCs) and traits mediating plant colonization and stress adaptation [24, 25].

Here, we conduct an extensive genome analysis of 25 *Bacilli* endophytes, isolated from diverse CWRs and halophytes across Crete and the Chrysi islands, Greece. By integrating hybrid Illumina–PacBio genome sequencing, comparative genomics and phenotypic assays, we examine their taxonomy, biosynthetic diversity and potential plant-beneficial traits. Our study uncovers both the genomic novelty of these microbes and their unique functional arsenal, shedding light on the potential metabolic strategies of *Bacilli* endophytes in extreme plant-associated habitats. By characterizing genomic and functional features of endophytes associated with hosts living in extreme environments, we provide endophyte-oriented genomic insights into plant–microbe interactions and establish a foundation for future studies leveraging novel, naturally adapted microbes and microbial traits that could be tested for their effectiveness in putative applications aimed at enhancing crop resilience and supporting sustainable agricultural practices.

METHODS

Samplings

During a 3 year period (2018–2021) nine samplings were performed across Crete and Chrysi island, for the collection of plant individuals and/or tissues (Fig. 1). Plant habitats included olive groves, sandy beaches and coastal shrublands; hosts comprised cultivated species (e.g. olive trees) and CWR, with emphasis on halophytes from high-salinity areas. Sampling-site details, including GPS coordinates in decimal degrees, host plant species and sampling locations, are provided in Table S1, available in the online supplementary material. Samples were collected with sterile tools, bagged separately and transported at 4 °C to the laboratory for processing.

Microbial isolation, identification and characterization

Plant surface sterilization and endophytic bacterial isolation and identification

Endophytic *Bacilli* were isolated from surface-sterilized roots and leaves of selected plants based on a previously described method [9], plated on three different media [Nutrient Agar (NA), R2A, NA1/2] and incubated at 28 °C. Pure cultures were preserved in 50%

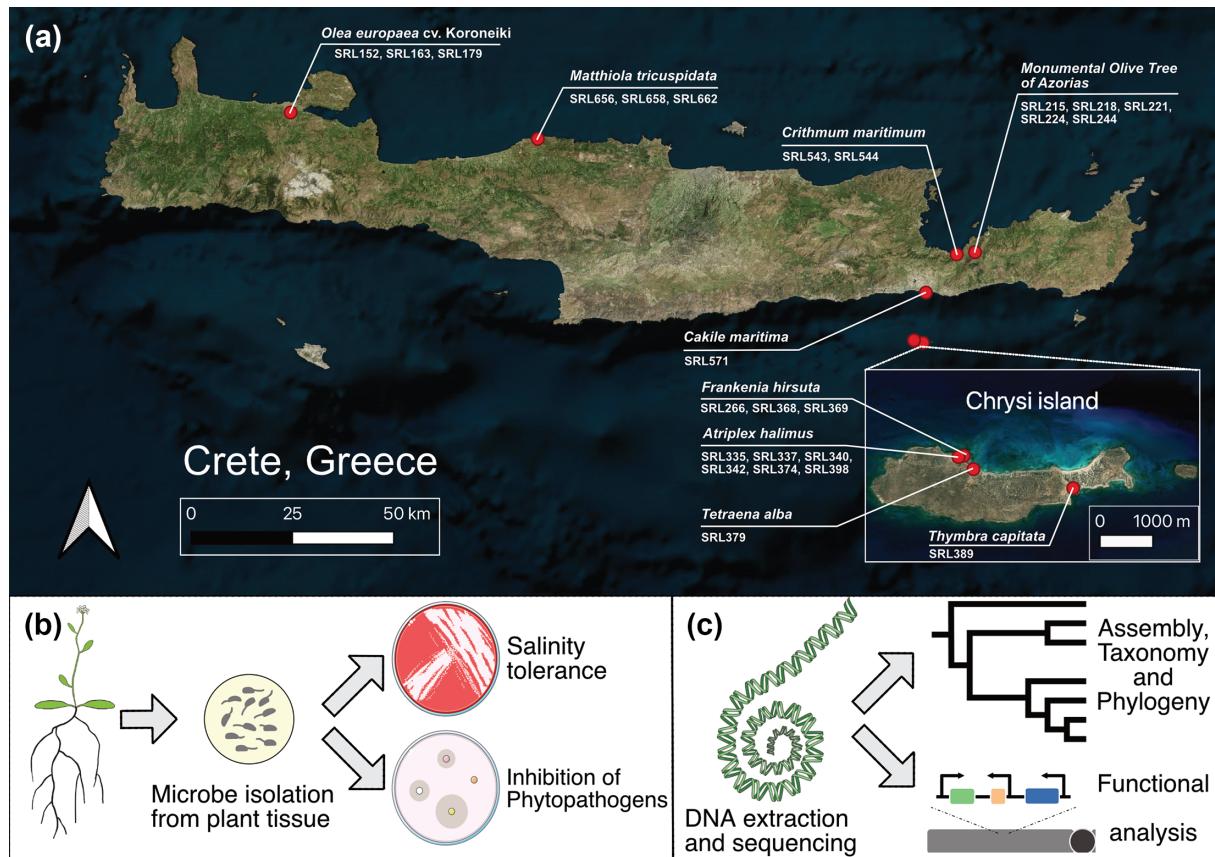


Fig. 1. The three components of the project. (a) Field sampling from different CWR plants across Crete and Chrysi islands. The plant species and the microbial endophytes ID are displayed. (b) The experimental procedure starts from plant tissue sterilization and microbe isolation. Then each microbe is tested against multiple phytopathogens and its maximum salinity tolerance is evaluated. (c) DNA is extracted and sequenced using hybrid technology to investigate the genome, the taxonomy and phylogeny of each isolate. Functional and pangenome analysis is subsequently conveyed. Icons are downloaded from Bioicons except from the dendrogram which is created by Leslie Coonrod from the Noun Project.

(v/v) glycerol at -80°C and preliminarily identified by 16S rRNA gene sequencing. A detailed description of the sterilization, isolation, plating and culturing protocol, as well as the 16S rRNA gene-based identification of the isolates, is provided in (SM 1).

Salt tolerance assay

The salt tolerance of each isolate was assessed by monitoring growth on NA medium supplemented with increasing NaCl concentrations: 0.086, 0.428, 0.856, 1.283, 1.711, 2.139, 2.567 and 2.995 M, corresponding to 0.5, 2.5, 5, 7.5, 10, 12.5, 15 and 17.5% NaCl (wt/v), respectively. Inocula from overnight cultures were streaked onto the plates and incubated at 28°C . Growth was evaluated at 24–96 h post-inoculation.

Growth inhibition of phytopathogens

The antibacterial activity of the isolates was evaluated against four phytopathogenic bacteria: *Ralstonia solanacearum* GMI1000 (INRA CNRS Research Unit, France, kindly provided by Professor Nemo Peeters), *Clavibacter michiganensis* HMU4521 (Collection of the Department of Agriculture, Hellenic Mediterranean University, HMU, kindly provided by Professor Dimitris Goumas), *Paracitovorax citrulli* (formerly *Acidovorax citrulli*) ATCC29625 (Department of Plant Pathology, University of Georgia, kindly provided by Professor Ronald Walcott) and *Xanthomonas campestris* pv. *campestris* strain 8004 [26]. Pathogens were grown in Nutrient Broth (NA without agar) at 28°C , 200 r.p.m. for 48 h. Pathogen lawns were prepared by spreading cultures onto NA plates using sterile cotton swabs and *Bacilli* isolates were spot-inoculated on the same plates and incubated at 28°C . Zones of inhibition (clear areas around the colonies of the isolates) were measured in millimetres (mm) at 72–96 h after incubation. Three biological replicates were performed for each isolate against each bacterial phytopathogen.

The antifungal activity of the isolates was assessed *in vitro* against four phytopathogenic fungi from an established phytopathogen culture collection maintained at the Department of Agriculture, Hellenic Mediterranean University: *Alternaria* sp., *Botrytis cinerea*,

Fusarium oxysporum f.sp. *radicis-cucumerinum* (Forc) and *Verticillium dahliae*. The highly virulent *V. dahliae* isolate 999-1, originally isolated from a symptomatic eggplant (*Solanum melongena* L.) [27] and the highly virulent Forc isolate Afu-68A, originally isolated from a symptomatic cucumber plant [28], were included in the study. The *B. cinerea* and *Alternaria* isolates originated from symptomatic vegetable plants collected from greenhouse crops in the major vegetable-growing areas of Crete, Greece [29]. The isolates were tested in five biological replicates per fungus in dual-culture (confrontation test) and dual-plate (volatile test) assays following the methodology described in [9]. The effect of bacterial isolates on mycelial radial growth (mm), conidia production and hyphae width (µm) and/or (micro)sclerotia formation was evaluated and measured. For the radial growth, growth rates of fungal cultures were calculated as the mean of sequential differences in colony diameter over time per plate. Specifically, the growth step was calculated as the difference in culture diameter over the corresponding day post-inoculation interval and the mean rate per plate was derived by averaging over all steps. Microsclerotial area (cm²) of *V. dahliae* was estimated by scanning the underside of each plate at the end of the bioassays and determining the black pigment area using the ImageJ 1.64 r software [30]. The sclerotia number of *B. cinerea* per plate was also measured. All fungal parameters were calculated in relation to the respective controls (each fungus alone). The percentage change was calculated using the formula:

$$\text{Change (\%)} = \frac{(\text{Value}_t - \text{Value}_c)}{\text{Value}_c} \times 100,$$

where *Value_t* is the treatment mean and *Value_c* is the corresponding control mean of the batch. To assess the significance of microbial effects, one-way ANOVA was conducted. Post-hoc comparisons were performed using *Tukey's HSD* test. Significant microbial effects, defined by a negative percent change and significant ANOVA and *Tukey P*-values (*P*<0.05). Calculations were performed in R (version 4.4) with the *ggplot2* package [31–33]. The raw data of the *in vitro* bioassays to evaluate the antifungal activity of the isolates are available in (SM 2).

DNA extraction for whole-genome sequencing

Genomic DNA was extracted following a procedure (see SM 1) based on a publicly available protocol by A. Kiledal and J. A. Maresca [34], modified from a method by T. Wecke. DNA samples were sequenced by Novogene (Netherlands) using both Illumina NovaSeq (2×150 bp) and PacBio Sequel II platforms. Information about the sequencing data is available in SM 3. The sequences of the 25 isolates were uploaded to the European Nucleotide Archive under the PRJEB93985 study accession number. The accession number of each isolate is provided in Table 1. The metadata of the samples follow the MxS-host-associated checklist of the Genomic Standards Consortium (SM 4) [35].

Genome assembly and annotation

Illumina short reads were quality-controlled with *fastp* v0.23.4 [36], using default settings. PacBio long reads were length-checked with *Filtlong* v0.2.1 (<https://github.com/rrwick/Filtlong>) using settings '--min_length 1,000' and '--keep_percent 95'. Hybrid assemblies were performed with *Unicycler* v0.5.0 [37] using default settings. For SRL389, SRL543 and SRL662, a first assembly was produced using *Flye* v2.9.2-b1786 [38]. Then, this assembly was used in *Unicycler* with the '--existing_long_read_assembly' option, using both the Illumina and PacBio processed reads. The structural statistics and functional completeness of the assemblies were evaluated using *QUAST* v5.0.2 [39] with default options and *BUSCO* 5.8.0 [40] with the *Firmicutes* database (*firmicutes_odb10*), respectively (see SM 5). Part of the assemblies were performed on IMBBC-HCMR's High Performance Computing Cluster Zorba [41].

Genome annotations and genome maps (Fig. S1) were generated using *Bakta* v1.11.2 with its full database v6.0 [42]. Plasmid detection was performed with *RFPlasmid* v0.0.18 using the parameter '--species *Bacillus*' [43]. Contigs were considered as plasmids if they were predicted with high confidence by *RFPlasmid* and confirmed to be circular (SM 5). The predicted proteomes of the isolates were obtained from the *Bakta* output and used as input for *OrthoFinder* v2.5.5 with default options [44].

Taxonomic and phylogenetic analyses

A species tree was inferred from the same 120 marker genes used in the GTDB-Tk pipeline [45, 46]. Unlike GTDB-Tk, which relies on a subset of sites, full-length marker genes were employed. Genome Taxonomy Database (GTD) reference genomes, restricted to the same putative genera as the query assemblies according to GTDB taxonomy, were selected. From this set, to avoid redundancy, highly similar genomes (≥95% average sequence identity across all marker genes; see identity matrix: SM 6) were clustered, and one genome from each cluster was retained for further analysis, resulting in overall 280 genomes, including our 25 genomes and the genomes obtained from the GTDB. Next, the corresponding marker genes were retrieved from the annotated query assemblies. Protein sequences of the 120 marker genes from both query assemblies and GTDB genomes were aligned using *MUSCLE* v5.1 [47] and used to project gaps onto coding DNA sequences producing codon-aware alignments. Gene trees were inferred from these DNA alignments using *IQ-TREE* 2 v2.4.0 [48] and combined into a species tree with *ASTRAL-IV* v1.22.4.6 [49]. Ballistic distances, defined as the sum of branch lengths along the unique path between two tips, were calculated between each assembly and its closest putative relative, as estimated by *FastANI* [50]. Species delimitation was performed with *mPTP* v0.2.5 [51] under maximum likelihood optimization in single-rate mode. Phylogenetic trees were visualized using *TreeViewer* [52]. All analyses were executed through a Snakemake workflow available on GitHub (https://github.com/FranziskaReden/bacillus_project_phylogeny.git) under the Creative Commons license (CC0 1.0) and described in more detail in (SM 1).

Table 1. Selected features of the 25 *Bacilli* isolates and their assemblies

Isolate ID	ENA accession no.	Assembly accession	Closest relative	Ballistic distance	GTDBk FastANI %	Genome size (Mbp)	G+C content (mol %)	Contigs	Assembly completeness	No. of genes	Host plant species	Plant tissue
SRL152	ERS25299874	GCA_979885685	<i>Peribacillus frigoritolerans</i>	0.018605	97.47	5.72	40.50	1	Complete circular	5,598	<i>Olea europaea</i>	Root
SRL163	ERS25299875	GCA_979889575	<i>Bacillus velezensis</i>	0.01077	99.04	3.99	46.51	1	Complete circular	3,907	<i>O. europaea</i>	Root
SRL179	ERS25299876	GCA_984790095	<i>Nochacillus jiddahensis</i>	0.27468	80.14	5.06	39.05	1	Complete circular	4,978	<i>O. europaea</i>	Root
SRL215	ERS25299877	GCA_979895345	<i>Bacillus_A thuringiensis_S</i>	0.03175	99.21	6.08	34.97	2	Complete circular	6,177	<i>O. europaea</i>	Root
SRL218	ERS25299878	GCA_979893815	<i>Bacillus_A thuringiensis_S</i>	0.03175	99.07	6.08	34.97	2	Complete circular	6,176	<i>O. europaea</i>	Root
SRL221	ERS25299879	GCA_979902625	<i>B. velezensis</i>	0.01077	99.04	3.99	46.51	1	Complete circular	3,911	<i>O. europaea</i>	Root
SRL224	ERS25299880	GCA_979901285	<i>Bacillus_A thuringiensis_S</i>	0.03175	99.22	6.08	34.97	2	Complete circular	6,176	<i>O. europaea</i>	Root
SRL244	ERS25299881	GCA_979901685	<i>B. velezensis</i>	0.01077	99.03	3.99	46.51	1	Complete circular	3,909	<i>O. europaea</i>	Root
SRL266	ERS25299882	GCA_979908925	<i>P. frigoritolerans</i>	0.00624	96.51	5.72	40.65	1	Complete circular	5,584	<i>Frankenia hirsuta</i>	Root
SRL335	ERS25299883	GCA_979905105	<i>Cyrobacillus oceanisdiminis_B</i>	0.12969	95.33	5.42	41.50	1	Complete circular	5,514	<i>Atriplex halimus</i>	Root
SRL337	ERS25299884	GCA_984799155	<i>Bacillus salicetis</i>	0.12559	83.69	4.65	42.92	1	Complete circular	4,693	<i>A. halimus</i>	Root
SRL340	ERS25299885	GCA_979910525	<i>P. frigoritolerans</i>	0.026	95.51	5.87	40.31	4	Draft	5,829	<i>A. halimus</i>	Root
SRL342	ERS25299886	GCA_984798905	<i>Penibacillus sp001955855</i>	0.0113	96.08	7.36	45.64	2	Complete circular	6,651	<i>A. halimus</i>	Root
SRL368	ERS25299887	GCA_979908215	<i>Bacillus_A thuringiensis_S</i>	0.01622	99.32	6.67	34.97	14	Draft	7,021	<i>F. hirsuta</i>	Leaves
SRL369	ERS25299888	GCA_979912685	<i>Bacillus_AB infantis</i>	0.01733	98.95	5.06	45.99	1	Complete circular	5,203	<i>F. hirsuta</i>	Leaves
SRL374	ERS25299889	GCA_979914865	<i>B. velezensis</i>	0.01077	99.04	3.99	46.52	1	Complete circular	3,907	<i>A. halimus</i>	Leaves
SRL379	ERS25299890	GCA_979914115	<i>B. velezensis</i>	0.01077	99.03	3.99	46.51	1	Complete circular	3,909	<i>Tetraena alba</i>	Leaves
SRL389	ERS25299891	GCA_979910765	<i>P. frigoritolerans</i>	0.00616	96.49	5.67	40.65	4	Draft	5,556	<i>Thymra capitata</i>	Root
SRL398	ERS25299892	GCA_984799295	<i>Penibacillus sp001955855</i>	0.0113	96.08	7.36	45.64	2	Complete circular	6,652	<i>A. halimus</i>	Root
SRL543	ERS25299893	GCA_984799205	<i>Bacillus_AB infantis</i>	0.06617	90.14	4.95	46.06	1	Complete circular	5,106	<i>Citellum maritimum</i>	Root
SRL544	ERS25299894	GCA_979925865	<i>Rosodonora hukouensis_A</i>	0.06832	99.99	4.66	42.41	1	Complete circular	4,642	<i>C. maritimum</i>	Root
SRL571	ERS25299895	GCA_979925135	<i>Bacillus altitudinis</i>	0.0273	97.98	3.82	41.29	1	Complete circular	3,980	<i>Cakile maritima</i>	Leaves
SRL656	ERS25299896	GCA_979930925	<i>Bacillus paralicheniformis</i>	0.01231	98.85	4.30	45.95	1	Complete circular	4,337	<i>Matthiola tricuspidata</i>	Root
SRL658	ERS25299897	GCA_979932065	<i>B. paralicheniformis</i>	0.01231	98.85	4.30	45.95	1	Complete circular	4,337	<i>M. tricuspidata</i>	Root
SRL662	ERS25299898	GCA_979929605	<i>B. paralicheniformis</i>	0.01231	98.85	4.26	45.88	14	Draft	4,310	<i>M. tricuspidata</i>	Root

Gene numbers are obtained from Bakla annotation. For SRL368, SRL389 and SRL462, the contig number refers to contigs larger than 1,000 bp. A more comprehensive version of this table, including additional info for every isolate, is available in S4d5.

Comparative genomics and functional profiling

Pangenome analysis of selected isolates

The three putative new species and SRL368 (increased genetic novelty) were selected for pangenome analysis along with their closest relatives identified via the identity matrix from taxonomic and phylogenetic analyses (SM 6; see: Taxonomic and phylogenetic analyses). For SRL179, SRL337 and SRL543, the three closest publicly available relatives were selected; for SRL543, SRL369 — identified as a close relative — was also included. For SRL368, five public genomes were selected along with our *Bacillus thuringiensis* isolates SRL215, SRL218 and SRL224. Genomes of the publicly available closest relatives were retrieved from NCBI using the command `ncbi-genome-download bacteria` with the flags `--formats fasta` and `--flat-output`, providing the desired accessions via the `--assembly-accessions` option [53, 54]. The pangenome analysis was performed using `anvi'o` v8 [55, 56]. Genomic and functional annotation was conducted on genome storage databases generated with `anvi-gen-genomes-storage` using `anvi-run-hmms` (with the flags: `-I Bacteria_71` and `--also-scan-trnas`) [57], `anvi-run-ncbi-cogs` [58], `anvi-run-kegg-kofams` [59], `anvi-run-pfams` [60] and `anvi-run-scg-taxonomy` [61, 62]. The `anvi-pan-genome` command was used with `--minbit 0.5`, `--mcl-inflation 10` and `--use-ncbi-blast` [55, 56]. Gene cluster presence/absence data were extracted from the SQLite PAN.db files of each pangenome analysis using `sqlite3` (version 3.45.3) and post-processed with `gawk` (GNU Awk 5.3.0) to count gene cluster occurrences per genome. The graphs were created in R (version 4.5.1) [31] using `tidyverse` v2.0.0 [33] and `UpSetR` v1.4.0 [63].

BGCs, manual curation and genome mining

All assemblies were uploaded to the antiSMASH version 8.0.1 platform for BGC prediction [64]. The detection strictness was set 'relaxed', and all additional features were enabled. Information about the BGC groups of each isolate and their similarity confidence with known BGCs was manually collected and groups were organized into categories (see SM 1).

Annotations were also retrieved from IMG/MER [65] and manually queried for gene terms, names, functions, products and mechanisms for plant growth promotion (PGP) characteristics. To ensure annotation fidelity, annotations were kept only when annotation information was identical in at least three of the following database outputs: KEGG orthology [66], Pfam [60], COG [58], SMART [67], Tigrfam [68], SUPERFamily [69] and CATH FunFam [70]. Genome annotations were then assigned to the following categories based on KEGG orthologies and literature: 'salt stress tolerance', 'disease resistance', 'antioxidants', 'colonization', 'plant hormone and VOCs production', 'PGP and nutrient acquisition', 'chemotaxis and flagella', 'osmoprotection' and 'secretion' [71–76].

PGP-related genes and gene products were derived from experimentally supported studies. Specifically, info for genes related to PGP under salt tolerance, antioxidants and resistance to oxidative stress were derived from: [72, 73]; genes for disease tolerance, plant hormones and volatile organic compounds (VOCs): [71, 73–75, 77]; osmoprotection: [72, 73]; general PGP and increased nutrient acquisition: [73–75]; plant colonization and chemotaxis: [71, 72] and secretion: [72, 76]. An annotation was accepted only if two or more systems supported the same function. Detailed gene lists and references are provided in SM 7.

Gene counts across functional categories and isolates were visualized using R (v4.5.1) [31] with the `pheatmap` package (v1.0.13) [78], while BGC plots were created using `ggplot2` v3.4.4 [32], `tidyverse` v2.0.0 [33], `patchwork` v1.3.1 [79] and `cowplot` v1.1.3 [80].

RESULTS AND DISCUSSION

Genomic characterization and taxonomic diversity

A total of 25 endophytic *Bacilli* were isolated from eight plant species across nine sampling sites in Crete and the nearby Chrysi satellite island (Fig. 1; Tables 1 and S1). Twenty isolates originated from root samples and five from leaves (Table 1). Thus, reflecting a broad environmental sampling from olive trees, halotolerant plants and halophytes from different locations, tissues and diverse lifestyles.

The hybrid sequencing strategy enabled high-quality assemblies with 21 out of 25 being assembled into complete circular genomes. The isolates SRL340, SRL368, SRL389 and SRL662 were draft assemblies, consisting of 2–14 contigs (≥ 1 kb) (Table 1; SM 5). Genome sizes ranged from 3.82 Mb (SRL571; *Bacillus altitudinis*) to 7.36 Mb (SRL342 and SRL398 assigned as *Paenibacillus sp001955855*), with a mean genome size of ~ 5.16 Mb (Table 1; SM 5). The genus *Bacillus* was the most prevalent (16 isolates), followed by *Peribacillus* (4), *Paenibacillus* (2) and one isolate each from the genera *Neobacillus*, *Cytobacillus* and *Rosellomorea* (Table 1; SM 5).

The G+C content spanned from 34.97 mol % (SRL215, SRL218, SRL224 and SRL368; all *B. thuringiensis*) to 46.52 mol % (SRL374, *Bacillus velezensis*), averaging 42.51 mol % across the dataset (Table 1; SM 5). The number of predicted genes, based on Bakta annotation, ranged from 3,907 (SRL163, *B. velezensis*) to 7,021 (SRL368, *B. thuringiensis*) (Table 1; SM 5). The diversity in genome size and features within closely related *Bacilli* species underscores potential genomic adaptations to host plant species or environmental niches, as well as traits possibly acquired through horizontal gene transfer [81, 82]. For example, *B. thuringiensis* SRL368 exhibited a notably large genome (6.67 Mb) and high gene content (7,021), which may be influenced by plasmid content and

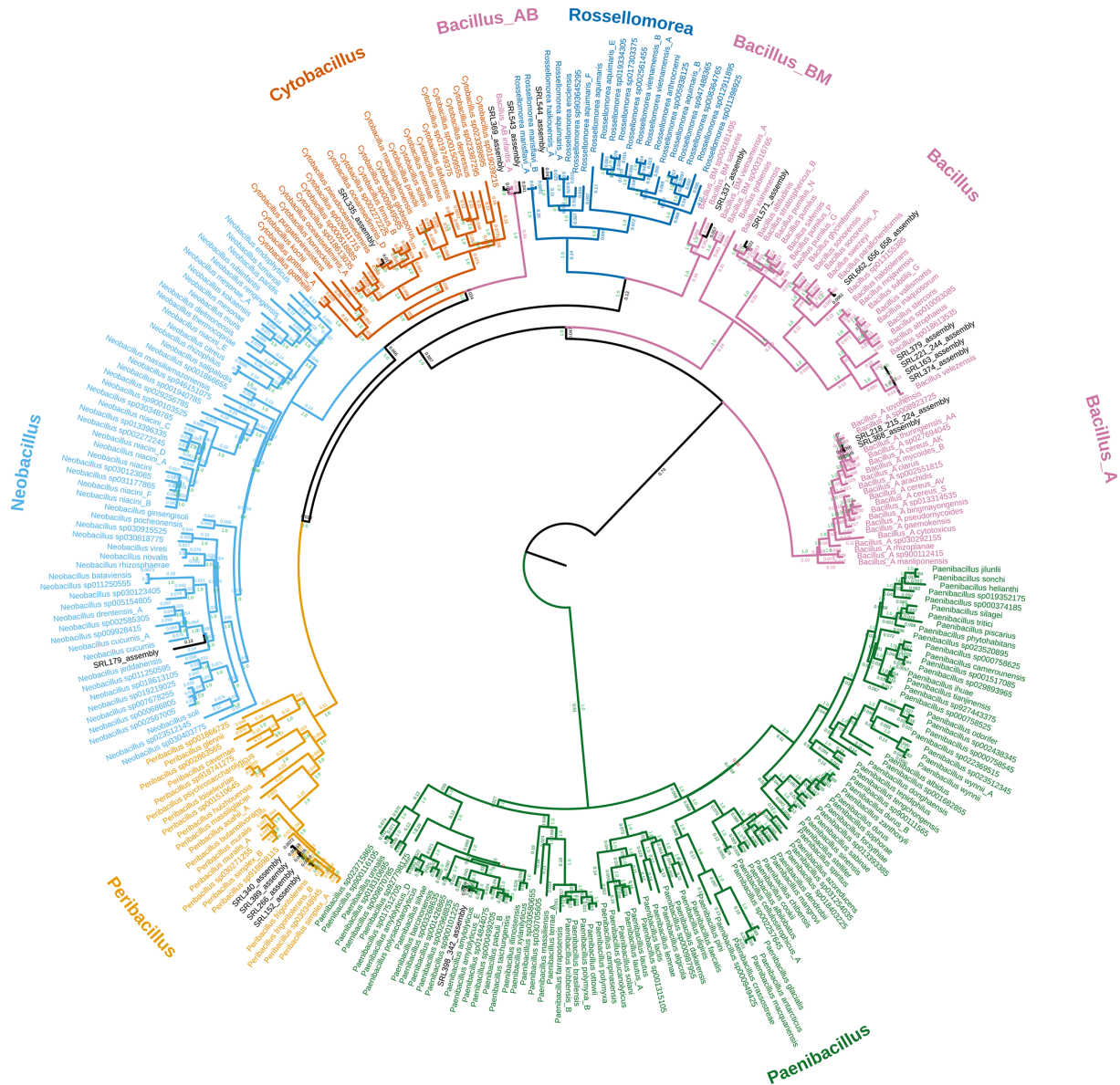


Fig. 2. Phylogenetic tree including the 25 endophytic *Bacilli*. Phylogenetic tree showing the placement of the 25 endophytic *Bacilli* assemblies (black font) among the publicly available reference genomes. Numbers above the internal branches (colour-coded from green to red according to their value) indicate branch support values (local posterior probabilities). Numbers below the branches represent branch lengths in coalescent units. The headers display the genera of the strains which share the same colour.

isolate-specific genomic expansions (Table 1; SM 5). Four putative plasmids were identified in SRL368; however, its fragmented assembly may have led to an underestimation (SM 5).

Phylogenetic analysis of *Bacilli* class including the 25 isolates

To explore the taxonomic diversity and placement among the assembled genomes, a species tree was constructed based on 120 GTDB marker genes (Fig. 2) [45, 46]. Although not intended as a definitive reconstruction of evolutionary history, the tree was used to visualize genome relationships and complement genome-wide similarity estimates, such as those derived from FastANI (Table 1) [50]. The topology of the tree generally reflects the established GTDB taxonomy (Fig. 2), with most assemblies clustering according to their genus-level classifications, some of which are, notably, GTDB-specific [46]. Ballistic distances calculated on the tree are largely consistent with FastANI values, suggesting that divergence in marker genes captures broader genomic similarity patterns reasonably well (Table 1). Taxonomic authority of the genus *Bacillus* is ambiguous and GTDB has the most comprehensive phylogeny. For example, *Bacillus_AB* is closer to *Cyatobacillus* than to *Bacillus* and it does not exist in the NCBI taxonomy.

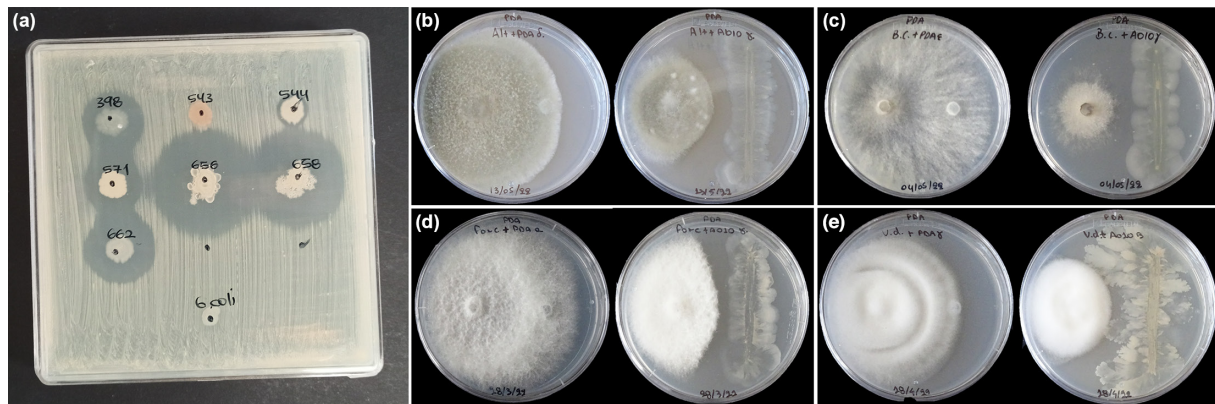


Fig. 3. *In vitro* bioassays against selected phytopathogens. (a) Inhibition of bacterial growth of *C. michiganensis* by the isolates SRL398, SRL571, SRL656, SRL658 and SRL662. Inhibition of fungal growth of (b) *Alternaria* sp., (c) *B. cinerea*, (d) *F. oxysporum* f.sp. *radicle-cucumerinum* and (e) *V. dahliae* by the most effective *Bacillus* isolate (SRL163) in dual-culture assays.

Three isolates, SRL179 (*Neobacillus* sp.), SRL337 (*Bacillus* sp.) and SRL543 (*Bacillus* sp.), exhibited low FastANI values (80.14, 83.69 and 90.14%, respectively) compared to their closest reference genomes (Table 1). They also appear to have substantially diverged from their closest relatives in the inferred phylogenetic tree, with ballistic distances of 0.27468, 0.12559 and 0.06617, respectively, providing strong evidence that they may represent novel species (Table 1; Fig. 2). Consistent with these findings, species delimitation with mPTP [51] estimated SRL179, SRL337 and SRL544 to be unique species, providing additional evidence that at least two of these lineages likely represent previously undescribed taxa.

While the species tree was primarily used to assess taxonomic placement and for visualization, it provides useful insights into the diversity and divergence present within the dataset. To better resolve evolutionary relationships within *Bacillota* genera, future research could explore whole-genome-based inference or incorporate additional genus-specific marker sets for phylogenetic analysis.

***In vitro* assays for salinity tolerance and growth inhibition of selected phytopathogens**

Salinity tolerance

Plant-associated salt-tolerant bacteria can contribute to salt stress alleviation and growth promotion of crops [9, 10, 83]. Most isolates (18) grew at salinity of 7.5% wt/v or higher, while all remaining isolates tolerated salinity levels of up to 5% wt/v (SM 5). The putative novel species SRL179 and SRL337 grew at salinity levels up to 17.5% wt/v, whereas SRL543 tolerated up to 10% wt/v salinity (SM 5). Notably, the *Peribacillus frigoritolerans* isolate SRL266 tolerated salinity as high as 17.5% wt/v, while the other *Peribacillus* isolates SRL152 and SRL340 tolerated up to 5% wt/v and SRL389 tolerated up to 7.5% wt/v salinity (SM 5). Similarly, while *B. thuringiensis* SRL368 grew at salinity levels of up to 17.5% wt/v, other isolates of the same species, SRL215, SRL218 and SRL224, could not tolerate salinity levels above 5% wt/v (SM 5). These properties could confer an advantage to SRL266 and SRL368 for potential applications and underscore their divergence from their close phylogenetic relatives.

Antibacterial activity

The 25 *Bacilli* isolates were evaluated for their ability to inhibit the economically important bacterial phytopathogens *C. michiganensis*, *R. solanacearum*, *P. citrulli* and *X. campestris* pv. *campestris*. Inhibition was most prevalent against the Gram-positive *C. michiganensis*, with 11 isolates displaying clear zones of inhibition ranging from 2 to 20 mm in diameter (Fig. 3a; SM 5). Notably, SRL656, SRL658 and SRL662 exhibited the strongest activity (Fig. 3a; SM 5). Among the *B. thuringiensis* isolates, SRL215, SRL218 and SRL224 inhibited *C. michiganensis*, whereas SRL368 showed no inhibition, again highlighting the variation in the characteristics and properties of separate *B. thuringiensis* isolates (SM 5). In contrast, the other bacterial pathogens, all being Gram-negative, were inhibited by fewer isolates and the resulting zones of inhibition were minimal or barely detectable (1–2 mm) (SM 5).

Antifungal activity

Based on preliminary *in vitro* screening, 15, 17, 16 and 17 bacterial strains exhibited antagonistic activity against *Alternaria* sp., *B. cinerea*, *F. oxysporum* f.sp. *radicle-cucumerinum* and *V. dahliae*, respectively (Table S2). Thus, they were selected for further testing.

In thorough dual-culture assays, the mycelial growth of *Alternaria* sp., *B. cinerea*, *F. oxysporum* f.sp. *radicle-cucumerinum* and *V. dahliae* was inhibited by 12, 15, 13 and 15 isolates, while conidial production of the abovementioned fungi was reduced by 0, 12, 13

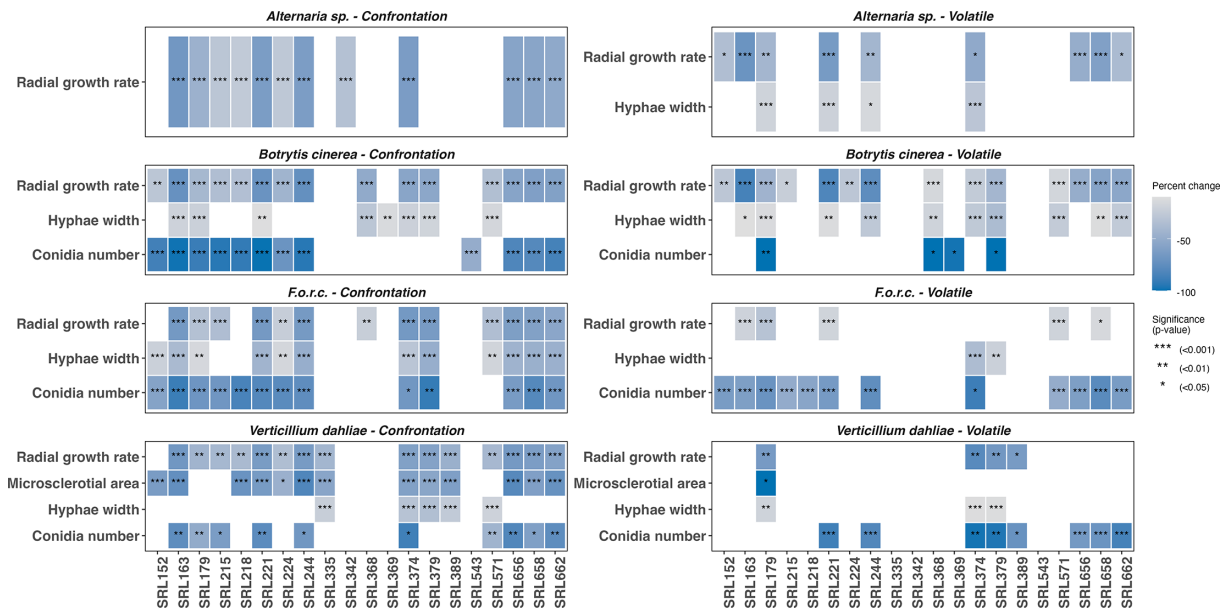


Fig. 4. Graph of antifungal activity of bacterial isolates. Antifungal activity of bacterial isolates against four fungal plant pathogens (*Alternaria* sp., *B. cinerea*, *F. oxysporum* f.sp. *radicum-cucumerinum* and *V. dahliae*) in *in vitro* assays (confrontation and volatile), estimated with four variables. The colour showcases the percent change based on the control of the assay and stars indicate the statistical significance based on the Tukey HSD post-hoc test.

and 10 isolates, respectively (Figs 4, S2 and S3). Moreover, the hyphal width of *B. cinerea*, *F. oxysporum* f.sp. *radicum-cucumerinum* and *V. dahliae* was affected by 8, 12 and 5 isolates respectively; 13 isolates reduced *V. dahliae* microsclerotial area, whereas the number of *B. cinerea* sclerotia was not affected by any of the *Bacillus* isolates tested (Figs 4, S4). These are statistically significant findings which align with earlier studies highlighting the broad-spectrum antifungal activity of *Bacillus* spp., commonly attributed to cyclic lipopeptides (CLPs) and hydrolytic enzymes [84]. Specifically, strains of *B. pumilus* and *Bacillus subtilis* producing surfactin, iturin and fengycin showed antimicrobial activity against several phytopathogens, including *B. cinerea*, *F. oxysporum*, *Rossellomorea solani*, *Sclerotinia sclerotiorum*, *Pythium ultimum* and *Phytophthora capsici* [85].

In thorough dual-plate assays for activity via VOCs, 9 isolates inhibited *Alternaria* sp., 14 *B. cinerea*, 5 *F. oxysporum* f.sp. *radicum-cucumerinum* and 4 *V. dahliae* (Figs 4, S5). These results suggest that VOCs produced by *Bacillus* spp. also exert indirect antifungal effects against certain plant pathogenic fungal species [86]. Such VOC-inducing effects have also been reported against *B. cinerea* in previous studies [87]. VOCs that are commonly produced by *Bacillus* spp. include ketones, alcohols, aldehydes, acids, pyrazines and terpenes. Some of these compounds have been proved effective in inhibiting fungal spore germination and mycelial radial growth *in vitro* [88, 89]. Similarly, *B. velezensis* and *B. subtilis* have been shown to emit VOCs that reduce fungal growth, sporulation and change hyphal morphology [87, 88]. In the present study, VOCs emitted by 4 *Bacillus* isolates affected conidia production of *B. cinerea*, whereas VOCs of 12 and 8 isolates affected conidial production of *F. oxysporum* f.sp. *radicum-cucumerinum* and *V. dahliae*, respectively. However, none of the tested isolates significantly reduced conidial production of *Alternaria* sp. via VOCs-inciting effects. Hyphal width of *Alternaria* sp., *B. cinerea*, *F. oxysporum* f.sp. *radicum-cucumerinum* and *V. dahliae* was significantly altered by 4, 10, 2 and 3 isolates, respectively (Figs 4; Fig. S6). Moreover, one isolate significantly reduced *V. dahliae* microsclerotial area, but none affected *B. cinerea* sclerotia number (Fig. 4; Fig. S7).

Overall, the findings of the present study suggest significant antifungal potential of several *Bacillus* spp. isolates against major plant pathogenic fungi, through both diffusible and volatile-mediated compounds. The identification and the chemical characterization of the active compounds would be of great interest and it remains to be addressed in future work.

Genomics and functional analysis

Comparative genomics with OrthoFinder

OrthoFinder identified 12,162 orthogroups across the 25 *Bacilli* isolates, encompassing 120,736 genes (97.1% of the 124,385 genes in total), including 804 single-copy orthogroups (Table S3). The remaining 3,649 genes (2.9%) were unassigned (not classified into any orthogroup) and 253 isolate-specific orthogroups (681 genes, 0.5%) were also detected (Table S3).

Isolate-specific orthogroups served as a measure for assessing the genetic novelty across the 25 isolates (Fig. S8A). All three putative new species – SRL179, SRL337 and SRL543 – harboured high numbers of unique orthogroups. SRL543 had 8 isolate-specific

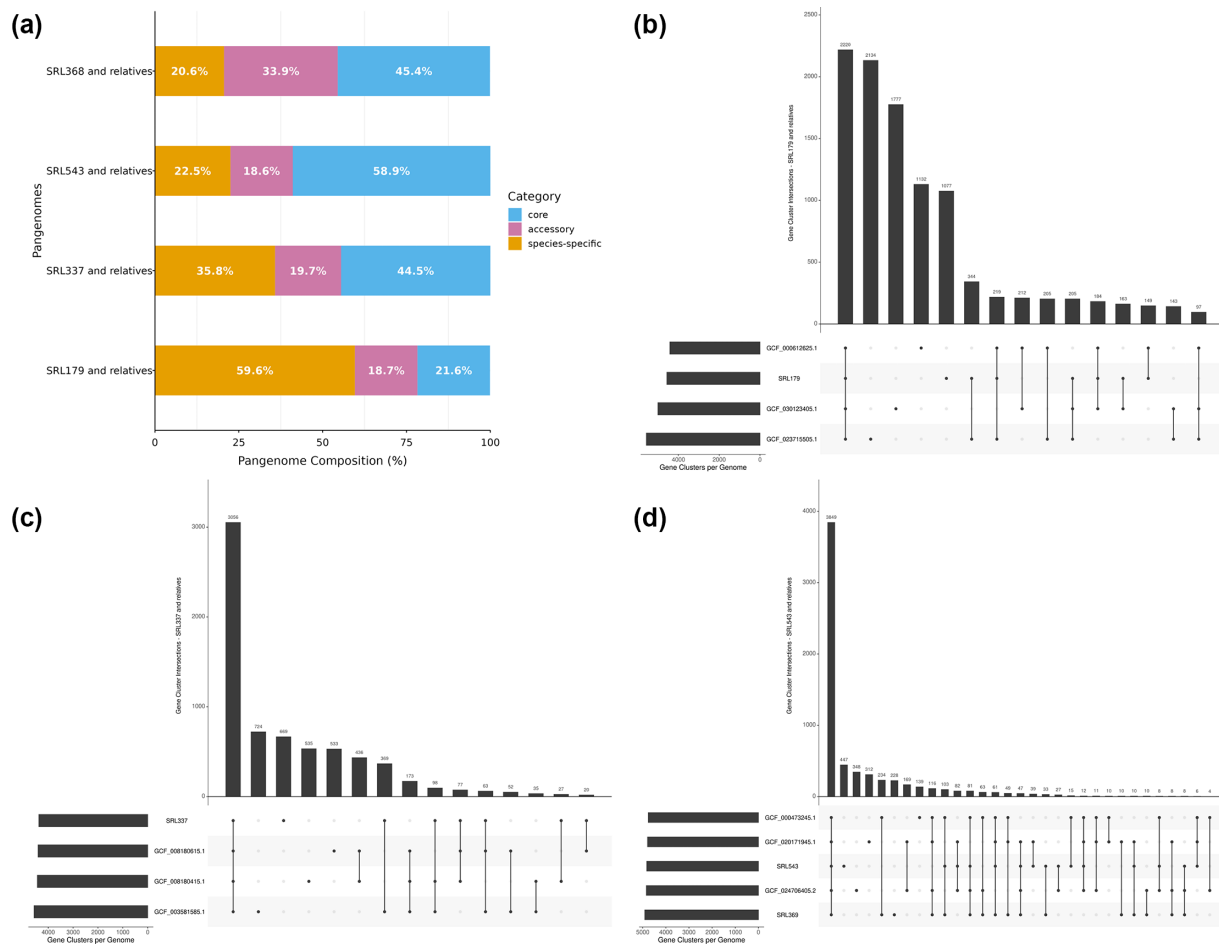


Fig. 5. Pangenome analysis of selected isolates with their closest relatives. (a) Distribution of core, accessory and species-specific genes (%) in the pangenome analyses of isolates SRL179, SRL337, SRL543 and SRL368 with their closest relatives. (b–d) UpSet plots for the pangenome analyses of the three putative new species: (b) SRL179, (c) SRL337 and (d) SRL543, with their closest relatives. In the case of SRL543, the *Bacillus infantis* isolate SRL369 was also identified as a close relative and included in the analysis. The UpSet plot for the pangenome analysis of the isolate SRL368 with its closest relatives is provided in Fig. S6. Horizontal bars (left) indicate the total number of gene clusters per genome, while vertical bars (top) display the number of shared clusters for each intersection (genome combination) indicated by the connected dots below.

orthogroups (18 genes, 0.4% of its genes in isolate-specific orthogroups); SRL337 had 27 (104 genes, 2.3%), the third highest among all isolates; and SRL179 had 48 (113 genes, 2.3%), the second highest number of isolate-specific orthogroups (Fig. S8A; SM 8). Notably, *B. thuringiensis* SRL368 exhibited the highest number of isolate-specific orthogroups: 105 orthogroups (293 genes; 4.3%) (Fig. S8A; SM 8). Despite SRL368 belonging to the same species as SRL215, SRL218 and SRL224, it harbours a markedly greater number of unique orthogroups, suggesting additional intra-species genomic divergence (Fig. S8A; SM 8).

In most cases, isolates with many isolate-specific orthogroups also showed higher proportions of unassigned genes (Fig. S8B). This trend may reflect greater genomic divergence or that certain lineage-specific genes are not captured in shared or isolate-specific orthogroups, highlighting the potential for even more undiscovered genetic novelty within *Bacilli*. In addition to 1,202 core orthogroups shared across all isolates, the number of accessory orthogroups (shared by two or more isolates) varied substantially (Table S3; Fig. S8C; SM 8).

Pangenome analysis of the three putative novel species and SRL368 with their closest relatives

Pangenome analysis enabled comparisons of the genetic features of the three putative new species and isolate SRL368 with their closest relatives. The pangenome of SRL179 and its three closest relatives contains 10,261 gene clusters. This pangenome shows extensive divergence among SRL179 and its closest relatives, with nearly 60% of the gene clusters being species-specific and only 21.6% constituting the core genome (Fig. 5a). SRL179 also harbours 1,077 species-specific gene clusters absent from the other genomes, underscoring its strong genomic uniqueness (Fig. 5b). The pronounced divergence between SRL179 and its relatives

highlights the potential of this clade – and SRL179 in particular – to encompass putatively novel functional and genetic features [90].

The pangenome of SRL337 and its three closest relatives includes 6,867 gene clusters. It also exhibits substantial divergence, with 35.8% species-specific and 44.5% core gene clusters (Fig. 5a). SRL337 possesses 669 species-specific gene clusters, the second highest among its relatives, highlighting pronounced differences in gene content relative to its closest relatives (Fig. 5c).

SRL543 was compared with its three closest publicly available relatives and the closely related *Bacillus infantis* SRL369. Its pangenome (6,539 gene clusters) includes 22.5% species-specific and 58.9% core genes (Fig. 5a). Interestingly, SRL543 has the most species-specific clusters (447) among its relatives, indicating a high degree of isolate-specific gene content (Fig. 5d).

In *B. thuringiensis* SRL368, 33.9% of clusters were accessory – more than in the other pangenomes – possibly suggesting a broader distribution of certain traits shared across separate *B. thuringiensis* strains and isolates (Fig. 5a). A total of 4,308 clusters (45.4%) formed the core genome, while 20.6% were species-specific (Fig. 5a). Notably, SRL368 contained the highest number of gene clusters and 1,154 uniquely shared clusters (the second-largest gene cluster intersection following the core genome) with the accession GCF_020775355.1, the genome with the second-highest number of gene clusters (Fig. S9). Additionally, SRL368 contained 15 unique clusters absent from all other genomes (Fig. S9). These results indicate that SRL368 harbours novel gene clusters and features not widely distributed among most of its relatives. Meanwhile, SRL215, SRL218 and SRL224 – from the same olive tree root sample – (Table 1; SM 5) – did not have species-specific gene clusters individually, but collectively shared 238 unique clusters absent from the other genomes (Fig. S9). These shared clusters may also encode novel features that warrant further investigation.

BGCs prediction

antiSMASH identified and annotated 312 BGCs across the 25 *Bacilli* isolates (SM 9). Isolates SRL215, SRL218, SRL224 and SRL368 (*B. thuringiensis*) and SRL662 had the highest BGC number (17) (Fig. S10; SM 9). Although SRL662 has a fragmented genome (Table 1; SM 9), which may inflate BGC predictions, its increased BGC count compared to SRL656 and SRL658, all *B. paralicheniformis* from the same *Matthiola tricuspidata* root system, suggests that subtle genetic, functional and biosynthetic differences can occur even within a single host microenvironment (Table 1; Fig. S10, SM 5, SM 9). Analogously, phenotypic and functional divergence has been reported among *B. licheniformis* isolates from the same source [91].

Mean BGC counts per genus revealed *Bacillus* and *Paenibacillus* as the most BGC-rich, followed by *Peribacillus* and then *Neobacillus*, *Rossellomorea* and *Cytobacillus* (Fig. S11), aligning with previous large-scale findings on *Bacillales* genomes [25, 92]. Among the eight BGC classes (SM 1), terpenes were the most frequent (23.7%) (Fig. S12; SM 9), highlighting their widespread role in the metabolite arsenal of *Bacilli* (SM 9). Known for their antimicrobial properties and potential applications against multidrug-resistant pathogens [93, 94], terpenes are naturally produced by *Bacilli* species which are promising microbial platforms for large-scale terpene production [95, 96]. The ‘others’ class was the second most frequent (19%), underscoring broad chemical diversity in *Bacilli* isolates, while NRPS (14.7%), PKS (12.5%) and RiPPs (11.2%) were also notably represented (SM 9).

AntiSMASH also assessed BGC similarity to known clusters (Figs 6, S10; SM 9). About 25% of the BGCs exhibited high similarity, mainly with NRPS-other hybrids, 5% medium similarity, mostly NI-siderophores (SM 9) and 11% low (mostly terpenes and betalactones) (SM 9). Notably, 59% showed no detectable similarity to any known cluster, highlighting a considerable novel biosynthetic potential (SM 9). Among these BGCs, terpenes were the most frequent, followed by NRPS and PKS clusters (SM 9), which are associated with antifungal biocontrol and plant-growth promotion [97]. These clusters underpin plant–microbe interactions and offer promising targets for further metabolic studies and biosynthetic engineering [97].

Comparisons between isolates of the same species revealed putative niche- or host-associated divergence and potential horizontal BGC transfer. In some cases, isolates showed highly conserved profiles. *B. velezensis* isolates (SRL163, SRL221, SRL244, SRL374, SRL379) from different hosts shared identical repertoires, enriched in terpenes, PKS and ‘other’ types (e.g. lanthipeptides, phosphonates). Notably, half of their clusters were of undefined similarity, highlighting novel biosynthetic potential despite any taxonomic and ecological homogeneity (Table 1; Figs 6, S10, S12; SM 9). Similarly, *B. infantis* SRL369 and SRL543, from distinct hosts, displayed identical profiles dominated by terpenes and one PKS cluster, despite SRL543 likely representing a novel species (Table 1; Figs 6, S10 and S12; SM 9). All but one of their BGCs had undefined similarity, positioning *B. infantis* and especially SRL543 as a promising candidate for further study (Figs 6, S10; SM 9).

In contrast, *B. thuringiensis* isolates possessed both the highest total and undefined-similarity BGC counts, indicating strong biosynthetic potential (SM 9; Fig. S10). SRL215, SRL218 and SRL224 (from olive roots) shared similar repertoires, whereas SRL368 (from *Frankenia hirsuta* leaves) showed distinct BGC diversity (Table 1; Figs 6, S10 and S12; SM 9). These differences may reflect host-associated structuring or isolate-specific acquisition of biosynthetic functions, a hypothesis that could be validated in the future using pipelines to detect putative horizontal gene transfer events.



Fig. 6. Distribution of BGCs across isolates by category and similarity confidence to known BGCs. Stacked bar plots showing the number of BGCs identified per isolate, grouped by similarity confidence levels to known BGCs (from top to bottom: High, Medium, Low, Undefined) as assigned by antiSMASH. BGCs are further grouped into eight categories: Ni-siderophores, NRPS, NRPS-PKS hybrids, NRPS-other hybrids, others, PKS, RiPPs and terpenes (see SM 1 – Supplementary Methods), indicated in the plot by different colours. 'Undefined similarity' refers to clusters without close matches to known BGCs in the antiSMASH database.

The putative novel species SRL179 and SRL337 also showed exceptional BGC novelty. All BGCs in SRL337 and all but one in SRL179 lacked similarity to known clusters (Figs 6, S10; SM 9). Both carried one PKS- and one RiPP-associated BGC; SRL179 also harboured a phosphonate BGC of undefined similarity and a low-similarity betalactone cluster (Figs 6, S12; SM 9). These chemically rich and phylogenetically distinct strains are strong candidates for future biosynthetic and functional studies [24, 98].

Overall, the 25 *Bacilli* isolates revealed extensive BGC diversity and novelty, with most clusters unrelated to known BGCs. *Bacillus* and *Paenibacillus* were the most BGC-rich, with *B. thuringiensis* standing out for both abundance and the presence of numerous uncharacterized clusters, including isolate-specific variation likely shaped by ecological factors. The putative new species SRL179 and SRL337 also exhibited pronounced divergence in their biosynthetic repertoires. Together, these findings underscore the widespread presence of previously uncharacterized BGCs among endophytic *Bacilli* and highlight the potential of both well-studied and novel lineages as reservoirs of unexplored secondary metabolites.

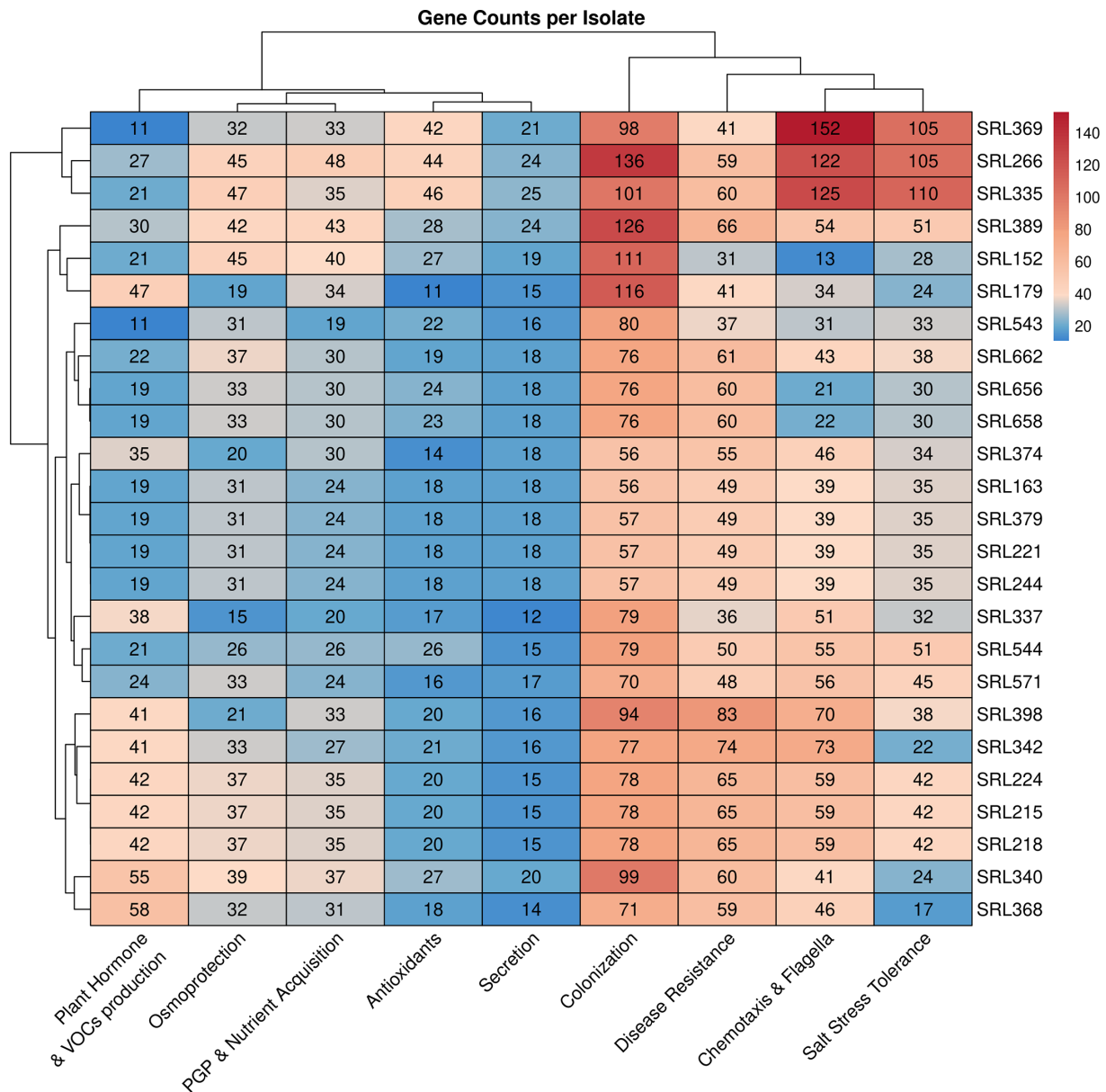


Fig. 7. Heatmap of gene counts associated with plant–microbe interactions across the 25 bacterial isolates. Heatmap showing the distribution of gene counts across nine plant–microbe associated functional categories: Plant hormone and VOCs production, osmoprotection, PGP and nutrient acquisition, antioxidants, secretion, colonization, disease resistance, chemotaxis and flagella and salt stress tolerance (see: BGCs, manual curation and genome mining). Rows represent isolates and columns represent functional categories. Numbers within cells indicate the gene counts, while the colour gradient reflects relative abundance. Hierarchical clustering is applied to both isolates (rows) and functional categories (columns).

Gene mining for secondary metabolites and genes involved in host–microbe interactions

The overall gene count distribution across nine functional categories revealed substantial functional diversity among the 25 isolates. Genes associated with salt and stress tolerance, chemotaxis, disease resistance and plant colonization were the most abundant and widely distributed.

Gene mining revealed that the closely related *B. thuringiensis* SRL215, SRL218 and SRL224 contained identical numbers of genes involved in host–microbe interactions and secondary metabolites, as did *B. velezensis* isolates SRL163, SRL221, SRL244 and SRL379 (Fig. 7). Combined with the phylogeny, these data indicate that these *B. thuringiensis* isolates are likely closely related isolates of the same initial microbe originated from the same plant. While the four *B. velezensis* isolates exhibit notable similarities, they originated from different plant samples (Table 1; SM 5). In contrast, *B. paralicheniformis* SRL662, from the same plant sample as SRL656 and SRL658, showed a distinct profile with markedly higher counts in secondary metabolite, chemotaxis and

salt-tolerance genes (Fig. 7). Differences between SRL662 and SRL656/658 were also evident in their BGC profiles, as discussed above. Notably, OrthoFinder did not differentiate these isolates through isolate-specific orthogroups (Fig. S8A), emphasizing that BGC prediction and manual genome mining can uncover diversity overlooked by orthogroup analysis alone.

P. frigiditolerans SRL266 exhibited stronger functional potential than other isolates, including those of the same species, especially in chemotaxis, plant colonization, salt stress tolerance and PGP- and nutrient-acquisition (Fig. 7). *B. infantis* SRL369 contained the most chemotaxis-related genes (152) and ranked among the top isolates in salt tolerance (105) and colonization (98), suggesting strong host-sensing capacity and adaptation (Fig. 7). Its close relative SRL543, likely a novel species, also had many colonization genes but fewer genes in other categories, highlighting a distinct functional profile. The inability of BGC profiles to differentiate these functional traits underscores the value of a more holistic approach that integrates genome mining with functional analysis. Among all isolates, *Cytobacillus* SRL335 contained the highest osmoprotection, salt stress and secretion gene counts and the second-highest Colonization count (Fig. 7). Its elevated salt stress tolerance gene count could be largely attributed to an abundance of heat shock protein genes, suggesting a probable physiological adaptation to salt stress (SM 10).

B. thuringiensis SRL368, with the most isolate-specific orthogroups also ranked high in plant hormone and VOC production genes (Figs 7, S8A), followed by SRL340 (Fig. 7). The putative novel species SRL179 (second highest number of isolate-specific orthogroups) showed enrichment in plant colonization and hormone/VOC genes, whereas the new species SRL337 (third highest number of isolate-specific orthogroups) had the fewest osmoprotection and secretion genes (Figs 7, S8A). *Paenibacillus* SRL342 and SRL398 carried many siderophore and iron-transport genes (disease resistance category), suggesting disease-protective potential for their host, and SRL342 also had the most antioxidants and spermidine synthesis genes (Fig. 7; SM 10). These two isolates, both extracted from the same *Atriplex halimus* root sample, were isolated using different culture media (Table 1; SM 5). Their subtle differences may reflect microdiversity arising even within a single host microhabitat and/or culture-medium effects resulting from repeated subculturing of a diverse population of isolates of the same species in different media.

Interestingly, gene mining uncovered genes involved in the synthesis of enzymes and compounds associated with antifungal activity (e.g. CLPs, such as surfactin, iturin and fengycin, as well as chitinases), consistent with the antifungal effects observed in the *in vitro* dual-culture assays (Fig. 4, SM 10). Gene mining further revealed a broad distribution among the 25 isolates of genes associated with VOC biosynthesis: sulphur VOCs (methanethiol, dimethyl disulfide, dimethyl trisulfide) known for their antifungal and biocontrol activities [99] and alcohol/ketone VOCs (acetoin, 2,3-butanediol, diacetyl) and their precursors, known for fungal pathogen inhibition and PGP [100] (SM 10). These genomic features are consistent with the volatile-mediated antifungal effects observed in the dual-plate bioassays (Figs 4, S5).

CONCLUDING REMARKS

The genomic, phylogenetic and functional characterization of 25 endophytic *Bacilli* isolates from plants with different lifestyles revealed striking novelty across multiple levels. The phylogenomic analysis placed the isolates within a wide diversity of *Bacilli* lineages, spanning *Bacillus*, *Paenibacillus*, *Peribacillus*, *Neobacillus*, *Cytobacillus* and *Rosellomorea*. The reconstruction of the phylogenetic inference revealed a clear separation of several isolates, supporting the delineation of three novel species and expanding the known diversity of plant-associated *Bacilli*.

BGC mining uncovered 312 BGCs, nearly 60% of which were novel, highlighting an exceptional and unexplored reservoir of secondary metabolites. The high proportion of BGCs showing no or very low similarity to known clusters suggests that such unique and novel biosynthetic repertoires may constitute a distinctive adaptive feature of these endophytic *Bacilli* and could also indicate a distinctive trait of endophytes in general, facilitating their colonization and interaction with host plants. Similarly, genome mining identified the presence of diverse genes linked to colonization and stress tolerance reflecting ecological strategies that enable *Bacilli* to thrive in saline and nutrient-limited habitats. Bioassays validated the high-salinity tolerance and strong antagonism against diverse phytopathogens, underscoring their potential as bioinoculants in agrifood.

Our study demonstrates that *Bacilli* endophytes from wild and halophytic plants combine taxonomic novelty with functional and biosynthetic innovation. Together, these findings underscore the dual importance of *Bacilli* endophytes: as promising resources for further investigation of their potential roles in crop resilience and biocontrol and model systems for advancing our understanding of microbial adaptation, symbiosis and the evolution of secondary metabolism.

Funding information

N.P.A. was supported by a PhD scholarship of the Research Funding Programme of the Research Committee of the University of Crete managed through the Financial and Administrative Support Unit of the Special Account for Research Funds (ELKE) of the University of Crete (No. 11643). C.A.C., S.P. and S.S. were supported by the Biosolutions project 'Sustainable solutions to address the effects of climate change in inducing crop tolerance to biotic and abiotic stresses' funded by Prasino Tameio (1721/23-03-2023). For parts of their work C.A.C., S.S. and N.P.A. were supported by the fund RESEARCH-CREATE-INNOVATE (Grant code/Acronym: T2EAK-01859/BIOCONTROL) cofinanced by the European Union and Greek national funds through the Operational Programme Competitiveness, Entrepreneurship and Innovation. F.R. and A.S. are funded by the European Union (EU) under Grant Agreement No 101087081 (CompBiodivGR).

Acknowledgements

We thank Professor A.M. Eren for his valuable help and insightful suggestions regarding the anvi'o pangenome analysis pipeline. *Ralstonia solanacearum* GMI1000 was kindly provided by Professor Nemo Peeters (INRA CNRS Research Unit, France), *Clavibacter michiganensis* HMU4521 was kindly provided by Professor Dimitris Goumas (Collection of the Department of Agriculture, Hellenic Mediterranean University, HMU) and *Paracitovorax citrulli* (formerly *Acidovorax citrulli*) ATCC29625 was kindly provided by Professor Ronald Walcott (Department of Plant Pathology, University of Georgia).

Author contributions

Conceptualization: P.F.S. and C.A.C. Data curation: P.F.S., S.P., N.P.A., C.A.C. and S.S. Formal analysis: P.F.S., N.P.A., C.A.C., S.P., S.S. and F.R. Funding acquisition: P.F.S. Investigation: N.P.A., C.A.C., S.P., S.S., C.P. and E.A. Methodology: C.A.C., N.P.A., S.P., S.S., F.R. and P.F.S. Project administration: P.F.S., S.P. and C.A.C. Resources: P.F.S., E.A.M. and A.S. Software: N.P.A., S.P., F.R. and C.A.C. Supervision: P.F.S., E.A.M. and A.S. Validation: P.F.S., C.A.C., N.P.A. and S.P. Visualization: N.P.A., S.P., F.R. and S.S. Writing – original draft: N.P.A., P.F.S., S.P., C.A.C., S.S., F.R., E.A.M. and A.S. Writing – review and editing: All authors reviewed and edited the manuscript.

Conflicts of interest

The author(s) declare that there are no conflicts of interest.

References

- Ingrao C, Strippoli R, Lagioia G, Huisinigh D. Water scarcity in agriculture: an overview of causes, impacts and approaches for reducing the risks. *Heliyon* 2023;9:e18507.
- Sünnemann M, Beugnon R, Breitreuz C, Buscot F, Cesarz S, et al. Climate change and cropland management compromise soil integrity and multifunctionality. *Commun Earth Environ* 2023;4:394.
- FAO, IFAD, UNICEF, WFP, WHO. The state of food security and nutrition in the world 2025 – addressing high food price inflation for food security and nutrition. *Rome* 2025.
- Königer J, Panagos P, Jones A, Briones MJ, Orgiazzi A. In defence of soil biodiversity: towards an inclusive protection in the European Union. *Biol Conserv* 2022;268:109475.
- Poppeliers SW, Sánchez-Gil JJ, de Jonge R. Microbes to support plant health: understanding bioinoculant success in complex conditions. *Curr Opin Microbiol* 2023;73:102286.
- Reinhold-Hurek B, Hurek T. Living inside plants: bacterial endophytes. *Curr Opin Plant Biol* 2011;14:435–443.
- Santoyo G, Moreno-Hagelsieb G, Orozco-Mosqueda M del C, Glick BR. Plant growth-promoting bacterial endophytes. *Microbiol Res* 2016;183:92–99.
- Afzal I, Shinwari ZK, Sikandar S, Shahzad S. Plant beneficial endophytic bacteria: mechanisms, diversity, host range and genetic determinants. *Microbiol Res* 2019;221:36–49.
- Christakis CA, Daskalogiannis G, Chatzaki A, Markakis EA, Mermigka G, et al. Endophytic bacterial isolates from halophytes demonstrate phytopathogen biocontrol and plant growth promotion under high salinity. *Front Microbiol* 2021;12:681567.
- Etesami H, Beattie GA. Mining halophytes for plant growth-promoting halotolerant bacteria to enhance the salinity tolerance of non-halophytic crops. *Front Microbiol* 2018;9:148.
- Llorens E, Sharon O, Camañes G, García-Agustín P, Sharon A. Endophytes from wild cereals protect wheat plants from drought by alteration of physiological responses of the plants to water stress. *Environ Microbiol* 2019;21:3299–3312.
- Wu W, Chen W, Liu S, Wu J, Zhu Y, et al. Beneficial relationships between endophytic bacteria and medicinal plants. *Front Plant Sci* 2021;12:646146.
- Blake C, Christensen MN, Kovács ÁT. Molecular aspects of plant growth promotion and protection by *Bacillus subtilis*. *Mol Plant Microbe Interact* 2021;34:15–25.
- Ragasruthi M, Balakrishnan N, Murugan M, Swarnakumari N, Harish S, et al. *Bacillus thuringiensis* (Bt)-based biopesticide: navigating success, challenges, and future horizons in sustainable pest control. *Sci Total Environ* 2024;954:176594.
- Siddika A, Rashid AA, Khan SN, Khatun A, Karim MM. Harnessing plant growth-promoting rhizobacteria, *Bacillus subtilis* and *B. aryabhattai* to combat salt stress in rice: a study on the regulation of antioxidant defense, ion homeostasis, and photosynthetic parameters. *Front Plant Sci* 2024;15.
- Nester EW, Thomashow LS, Metz M, Gordon M. 100 years of *Bacillus thuringiensis*: a critical scientific assessment. *Am Soc Microbiol* 2002.
- Priest FG, Goodfellow M, Todd C. A numerical classification of the genus *Bacillus*. *Microbiology* 1988;134:1847–1882.
- Priest FG, Barker M, Baillie LWJ, Holmes EC, Maiden MCJ. Population structure and evolution of the *Bacillus cereus* group. *J Bacteriol* 2004;186:7959–7970.
- Ash C, Farrow JAE, Wallbanks S, Collins MD. Phylogenetic heterogeneity of the genus *Bacillus* revealed by comparative analysis of small-subunit-ribosomal RNA sequences. *Lett Appl Microbiol* 1991;13:202–206.
- Kunst F, Ogasawara N, Moszer I, Albertini AM, Alloni G, et al. The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*. *Nature* 1997;390:249–256.
- Fajardo-Cavazos P, Maughan H, Nicholson WL. Evolution in the *Bacillaceae*. *Microbiol Spectr* 2014;2.
- Patel S, Gupta RS. 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. and *Alkalihalobacillus* gen. nov. *Int J Syst Evol Microbiol* 2020;70:406–438.
- Logan NA, Berkeley RCW. Identification of *Bacillus* strains using the API system. *J Gen Microbiol* 1984;130:1871–1882.
- Xia L, Miao Y, Cao A, Liu Y, Liu Z, et al. Biosynthetic gene cluster profiling predicts the positive association between antagonism and phylogeny in *Bacillus*. *Nat Commun* 2022;13:1023.
- Song L, Nielsen LJD, Xu X, Mohite OS, Nuhamunada M, et al. Expanding the genome information on Bacillales for biosynthetic gene cluster discovery. *Sci Data* 2024;11:1267.
- Jiang W, Jiang B-L, Xu R-Q, Huang J-D, Wei H-Y, et al. Identification of six type III effector genes with the PIP box in *Xanthomonas campestris* pv. *campestris* and five of them contribute individually to full pathogenicity. *Mol Plant Microbe Interact* 2009;22:1401–1411.
- Markakis EA, Ligoixakis EK, Avramidou EV, Tzanidakis N. Survival, persistence, and infection efficiency of *Verticillium dahliae* passed through the digestive system of sheep. *Plant Dis* 2014;98:1235–1240.
- Vakalounakis DJ, Doulis AG, Klironomou E. Characterization of *Fusarium oxysporum* f. sp. *radicis-cucumerinum* attacking melon under natural conditions in Greece. *Plant Pathol* 2005;54:339–346.
- Malandrakis AA, Krasagakis N, Kavroulakis N, Ilias A, Tsagkarakou A, et al. Fungicide resistance frequencies of *Botrytis cinerea* greenhouse isolates and molecular detection of a novel SDHI resistance mutation. *Pestic Biochem Physiol* 2022;183:105058.
- Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* 2012;9:671–675.

31. R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria; 2024. <https://www.R-project.org>
32. Wickham H. *Ggplot2: Elegant Graphics for Data Analysis*. Springer New York, 2016.
33. Wickham H, Averick M, Bryan J, Chang W, McGowan L, et al. Welcome to the Tidyverse. *JOSS* 2019;4:1686.
34. Kiledal A, A Maresca J. Chromosomal DNA extraction from Gram-positive bacteria v1. 2021. DOI: 10.17504/protocols.io.bs7knkhk.
35. Yilmaz P, Kottmann R, Field D, Knight R, Cole JR, et al. Minimum information about a marker gene sequence (MIMARKS) and minimum information about any (x) sequence (MlxS) specifications. *Nat Biotechnol* 2011;29:415–420.
36. Chen S. Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp. *iMeta* 2023;2.
37. Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput Biol* 2017;13:e1005595.
38. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol* 2019;37:540–546.
39. Mikheenko A, Pribelski A, Saveliev V, Antipov D, Gurevich A. Versatile genome assembly evaluation with QUAST-LG. *Bioinformatics* 2018;34:i142–i150.
40. Seppey M, Manni M, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness. In: *Gene Prediction: Methods and Protocols*. New York, NY: Springer New York, . pp. 227–245.
41. Zafeiropoulos H, Gioti A, Ninidakis S, Potirakis A, Paragkamian S, et al. 0s and 1s in marine molecular research: a regional HPC perspective. *Gigascience* 2021;10:giab053.
42. Schwengers O, Jelonek L, Dieckmann MA, Beyvers S, Blom J, et al. Bakta: rapid and standardized annotation of bacterial genomes via alignment-free sequence identification. *Microb Genom* 2021;7:000685.
43. van der Graaf-van Bloois L, Wagenaar JA, Zomer AL. RFPlasmid: predicting plasmid sequences from short-read assembly data using machine learning. *Microb Genom* 2021;7.
44. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol* 2019;20:238.
45. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk v2: memory friendly classification with the genome taxonomy database. *Bioinformatics* 2022;38:5315–5316.
46. Parks DH, Chuvochina M, Rinke C, Mussig AJ, Chaumeil P-A, et al. GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. *Nucleic Acids Res* 2022;50:D785–D794.
47. Edgar RC. Muscle5: high-accuracy alignment ensembles enable unbiased assessments of sequence homology and phylogeny. *Nat Commun* 2022;13:6968.
48. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol* 2020;37:1530–1534.
49. Zhang C, Rabiee M, Sayyari E, Mirarab S. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics* 2018;19:153.
50. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat Commun* 2018;9:5114.
51. Kapli P, Lutteropp S, Zhang J, Kobert K, Pavlidis P, et al. Multi-rate Poisson tree processes for single-locus species delimitation under maximum likelihood and Markov chain Monte Carlo. *Bioinformatics* 2017;33:1630–1638.
52. Bianchini G, Sánchez-Baracaldo P. TreeViewer: flexible, modular software to visualise and manipulate phylogenetic trees. *Ecol Evol* 2024;14:e10873.
53. Blin K. ncbi-genome-download (Version 0.3.3) [Computer software]; 2023. <https://github.com/kblin/ncbi-genome-download>
54. O'Leary NA, Cox E, Holmes JB, Anderson WR, Falk R, et al. Exploring and retrieving sequence and metadata for species across the tree of life with NCBI Datasets. *Sci Data* 2024;11:732.
55. Eren AM, Kiehl E, Shaiber A, Veseli I, Miller SE, et al. Community-led, integrated, reproducible multi-omics with anvi'o. *Nat Microbiol* 2020;6:3–6.
56. Delmont TO, Eren AM. Linking pangenomes and metagenomes: the *Prochlorococcus* metapangenome. *PeerJ* 2018;6:e4320.
57. Eddy SR. Accelerated profile HMM searches. *PLoS Comput Biol* 2011;7:e1002195.
58. Galperin MY, Wolf YI, Makarova KS, Vera Alvarez R, Landsman D, et al. COG database update: focus on microbial diversity, model organisms, and widespread pathogens. *Nucleic Acids Res* 2021;49:D274–D281.
59. Kanehisa M, Furumichi M, Sato Y, Ishiguro-Watanabe M, Tanabe M. KEGG: integrating viruses and cellular organisms. *Nucleic Acids Res* 2021;49:D545–D551.
60. Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, et al. Pfam: the protein families database in 2021. *Nucleic Acids Res* 2021;49:D412–D419.
61. Parks DH, Chuvochina M, Waite DW, Rinke C, Skarshewski A, et al. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat Biotechnol* 2018;36:996–1004.
62. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. *Nat Methods* 2015;12:59–60.
63. Conway JR, Lex A, Gehlenborg N. UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics* 2017;33:2938–2940.
64. Blin K, Shaw S, Vader L, Szenei J, Reitz ZL, et al. antiSMASH 8.0: extended gene cluster detection capabilities and analyses of chemistry, enzymology, and regulation. *Nucleic Acids Res* 2025;53:W32–W38.
65. Chen I-MA, Chu K, Palaniappan K, Pillay M, Ratner A, et al. IMG/M v5.0: an integrated data management and comparative analysis system for microbial genomes and microbiomes. *Nucleic Acids Res* 2019;47:D666–D677.
66. Kanehisa M, Goto S. K. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res* 2000;28:27–30.
67. Letunic I, Khedkar S, Bork P. SMART: recent updates, new developments and status in 2020. *Nucleic Acids Res* 2021;49:D458–D460.
68. Haft DH. TIGRFAMs: a protein family resource for the functional identification of proteins. *Nucleic Acids Res* 2001;29:41–43.
69. Pandurangan AP, Stahlhacke J, Oates ME, Smithers B, Gough J. The SUPERFAMILY 2.0 database: a significant proteome update and a new webserver. *Nucleic Acids Res* 2019;47:D490–D494.
70. Sillitoe I, Bordin N, Dawson N, Waman VP, Ashford P, et al. CATH: increased structural coverage of functional space. *Nucleic Acids Res* 2021;49:D266–D273.
71. De Vrieze M, Varadarajan AR, Schneeberger K, Bailly A, Rohr RP, et al. Linking comparative genomics of nine potato-associated *Pseudomonas* isolates with their differing biocontrol potential against late blight. *Front Microbiol* 2020;11.
72. Eida AA, Bougouffa S, L'Haridon F, Alam I, Weisskopf L, et al. Genome insights of the plant-growth promoting Bacterium *Cronobacter muytjensii* JZ38 with volatile-mediated antagonistic activity against phytophthora infestans. *Front Microbiol* 2020;11.
73. Zamanzadeh-Nasrabadi SM, Mohammadiapanah F, Hosseini-Mazinani M, Sarikhan S. Salinity stress endurance of the plants with the aid of bacterial genes. *Front Genet* 2023;14.
74. Adedayo AA, Fadiji AE, Babalola OO. Unraveling the functional genes present in rhizosphere microbiomes of *Solanum lycopersicum*. *PeerJ* 2023;11:e15432.

75. Jiang L, Seo J, Peng Y, Jeon D, Park SJ, et al. Genome insights into the plant growth-promoting bacterium *Saccharibacillus brassicae* ATSA2^T. *AMB Express* 2023;13:9.
76. Harwood CR, Cranenburgh R. Bacillus protein secretion: an unfolding story. *Trends Microbiol* 2008;16:73–79.
77. Bruto M, Prigent-Combaret C, Muller D, Moënné-Loccoz Y. Analysis of genes contributing to plant-beneficial functions in plant growth-promoting *Rhizobacteria* and related *Proteobacteria*. *Sci Rep* 2014;4:6261.
78. Kolde R. pheatmap: pretty heatmaps. R package version 1.0.13; 2025. <https://cran.r-project.org/web/packages/pheatmap/index.html>
79. Pedersen TL. patchwork: the composer of plots. R package version 1.3.1; 2025. <https://cran.r-project.org/web/packages/patchwork/index.html> [accessed 12 September 2025].
80. Wilke CO. cowplot: streamlined plot theme and plot annotations for 'ggplot2'. R package version 1.1.3; 2025. <https://cran.r-project.org/web/packages/cowplot/index.html> [accessed 12 September 2025].
81. Fu X, Gong L, Liu Y, Lai Q, Li G, et al. *Bacillus pumilus* group comparative genomics: toward pangenome features, diversity, and marine environmental adaptation. *Front Microbiol* 2021;12:571212.
82. Das VA, Gautam B, Yadav PK, Varadwaj PK, Wadhwa G, et al. Computational approach to identify novel genomic features conferring high fitness in *Bacillus atrophaeus* CNY01 and *Bacillus velezensis* AK-0 associated with plant growth promotion (PGP) in apple. *BMC Plant Biol* 2024;24:1127.
83. Ramasamy KP, Mahawar L. Coping with salt stress-interaction of halotolerant bacteria in crop plants: a mini review. *Front Microbiol* 2023;14:1077561.
84. Karačić V, Miljaković D, Marinković J, Ignjatov M, Milošević D, et al. *Bacillus* species: excellent biocontrol agents against tomato diseases. *Microorganisms* 2024;12:457.
85. Bouchard-Rochette M, Machrafi Y, Cossus L, Thuy An Nguyen T, Antoun H, et al. *Bacillus pumilus* PTB180 and *Bacillus subtilis* PTB185: production of lipopeptides, antifungal activity, and biocontrol ability against *Botrytis cinerea*. *Biol Control* 2022;170:104925.
86. Etesami H, Jeong BR, Glick BR. Biocontrol of plant diseases by *Bacillus* spp. *Physiol Mol Plant Pathol* 2023;126:102048.
87. Jiang C-H, Liao M-J, Wang H-K, Zheng M-Z, Xu J-J, et al. *Bacillus velezensis*, a potential and efficient biocontrol agent in control of pepper gray mold caused by *Botrytis cinerea*. *Biol Control* 2018;126:147–157.
88. Chen H, Xiao X, Wang J, Wu L, Zheng Z, et al. Antagonistic effects of volatiles generated by *Bacillus subtilis* on spore germination and hyphal growth of the plant pathogen, *Botrytis cinerea*. *Biotechnol Lett* 2008;30:919–923.
89. Massawe VC, Hanif A, Farzand A, Mburu DK, Ochola SO, et al. Volatile compounds of endophytic *Bacillus* spp. have biocontrol activity against *Sclerotinia sclerotiorum*. *Phytopathology* 2018;108:1373–1385.
90. Kim Y, Koh I, Young Lim M, Chung W-H, Rho M. Pan-genome analysis of *Bacillus* for microbiome profiling. *Sci Rep* 2017;7:10984.
91. Idelevich EA, Pogoda CA, Ballhausen B, Wüllenweber J, Eckardt L, et al. Pacemaker lead infection and related bacteraemia caused by normal and small colony variant phenotypes of *Bacillus licheniformis*. *J Med Microbiol* 2013;62:940–944.
92. Yin Q-J, Ying T-T, Zhou Z-Y, Hu G-A, Yang C-L, et al. Species-specificity of the secondary biosynthetic potential in *Bacillus*. *Front Microbiol* 2023;14:1271418.
93. El Fannassi Y, Gharsallaoui A, Khelissa S, El Amrani MA, Suisse I, et al. Complexation of terpenes for the production of new antimicrobial and antibiofilm molecules and their encapsulation in order to improve their activities. *Appl Sci* 2023;13:9854.
94. Liu X, Xin J, Sun Y, Zhao F, Niu C, et al. Terpenoids from marine sources: a promising avenue for new antimicrobial drugs. *Mar Drugs* 2024;22:347.
95. Caulier S, Nannan C, Gillis A, Licciardi F, Bragard C, et al. Overview of the antimicrobial compounds produced by members of the *Bacillus subtilis* group. *Front Microbiol* 2019;10:302.
96. Pramastya H, Song Y, Elfahmi EY, Sukrasno S, Quax WJ. Positioning *Bacillus subtilis* as terpenoid cell factory. *J Appl Microbiol* 2021;130:1839–1856.
97. Zheng Y, Liu T, Wang Z, Wang X, Wang H, et al. Whole-genome sequencing and secondary metabolite exploration of the novel *Bacillus velezensis* BN with broad-spectrum antagonistic activity against fungal plant pathogens. *Front Microbiol* 2025;15.
98. Grubbs KJ, Bleich RM, Santa Maria KC, Allen SE, Farag S, et al. Large-scale bioinformatics analysis of *Bacillus* genomes uncovers conserved roles of natural products in bacterial physiology. *mSystems* 2017;2:e00040-17.
99. Ballot A, Dore J, Rey M, Meiffren G, Langin T, et al. Dimethylpolysulfides production as the major mechanism behind wheat fungal pathogen biocontrol, by *Arthrobacter* and *Microbacterium* actinomycetes. *Microbiol Spectr* 2023;11:e0529222.
100. Wu Y, Zhou J, Li C, Ma Y. Antifungal and plant growth promotion activity of volatile organic compounds produced by *Bacillus amyloliquefaciens*. *Microbiologyopen* 2019;8:e00813.

The Microbiology Society is a membership charity and not-for-profit publisher.

Your submissions to our titles support the community – ensuring that we continue to provide events, grants and professional development for microbiologists at all career stages.

Find out more and submit your article at microbiologyresearch.org