

Description of *Brachybacterium manguinensis* sp. nov., a novel species isolated from pharmaceutical production environment in Rio de Janeiro, Brazil

Nathalia Gonçalves Santos Caldeira

National Institute of Quality Control in Health, Oswaldo Cruz Foundation

Juliana Nunes Ramos

Fundação Oswaldo Cruz

Maria Luiza Soares de Souza

National Institute of Quality Control in Health, Oswaldo Cruz Foundation

Brena de Jesus Silva

Fundação Oswaldo Cruz

Rebeca Vitória da Silva Lage de Miranda

Fundação Oswaldo Cruz

Talita Bernardo Valadão

Fundação Oswaldo Cruz

Stephen James Forsythe

Foodmicrobe.com Ltd

Bruno Vicente

Fundação Oswaldo Cruz

Victor Midlej

Fundação Oswaldo Cruz

Luciana Veloso da Costa

luciana.costa@bio.fiocruz.br

Fundação Oswaldo Cruz

Viviane Zahner

Fundação Oswaldo Cruz

Marcelo Luiz Lima Brandão

Fundação Oswaldo Cruz

Research Article

Keywords: Brachybacterium, 16S rRNA, MALDI-TOF MS, novel taxon, genomic taxonomy

Posted Date: October 16th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10884953/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

A Gram-positive bacterial isolate, designated B718/18, was recovered from a pharmaceutical production environment in Rio de Janeiro, Brazil, during viable airborne particle monitoring. Identification using VITEK® 2 and two MALDI-TOF MS platforms was unsuccessful, suggesting that the isolate was not represented in current commercial databases. Full-length 16S rRNA gene sequencing assigned B718/18 to the genus *Brachybacterium*, showing 98.68% sequence similarity to *Brachybacterium sacelli* DSM 14566T. Phylogenetic and phylogenomic analyses consistently placed the isolate within a distinct, well-supported lineage. Whole-genome sequencing yielded a 3.31 Mb genome with a G + C content of 72.4%. Comparative genomic analyses showed low relatedness to all recognized *Brachybacterium* species, with ANI values ranging from 74.1% to 81.0% and dDDH values from 20.1% to 25.7%, below the accepted species delineation thresholds. Phenotypically, B718/18 produced yellow colonies and consisted of non-motile, catalase-positive, oxidase-negative coccobacilli with a distinctive biochemical profile. The isolate did not form biofilms on the surfaces tested. Genome mining identified genes potentially associated with stress adaptation and antimicrobial resistance, as well as putative biosynthetic gene clusters, including a terpene-related cluster potentially involved in carotenoid production. Based on the combined phylogenetic, genomic, and phenotypic evidence, B718/18 was proposed as the type strain of the novel species *Brachybacterium manguinensis*. The strain was deposited in the Fiocruz CBAS and CBP collections, and its 16S rRNA and complete genome sequences were deposited under accession numbers PV425877 and NZ_JBVDDO000000000, respectively.

Introduction

The genus *Brachybacterium* belongs to the class Actinomycetes, order *Micrococcales*, and family *Dermabacteriaceae*, according to List of Prokaryotic names in Standing in Nomenclature (LPSN) (Freese et al. 2025). It was originally described by Collins, Brown, and Jones (1988), with *Brachybacterium faecium* as the type species, based on biochemical and phenotypic characteristics and DNA hybridization studies performed on strains isolated from poultry deep litter (Collins et al. 1988). Currently, the genus comprises 28 species with validly published names; however, when considering species described in publications that have not yet been validated, this number can reach approximately 36 species, in addition to two invalid subspecies (<https://lpsn.dsmz.de/genus/brachybacterium>; accessed on July 06, 2026).

Brachybacterium species are Gram-positive, aerobic, non-motile, and non-spore-forming bacteria, exhibiting morphological variation associated with a rod-coccus cycle. During the exponential growth phase, irregular rod-shaped or corineform cells predominate, while in the stationary phase, rounded coccoid cells are more frequent. From a chemotaxonomic point of view, the genus is characterized by meso-diaminopimelic –D-Glu₂ (A4y variation), predominant menaquinone MK-7, absence of mycolic acids, and DNA G + C content ranging from 68 to 72 mol% (Takeuchi et al. 1995).

Species of this genus have been isolated from a wide variety of environments, including animal organs, plant associations, soils, lakes, coastal marine waters, and different food matrices, such as dairy products, seafood, and corn liquor (Hlaing et al. 2019; de Castro et al. 2023; Farooq et al. 2023). In addition, the isolation of *Brachybacterium nesterenkovi* from human blood culture highlighted the potential of this genus as an opportunistic pathogen (Brummaier et al. 2020).

Given the wide environmental distribution of *Brachybacterium* species and the ability of environmental Gram-positive bacteria to persist under adverse conditions, their occurrence in pharmaceutical production environments cannot be ignored. The presence of microorganisms in samples collected along a pharmaceutical production chain (as sterile intermediary and final products) can represent risk to product safety and quality. In this context, quality control methods must be capable of detecting and identifying contaminants from diverse sources, particularly in critical areas (grades A and B), media fill and sterility tests (WHO, 2011; EMA, 2022). The accurate identification of microorganisms isolated from less critical areas (grades C and D) and bioburden assay is also essential for investigating contamination sources and implementing effective preventive and corrective actions, especially for organisms that are difficult to control, such as spore-forming bacteria, filamentous fungi, biofilm producers, or degradative enzyme producers (Moreira et al. 2023; Caldeira et al. 2024; Mattoso et al. 2025; Vasconcellos et al. 2026). The formation of biofilms by bacterial communities in pharmaceutical production environments represents a major challenge for environmental control. The extracellular matrix associated with biofilms increases bacterial tolerance to cleaning and disinfection processes (CIP), enabling prolonged persistence on surfaces, equipment, and water distribution systems, thereby contributing to environmental contamination (Parental Drug Association 2015; Sharma et al. 2023; Silva-Santana et al. 2025).

Therefore, identification methodologies must combine accuracy and speed, as batch release depends on timely and reliable results. Although Matrix-Assisted Laser Desorption Ionization–Time of Flight Mass Spectrometry (MALDI-TOF MS) offers significant advantages over traditional phenotypic methods, its application remains limited by incomplete databases that insufficiently represent the microbial diversity of pharmaceutical production environments (Miranda et al. 2023; Caldeira et al. 2024; Ramos et al. 2025).

Comparison of 16S rRNA gene sequences is widely used for placing strains within a genus and for assessing similarity with described species; however, this marker often has limited resolution for discriminating phylogenetically closely related species. In these situations, the use of additional housekeeping genes and, above all, analyses based on complete genomes have proven essential (Costa et al. 2022a,b; Farooq et al. 2023).

The distinction between species of *Brachybacterium* is based on the integrated analysis of phenotypic, physiological, and chemotaxonomic characteristics, including the use of different sugars and organic compounds, hydrolytic and enzymatic activities, growth patterns under different conditions of temperature, pH, salinity, and oxygenation, as well as morphological and pigmentation variations

observed throughout growth. Additionally, the analysis of cellular fatty acid profiles, polar lipids, and phospholipids, as well as cell wall composition, including peptidoglycan amino acids and specific sugars, is a relevant tool for taxonomic discrimination, with G + C content being a complementary parameter that, alone, is not sufficient for species definition (Gvozdyak et al. 1992; Takeuchi et al. 1995; Farooq et al. 2023).

In recent decades, advances in next-generation sequencing (NGS), together with an improved understanding of genomic signatures and the development of robust bioinformatic tools for data analysis, have greatly expanded the study of microbial diversity. This progress has led to the description of novel taxa and the reclassification of previously described bacterial species. In this scenario, methods based on whole-genome comparisons have been widely applied, including metrics such as Average Nucleotide Identity (ANI), digital DNA–DNA hybridization (dDDH), and phylogenomic approaches based on conserved genes and core genome proteins. These strategies offer higher resolution for species delineation and are increasingly recognized as the standard in modern bacterial taxonomy (Riesco and Trujillo 2024).

This study reports a bacterial isolate (B718/18) recovered from a pharmaceutical industry, which was initially identified as *Brachybacterium* sp. by MALDI-TOF MS and *16S rRNA* gene sequencing. Based on this preliminary identification, the objective of this study was to perform a detailed taxonomic characterization of this isolate using a polyphasic approach, aiming to describe a new species.

Material and methods

Origin of bacterial isolate

The bacterial isolate B718/18 was obtained in 2018 through viable airborne particle counting in a pharmaceutical facility producing immunobiologicals, in Rio de Janeiro, Brazil. Initial phenotypic characterization of the isolate was performed using the VITEK®2 system (bioMérieux, Craponne, France) for biochemical analysis and the VITEK® MS system (bioMérieux, Craponne, France) for proteomic identification by MALDI-TOF MS, both according to the manufacturer's instructions. However, neither method was able to provide a species level identification of the isolate. Subsequently, B718/18 was stored in skim milk containing 30% glycerol at -70°C. In 2022, the isolate was cultured on Sheep Blood Agar (Biocen, São Paulo, Brazil) to allow a new attempt at identification.

Re-analysis by MALDI-TOF MS

A new analysis by MALDI-TOF MS was performed using two semi-automated systems: MALDI Biotyper® (MBT; Bruker Corporation, Billerica, USA) and VITEK®MS RUO (bioMérieux, Craponne, France), following the manufacturers' instructions. Both systems record mass spectra within a range of 2,000–20,000 Da, with laser frequencies of 60 Hz and 50 Hz, respectively. For the MALDI Biotyper® system, identification scores were interpreted as follows: 0.00–1.69, no reliable identification; 1.70–1.99, low-confidence species identification; and 2.00–3.00, high-confidence species identification. For the VITEK® MS RUO

system, identification confidence was expressed as percentages, with $\geq 75.0\%$ values considered acceptable for species-level identification.

For sample preparation, isolate B718/18 was streaked onto Tryptic Soy Agar (TSA) plates (Biocen, São Paulo, Brazil) and incubated at 30°C for 24–48 h. Following bacterial growth, a small portion of each colony was applied in duplicate onto the target plate specific to each instrument, followed by the addition of 1 μL of 70% formic acid for on-plate protein pre-extraction. After air drying, 1 μL of α -cyano-4-hydroxycinnamic acid (CHCA) matrix solution (VITEK MS-CHCA, bioMérieux, Craponne, France; Bruker Matrix HCCA, Bruker Corporation, Billerica, USA) was added to each spot. After matrix crystallization, the target plates were introduced into their respective instruments for spectral acquisition and analysis.

Full-length *16S rRNA* gene sequencing

The full-length *16S rRNA* gene sequencing was performed using the MicroSEQ™ Full Gene *16S rDNA* kit (Thermo Fisher Scientific, Waltham, USA), following the manufacturer's protocol. The sequence was obtained using a 3500 Series Genetic Analyzer (Applied Biosystems, Waltham, USA) and subsequently assembled and edited using SeqMan (v. 7.0.0), integrated into DNASTAR Lasergene package (DNASTAR Inc., USA).

The consensus sequence was deposited in the National Center for Biotechnology Information (NCBI) database. Identification was performed using the EzBioCloud database (updated on April 21, 2025; last accessed in June 2026) (Chalita et al. 2024).

Identification based on full-length *16S rRNA* gene sequencing was considered reliable at the genus level when sequence identity exceeded 95%. Species-level identification was considered reliable when sequence identity was $\geq 98.7\%$, as previously proposed by Chalita et al. 2024. Sequence identity values below 98.7% were interpreted as indicative of a potential novel species.

Phylogenetic analysis of the *16S rRNA* gene based on Sanger sequencing

For the phylogenetic analysis, *16S rRNA* gene sequences from the type strains of all species within the genus, both validly and non-validly published under the International Code of Nomenclature of Prokaryotes (INCP), were obtained from the LPSN (Freese et al. 2025), last accessed June 12, 2026. Sequence alignment was performed using the program MAFFT online (v. 7) (Kato and Standley 2013) and subsequently edited using the software MEGA (v. 11.0.13) (Tamura et al. 2021).

Phylogenetic trees were generated using PhyloSuite (v. 1.2.3) (Zhang et al. 2020), within which ModelFinder (v. 2.0.0) (Kalyaanamoorthy et al. 2017) was employed to determine the best-fit evolutionary model. IQ-TREE was used for maximum likelihood inference (Wong et al. 2025). Bayesian inference was also conducted using MrBayes (v. 3.2.7a) (Ronquist et al. 2012) through the same platform.

The phylogenetic trees were visualized and edited using FigTree (v. 1.4.4) (<https://github.com/rambaut/figtree/releases#release-v1.4.4>). The sequence of type strain of *Dermabacter hominis* strain DMS 7083^T (GenBank accession number: X91034) was used as an outgroup to root the trees.

Morphological and biochemical characterization

The morphological characteristics of colonies formed by the bacterial isolate B718/18 were investigated by culturing in TSA plates and in Tryptic Soy Broth (TSB) (Biocen, São Paulo, Brazil), at temperatures of 5.0–6.0 °C, 30 °C, 35 °C, and 42°C for 24 and 48 h. The cell morphology, evaluated by Gram staining and optical microscopy. The biochemical characterization was performed using the API 20 NE and API 50 CH systems (bioMérieux, Craponne, France) to evaluate enzymatic activities, carbohydrate assimilation profiles, nitrate reduction, indole production, gelatin hydrolysis, and urease activity. In addition, complementary enzymatic analyses were performed using the VITEK®2 system (bioMérieux, Craponne, France) with GP (Gram-positive cocci and non-spore-forming bacilli) and CBC (*Corynebacterium* and related bacteria) cards. All methodologies were performed according to the manufacturers' instructions, and the results for the same biochemical test were compared.

For the catalase test, a drop of a 3% hydrogen peroxide solution (Supelco, Bellefonte, United States of America) was added to the bacterial culture on a slide, and the formation of air bubbles was considered a positive reaction. For oxidase detection, a portion of the colony was placed on filter paper, and a drop of N,N-dimethyl-p-phenylenediamine dihydrochloride solution (Inlab, São Paulo, Brazil) was added. The motility test was performed in Sulfide Indole Motility (SIM) semi-solid culture medium (Kasvi, Brasil, Pinhais, Brazil) incubated at 30°C for 24–48 h.

Biofilm formation

Biofilm formation was examined using scanning electron microscopy (SEM) in four different surfaces: vinyl coating, stainless steel, acrylic, and polystyrene, according to the protocol described by Vasconcellos et al. 2026. The strains were cultured in 15 mL of brain heart infusion (BHI) broth (Merck, Darmstadt, Germany) and incubated at 30 °C for 24 h with shaking. A sterile 6-well polystyrene plate (Corning®, New York, USA) was prepared with 4.0 mL of bacterial suspension in each well, along with each different carrier surface. Negative control included carriers with BHI. The plate was incubated at 30 °C for 48 h. Following incubation, the wells were washed twice with 2.0 mL phosphate-buffered saline (PBS) (Gibco-BRL, Gaithersburg, USA). The biofilm was then fixed for 1 h using 2.5% glutaraldehyde in 0.1 M cacodylate buffer. After fixation, the biofilm was washed three times with PBS for 5 min, post-fixed for 15 min in 1% osmium tetroxide (OsO₄) (Electron Microscopy Sciences, Hatfield, USA) and washed again, 3 times in PBS/5 min each. The samples were dehydrated through a graded ethanol series (7.5, 15, 30, 50, 70%, 90, and 100%) for 15 min at each concentration, critical point dried with CO₂ using a Critical Point Dryer machine (Leica-Microsystems, Wetzlar, Germany). Then, sputter-coated with a 15-nm gold layer and examined with a Jeol JSM 6390 scanning electron microscope (Tokyo, Japan).

Genome sequencing, assembling and annotation

Genomic DNA from bacterial isolate was obtained using the DNeasy Blood & Tissue kit (Qiagen, Hilden, Germany), following the protocols recommended by the manufacturer. High-throughput sequencing was conducted at the Fiocruz Platform Network (Oswaldo Cruz Institute, Brazil) on an Illumina NextSeq 2000 system (Illumina®, San Diego, USA). Sequencing libraries were constructed using the Nextera DNA Flex kit (Illumina®, San Diego, USA) in accordance with the manufacturer's guidelines. Paired-end sequencing was performed using P2 XLEAP-SBSTM reagents (Illumina®, San Diego, USA), generating reads of 151 base pairs in each direction (2×151 bp).

The quality of raw sequencing data was initially evaluated with FastQC (v. 0.12.1) (Andrews 2010) to detect potential issues such as adapter contamination, overrepresented sequences, and low-quality bases. Common sequencing artefacts like poly-G and poly-X tails were removed using Fastp (v. 0.23.2) (Chen 2025). Additional quality filtering and trimming were applied with Trimmomatic (v. 0.40) (Bolger; Lohse; Usadel, 2014), including the removal of adapter sequences, low-quality bases (Phred score < 28) from both read ends, and reads shorter than 50 bp.

High-quality reads were assembled *de novo* using SPAdes (v. 4.2.0) (Prjibelski et al. 2020), and the resulting assembly was further refined through polishing with Pilon (v. 1.24) (Walker et al. 2014). Genome quality parameters, including completeness and contamination, were estimated using CheckM Taxonomy_wf (v. 1.2.5) (Parks et al. 2015) within the Galaxy Platform (v. 25.0). The presence of chimerism and contamination was further evaluated with GUNC (v. 1.0.5) (Orakov et al. 2021), considering Clade Separation Score values above 0.45 as indicative of potential chimeric or contaminated assemblies. Possible contamination in the *16S rRNA* gene sequence was assessed using the Contest16S algorithm (Lee et al. 2017). Genome completeness based on single-copy orthologs was examined with BUSCO (v. 6.0.0) (Simão et al. 2015), while assembly statistics were generated using QUAST (v. 5.3.0) (Mikheenko et al. 2018).

The detection of genes associated with antimicrobial resistance (AMR), virulence, and plasmid replicons was performed using ABRicate (v. 0.8.13) (<https://github.com/tseemann/abricate>) by screening the assembled genome against multiple reference databases. The analyses included the NCBI AMRFinderPlus (Feldgarden et al. 2019), Comprehensive Antibiotic Resistance Database (CARD) (Alcock et al. 2023), ResFinder (Bortolaia et al. 2020), ARG-ANNOT (Gupta et al. 2014), MEGARes (Bonin et al. 2023), Virulence Factors Database (VFDB) (Liu et al. 2022), and PlasmidFinder (Carattoli et al. 2014), all updated as of June 12, 2026. Searches were conducted using a minimum nucleotide identity threshold of 60% and a minimum coverage of 80%.

Molecular typing was performed by submitting the genome to the ribosomal multilocus sequence typing (rMLST) database (<https://pubmlst.org/species-id>), to determine the allelic profiles and ribosomal sequence type (rST) corresponding (Jolley et al. 2012). Additionally, CRISPR-Cas systems were identified using CRISPRCasFinder (v. 1.1.2), considering only CRISPR arrays supported by evidence levels 3 or 4 (Couvin et al. 2018).

Genomic mining approaches were employed to explore the biosynthetic potential of the analyzed genome. Biosynthetic gene clusters (BGCs) were predicted using antiSMASH (v. 7.1.0) (Blin et al. 2023) and DeepBGC (v. 0.1.31) (Hannigan et al. 2019). Cluster similarity scores equal to or above 0.7 were interpreted as indicative of more reliable relationships among BGCs, in accordance with thresholds commonly adopted in analyses integrating antiSMASH outputs with BiG-SCAPE and MIBiG datasets (Gren et al. 2021)..

The genome annotation was carried out using the NCBI Prokaryotic Genome Annotation Pipeline (Tatusova et al. 2016) and the Rapid Annotation using Subsystem Technology (RAST) Server (Aziz et al. 2008).

Genomic taxonomy, phylogeny of the core proteome and pangenome analysis

The taxonomic assignment of strain B718/18 was conducted using a polyphasic genomic approach that included the identification by Genome Taxonomy Database Toolkit (GTDB-Tk) (v.2.6.1) (Chaumeil et al. 2020), ribosomal multilocus sequence typing (rMLST) (Jolley et al. 2012), and the Type Strain Genome Server (TYGS) (Freese et al. 2025). ANI values were calculated using the JSpecies Web Server (v. 5.0.3) employing the BLAST-based (ANiB) algorithm (Richter et al. 2016). Digital DNA–DNA hybridization (dDDH) values were computed with the Genome-to-Genome Distance Calculator (GGDC) (v. 3.0) using the recommended formula $2 \times (\text{identities} / \text{HSP length})$ (Meier-Kolthoff et al. 2022).

Proteomic-based phylogenetic analyses were carried out using predicted protein sequences retrieved from the NCBI database by NCBI Datasets tool (v. 18.9.0) (O’Leary et al. 2024). Orthologous protein clusters were identified using OrthoFinder (v. 3.1) (Emms and Kelly 2019), followed by multiple sequence alignment with MAFFT (v. 7.525) (Kato and Standley 2013). A maximum-likelihood phylogenetic tree based on orthologous proteins was reconstructed using FastTree (v. 2.2.0) (Price et al. 2010) with the “-M msa” parameter to optimize alignment-based evolutionary modeling. The phylogenetic tree was subsequently visualized and manually edited using FigTree (v.1.4.4). The accession numbers of the reference genomes of all *Brachybacterium* species used in this study are provided in Table 1.

Table 1

List of genomes of type strains of *Brachybacterium* species used in this study and comparative genomic metrics (ANIb e dDDH) and 16S rRNA sequence similarities between the B718/18 isolate (GC content of 72.4%) and reference genomes of *Brachybacterium* species.

<i>Brachybacterium</i> spp.	Accession no.	ANIb (%)	dDDH (%)	16S rRNA identity (%)	GC content%
<i>B. alimentarium</i> JCM 11607	GCA_039526145.1	78.6	22.4	97.97	70.5
<i>B. aquaticum</i> DSM28796	GCA_014204755.1	80.4	23.7	98.16	72.0
<i>B. atlanticum</i> PhyBa_CO2_2	GCA_025669985.1	81.0	24.1	97.36	72.0
<i>B. avium</i> VR2415	GCA_002216795.1	78.5	21.8	97.27	70.5
<i>B. conglomeratum</i> CCM2589	GCA_042647605.1	81.0	24.4	97.12	72.0
<i>B. endophyticum</i> M1HQ-2	GCA_003130585.1	75.4	20.6	97.12	71.0
" <i>B. epidermidis</i> " Marseille-Q2903	GCA_015209585.1	77.1	21.4	97.76	70.5
" <i>B. equifaecis</i> " JHP9	GCA_023509525.1	75.5	20.7	96.78	71.0
<i>B. faecium</i> DSM 4810	GCA_000023405.1	79.5	22.9	97.41	72.0
<i>B. fresconis</i> DSM 14564	GCA_017876515.1	78.9	22.6	97.97	70.5
<i>B. ginsengisoli</i> DCY80	GCA_002407065.1	79.5	22.5	97.41	71.5
<i>B. hainanense</i> CICC 10874	GCA_042434575.1	75.8	20.8	97.20	72.0
<i>B. halotolerans</i> MASK1Z-5	GCA_016623465.1	78.0	21.4	96.78	72.0
<i>B. horti</i> JCM 31742 ^T	GCA_054777615.1	78.8	21.9	96.80	72.7
<i>B. huguangmaarensis</i> JCM 30544	GCA_054781315.1	74.5	20.4	96.58	73.0
<i>B. kimchii</i> CBA3104	GCA_023373525.1	76.0	21.5	96.85	71.5
" <i>B. massiliense</i> " mt5	GCA_900184245.1	79.3	22.6	98.18	70.5
<i>B. muris</i> CCM 7047	GCA_042433385.1	77.0	21.5	97.79	70.0
<i>B. nesterenkovii</i> CIP 104813	GCA_900163655.1	75.3	21.0	97.41	72.5
" <i>B. netajii</i> " DNPG3	GCA_054114245.1	78.4	22.5	97.76	72.5
<i>B. paraconglomeratum</i> KCTC9916	GCA_003994255.1	80.9	24.3	97.48	72.0

<i>Brachybacterium</i> spp.	Accession no.	ANIb (%)	dDDH (%)	16S rRNA identity (%)	GC content%
<i>B. phenoliresistens</i> JCM 15157	GCA_039530785.1	76.5	21.0	96.67	73.0
<i>B. rhamnosum</i> JCM 11650	GCA_039526185.1	78.8	21.9	97.55	73.0
<i>B. sacelli</i> DSM 14566	GCA_017876545.1	78.7	22.2	98.68	70.5
<i>B. saurashtrense</i> DSM 23186	GCA_003355475.1	79.8	22.9	97.83	73.0
<i>B. sillae</i> EF45031	GCA_025028335.1	74.4	20.1	96.84	71.0
<i>B. squillarum</i> M-6-3	GCA_000225825.2	78.2	21.9	97.97	73.0
<i>B. subflavum</i> CFH 10395	GCA_009299795.1	76.1	21.3	96.79	72.0
<i>B. timonense</i> Marseille-P4339	GCA_900258505.1	74.1	20.4	96.91	67.5
<i>B. tyrofermentans</i> JCM 11610	GCA_039526165.1	78.9	22.4	97.48	70.0
<i>B. vulturis</i> VM2412	GCA_002407185.1	78.6	22.0	97.62	71.0
<i>B. zhongshanense</i> JCM 15471	GCA_001314085.1	79.2	25.7	97.51	70.5
“ ”; non validly species published under the INCP, according to LPSN (Göker et al., 2025); ANIb, average nucleotide identity based on Blast algorithm; dDDH, digital hybridization DNA-DNA; ^T , type strains.					

Pangenome analysis was performed using PPanGGOLiN (v. 2.2.6) (Gautreau et al. 2020) based on genome annotations archives (gff) generated with Prokka (v. 1.14.6) (Seemann 2014), using the default parameters for identity and coverage (80%). Gene families were inferred through clustering of homologous protein sequences and subsequently structured into a graph-based representation of the pangenome. The classification of gene families into persistent (core), shell, and cloud partitions was achieved using a statistical framework that integrates gene frequency across genomes with their genomic neighborhood context. Downstream analyses, including the construction of pangenome accumulation curves and the evaluation of gene repertoire composition, were conducted using the resulting gene presence/absence matrix.

Results and discussion

Reanalysis of bacterial isolate B718/18 by MALDI-TOF MS using two instruments from different manufacturers, MALDI Biotyper® and VITEK® MS RUO systems, reproduced the results previously obtained in 2018, with no identification achieved for the isolate. MALDI-TOF MS enables reliable

discrimination among bacterial species, including phylogenetically closely related taxa, through comparison of the protein mass spectrum of an isolate with reference spectra contained in curated databases (Ashfaq et al. 2022; Calderaro and Chezzi 2024). Although MALDI Biotyper and VITEK MS employ different sample preparation protocols, spectral databases, and identification algorithms, both platforms considerably reduce the time and cost associated with microbial identification, representing an important advantage in applications such as pharmaceutical quality control and batch release processes (Ashfaq et al. 2022; Moreira et al. 2023; Song et al. 2024; Ferreira et al. 2025; Li et al. 2026).

Initially developed primarily for clinically relevant microorganisms (Calderaro and Chezzi 2024), MALDI-TOF MS databases have progressively expanded to include environmental, industrial, and niche-associated isolates, frequently validated by *16S rRNA* gene sequencing, thereby broadening the applicability of the technique (Costa et al. 2022b; Song et al. 2024). Nevertheless, the limited representation of environmental bacteria in commercial spectral databases remains a major limitation for accurate identification of environmental isolates by MALDI-TOF MS (Ashfaq et al. 2022; Costa et al. 2022a; Nunes Ramos et al. 2025). Despite this limitation, the possibility of incorporating spectra from isolates previously characterized by robust taxonomic approaches has strengthened the utility of this methodology as a rapid and cost-effective tool for the characterization and comparison of environmental microorganisms from specific ecosystems (Costa et al. 2022b; Song et al. 2024).

Following the non-identification results obtained with the aforementioned methodology, complete sequencing of the *16S rRNA* gene was performed to determine the taxonomic identity of bacterial isolate B718/18. Sequencing generated a 1,428-base pair (bp) fragment with 99.0% sequence integrity. Comparative analysis using the EzBioCloud database revealed 98.68% sequence identity with the type strain of *Brachybacterium sacelli* LMG 20345^T, a value below the 98.7% threshold proposed for species delineation (Chalita et al. 2024). Therefore, bacterial isolate B718/18 was considered a potential novel species.

Phylogenetic tree inference based on *16S rRNA* gene sequences from different species of the genus *Brachybacterium* revealed that bacterial isolate B718/18 formed a distinct monophyletic clade supported by high bootstrap values (97%) and posterior probability values (1.0), indicating strong clade robustness (Fig. 1).

Regarding morphological characteristics, bacterial isolate B718/18 formed yellow, elevated colonies with smooth margins, measuring approximately 1.8–2.5 mm in diameter, after 48 h of incubation under aerobic conditions at 30 and 35°C, on TSA plates and in TSB broth, which were considered the optimal growth conditions. No growth was observed at 42°C or at temperatures between 5.0 and 6.0°C. Microscopic examination revealed Gram-positive, coccobacilli.

Biochemical characterization revealed that bacterial isolate B718/18 was oxidase-negative, catalase-positive, and non-mobile. According to de Castro et al. (2023), these enzymatic traits are important for the taxonomic classification of members of the genus *Brachybacterium*. Biochemical assays associated with carbohydrate metabolism, including glucose, maltose, mannitol, mannose, raffinose, and trehalose,

were considered variable due to inconsistent results obtained using different methodological approaches (Table 2). Among these, glucose, maltose, mannitol, and trehalose yielded positive results in two or three methodologies and negative results in one, whereas mannose was negative in two methodologies and positive in one. In contrast, raffinose utilization was positive in one methodology and negative in another.

Table 2
Biochemical characteristics of *B. manguinensis* B718/18 and related species.

Features	1	2	3	4	5	6
Isolation source	Air of pharmaceutical facility	Poultry deep litter	Oil brine	Sea-water	Sea-water	Mural paintings
Colour	SY	W-PB	PB	Y	SY	C
DNA G + C content (%)	72.4	68.6	70.6	72.1	72	70.3
Enzymatic activity						
Catalase	+	+	+	+	+	+
Hydrolysis of gelatin	-	-	+	+	-	-
Indole production	-	+	+	-	-	-
Nitrate reduction	+	+	-	-	-	+
Oxidase	-	-	-	-	-	-
Urease	-	+	+	-	+	-
α -Galactosidase	+	-	-	-	-	ND
β -Galactosidase	+	+	-	-	+	+
β -Glucuronidase	+	+	+	-	-	ND
α -Glucosidase	+	+	-	-	-	ND
Carbohydrate metabolism						
Celobiose	+	-	+	+/-	-	ND
Fructose	+	-	+	+	+	+
Galactose	+	-	+	+	+	-
Glucose	V+	-	+	+	+	+
Maltose	V+	+	+	+/-	-	+
Mannitol	V+	-	-	-	-	+
Strains: 1, <i>B. manguinensis</i> B718/18^T; 2, <i>B. paraconglomeratum</i> KCTC9916^T; 3, <i>B. conglomeratum</i> VKM Ac-1975^T; 4, <i>B. atlanticum</i> PhyBa_CO2_2^T; 5, <i>B. aquaticum</i> DSM 28796^T; 6, <i>B. sacelli</i> DSM 14566^T. Y, yellow; PY, pale yellow; SY, straw yellow; PB, pale brown; C, cream; W, white; +, Positive; +/-, weakly positive; -, negative; ND, not determined; V, variable; V+, variable with two or more positive results; V-, variable with two or more negative results.						

Features	1	2	3	4	5	6
Mannose	V-	+	+	+	-	+
Raffinose	V	-	+	+/-	-	ND
Rhamnose	+/-	+	-	-	-	(+)
Salicin	+	+	-	+/-	-	ND
Sucrose	+	-	+	+/-	+	+
Trehalose	V+	-	+	-	-	ND
Strains: 1, <i>B. manguinensis</i> B718/18 ^T ; 2, <i>B. paraconglomeratum</i> KCTC9916 ^T ; 3, <i>B. conglomeratum</i> VKM Ac-1975 ^T ; 4, <i>B. atlanticum</i> PhyBa_CO2_2 ^T ; 5, <i>B. aquaticum</i> DSM 28796 ^T ; 6, <i>B. sacelli</i> DSM 14566 ^T . Y, yellow; PY, pale yellow; SY, straw yellow; PB, pale brown; C, cream; W, white; +, Positive; +/-, weakly positive; -, negative; ND, not determined; V, variable; V+, variable with two or more positive results; V-, variable with two or more negative results.						

The bacterial isolate B718/18 was unable to form biofilms on any of the four surfaces evaluated (Fig. 2). This finding is consistent with the results reported by Hussain et al. (2025), who also observed the absence of biofilm formation in a strain of *Brachybacterium paraconglomeratum*. However, biofilm formation has been reported in other members of the genus. For instance, lungin et al. (2024) observed the formation of a dispersed biofilm with low cellulose content and few amyloid proteins by a *Brachybacterium* sp. strain associated with Antarctic vascular plants. Likewise, Ashraf et al. (2024) reported biofilm formation on polystyrene by *Brachybacterium* sp. isolates recovered from biofilms associated with melanin-secreting corals. These findings suggest that biofilm-forming ability may vary among species and strains of *Brachybacterium*. Therefore, although strain B718/18 did not form biofilms under the conditions tested in this study, this trait cannot be excluded for other strains of the same specie.

Whole-genome sequencing generated approximately 23.3 million paired-end reads, yielding a total of 3.4 Gbp of raw sequencing data. Genome assembly resulted in a sequence of 3,308,393 bp across 22 contigs, with a GC content of 72.4% (within the 67.5–73.0% range of *Brachybacterium* reference genomes deposited in NCBI (Table 1). An average sequencing coverage of approximately 583,58x was obtained, with assembly metrics showing an N50 of 294,729 bp and an L50 of 4 contigs. Annotation performed by the NCBI pipeline identified 2,294 coding sequences (CDSs), along with 57 RNA genes, including 50 tRNAs and three complete rRNA operons. A total of 827 genes were assigned to functional subsystems in the genome annotation performed on the RAST server (Fig. 3). Most of the genes were associated with the categories of amino acids and derivatives (21.4%), carbohydrates (17.3%), and protein metabolism (14.9%). Approximately 22 genes were classified into the categories of virulence, disease, and defense; 15 of these genes are associated with resistance to antibiotics and toxic compounds, and 12 are associated with intracellular invasion and resistance.

Genome completeness and contamination were estimated at 90.19% and 8.38%, respectively, based on 876 conserved markers from the Dermabacteriaceae family. These values fall below the minimum criteria recommended for the taxonomic classification of prokaryotes (Riesco and Trujilo 2024). No contamination or chimerism were detected by GUNC analysis, which demonstrated a hits average identity of 0.77% and the reference representation scores of 0.69%. These data suggest that the strain B718/18 may belong to an underrepresented lineage within the genus *Brachybacterium*, what corroborates its potential taxonomic novelty. In addition, it was not possible to determine contamination by the 16S ribosomal gene, since only a 1,519-bp fragment was recovered from the genome. According to Teo et al. (2026), 16S rRNA gene sequences obtained by conventional Sanger sequencing should have matched with those recovered from the whole genome. In this context, alignment between the sequence obtained by Sanger sequencing and the extracted from the assembled genome revealed 99,93% identity, demonstrating high consistency between the different sequencing approaches and corroborating the reliability of the obtained genome assembly.

Based on the *Brachybacterium* genus dataset implemented in BUSCO, a total of 1,212 orthologous genes were assessed. Of these, 1,183 were identified as complete and single-copy, two as complete and duplicated, seven as fragmented, and twenty were classified as missing, resulting in an estimated genome completeness of 97.80%, indicating that the assembly is of sufficient quality for comparative genomic analyses. The number of missing orthologs is likely attributable to assembly fragmentation or natural genomic variation among *Brachybacterium* strains. Furthermore, the very low number of duplicate BUSCOs suggests the absence of serious assembly errors or contamination. The missing orthologs were identified using OrthoDB v12.2 (Tegenfeldt et al. 2025) based on their corresponding gene IDs. Most of the missing BUSCOs are associated with carbohydrate metabolism, secondary catabolism, transporters, and cell regulation and are listed in **Table S1**.

Genomic analysis revealed the presence of three antibiotic resistance-associated genes (Table 3), including *mphH*, which encodes the macrolide 2'-phosphotransferase H (MphH). This enzyme confers resistance to macrolide antibiotics, such as azithromycin, by enzymatic inactivation through phosphorylation. The *mphH* gene, originally described in *Brachybacterium faecium*, encodes a protein that specifically targets the 2'-hydroxyl group of the desosamine sugar moiety of macrolide molecules (Madej et al. 2014). In addition, genes related to resistance to the aminocoumarin antibiotics were identified. Aminocoumarins are natural products produced by soil-dwelling bacteria, such as *Streptomyces* spp., and are known as potent inhibitors of DNA gyrase, an essential bacterial enzyme and a well-established target of antimicrobial agents (Lawson and Stevenson 2012). No AMR genes were detected using the ResFinder and ARG-ANNOT databases. No plasmid replicon was detected.

Table 3

Antimicrobial resistance (AMR) and virulence genes detected by ABRicate in multiple databases.

Genes	Category	Database	Contig	Notes
<i>mphH</i>	AMR	CARD, NCBI, MEGARes	12	macrolide 2'-phosphotransferase
<i>parY</i>		MEGARes	1	aminocoumarin-resistant DNA topoisomerase
<i>novA</i>		CARD, MEGARes	5	aminocoumarin efflux pump
<i>bpID</i>	virulence	VFDB	16	UDP-N-acetylglucosamine 2-epimerase
<i>sodA</i>			1 and 4	superoxide dismutase
<i>narH</i>			1	nitrate reductase subunit beta
<i>relA</i>			1	probable GTP pyrophosphokinase
<i>phoP</i>			5	possible two component system response transcriptional positive regulator

Several genes associated with functions commonly linked to bacterial fitness, stress adaptation, and host interaction were identified (Table 3). One of them is the *bpID* gene, encoding a UDP-N-acetylglucosamine 2-epimerase, which is involved in cell envelope and polysaccharide biosynthesis and has been implicated in virulence-related processes in some bacteria (Campbell et al. 2000). The *sodA* gene, detected in two contigs, encodes the enzyme superoxide dismutase, which is widely distributed in various organisms and plays a key role in protecting against oxidative stress, enabling bacterial survival in the face of reactive oxygen species produced by the host's immune responses (Zhang et al. 2025). Respiratory nitrate reductase is an enzyme complex consisting of three subunits encoded by the *narG*, *narH*, and *narI* genes (Viana et al. 2017). The *narH* gene, responsible for encoding the β -subunit of the enzyme, was identified in the analyzed genome and has also been described in the genome of the type strain of *Brachybacterium massiliense*. However, although the presence of this gene suggests potential involvement in nitrate-dependent respiratory processes, its specific function in the metabolism of species of the genus *Brachybacterium* is not yet clearly understood, and further functional studies are needed for its complete characterization.

Molecular typing by rMLST revealed that the strain belongs to the profile rST-316643. Thirty-five new alleles were deposited in this database (**Table S2**). Strain B718/18 was not identified in the rMLST database, showing 32% support for the genus *Brachybacterium* and 52 exact-match alleles, forming a singleton at all clustering thresholds, highlighting its genetic distinctiveness and corroborating its classification as a potentially new species within the *Brachybacterium* genus.

The CRISPR-Cas system functions as an adaptive immune mechanism in bacteria and archaea, providing protection against invading genetic elements, such as bacteriophages and plasmids (Makarova et al. 2018). The type I-E CRISPR-Cas system has previously been reported in the genomes of *Brachybacterium equifaecis*, *Brachybacterium nesterenkovi*, *Brachybacterium squillarum*, and *Brachybacterium vulturis* (Farooq et al. 2023). In contrast, no CRISPR arrays with high-confidence evidence levels (3 or 4) were identified in the B718/18 isolate, indicating that the presence of this adaptive immune mechanism is variable within the genus. This characteristic is not uncommon among bacteria, and its absence could facilitate horizontal gene transfer, potentially contributing to genomic plasticity and the acquisition of accessory genes, including those associated with environmental adaptation or AMR (Ramos et al. 2023).

BGCs are responsible for the biosynthesis of secondary metabolites, also known as natural products, including pigments, antibiotics, antifungal compounds, and siderophores. These molecules play important roles in various bacterial survival strategies and are of great significance for human medicine, animal health, and the protection of agricultural crops (Li et al. 2021; Bruna et al. 2024). Analysis using antiSMASH revealed a terpene-type BGC located on contig 4, with 20,911 bp, showing 28% similarity to clusters associated with carotenoid biosynthesis. The low similarity observed suggests that this cluster may represent a divergent variant or potentially a new terpene-type BGC, which are frequently associated with the production of pigments such as carotenoids, which play important roles in protection against oxidative stress and cellular adaptation (Yamada et al. 2015). Of the 25 BGCs identified by the DeepBGC program (Fig. 4), 10 had scores higher than 0.8 (Fig. 4a). Regarding classification (Fig. 4b), the majority (n = 19) were classified as undefined, while six were distributed among the saccharide (n = 3), polyketide (n = 2), and RiPP (ribosomally synthesized and post-translationally modified peptides) (n = 1) classes. Notably, a saccharide-class BGC with 26,165 bp achieved the maximum score (1.0), although its activity was not determined (Fig. 4c). In total, 18 BGCs were associated with antibacterial activity (Fig. 4d), while seven remained without a defined function based on the database used. Among the BGCs with antibacterial activity, the majority (n = 14) had scores between 0.5 and 0.6, with only two classified by class - one as a polyketide and the other as a RiPP.

The taxonomic classification conducted in GTDB-Tk indicates that isolate B718/18 was classified with high confidence into the genus *Brachybacterium*, based on approximately 95.70% of the marker genes found in the genome; however, it does not correspond to any described and validated species and was therefore assigned to a placeholder species in the GTDB database, designated *Brachybacterium* sp. J153 (NCBI accession number: GCF_036311905.1), with which it shares 96.34% ANI, corroborating its inclusion in this genomic cluster.

Analysis performed by the TYGS server indicated that isolate B718/18 represents a potential new species (Fig. 5). According to the GBDP (Genome BLAST Distance Phylogeny) distance-based classification, the closest species are *Brachybacterium sacelli* and *Brachybacterium fresconis*, both of which have dDDH values of 22.2% and 22.6%, respectively, relative to the strain in this study. The NCBI Taxonomy database (Schoch et al. 2020) has assigned TaxID 3684135 to isolate B718/18, classifying it

as “unclassified” within the genus *Brachybacterium*, which indicates that, based on currently available data, its taxonomic status at the species level remains unresolved. The core proteome-based phylogeny (Fig. 6), inferred from 357 single-copy proteins, placed strain B718/18 in a distinct clade, supported by high bootstrap support (100%), confirming the robustness of the phylogenetic grouping.

The ANI and dDDH values were below the thresholds established for bacterial species delineation (Table 1), indicating that it does not belong to any of the currently described *Brachybacterium* species, which is consistent with the classification provided by TYGS. The ANI metric estimates the average nucleotide identity between whole genomes and is widely used as a standard for genome-based taxonomic inference. In general, ANI values above 95–96% indicate that two strains belong to the same species. Complementarily, the dDDH computationally replicates the classical DNA-DNA hybridization method, historically considered the gold standard for species delimitation, adopting a threshold of 70% or higher as indicative of belonging to the same species (Thompson et al. 2013; Riesco and Trujillo 2024). These two approaches show a high correlation and are frequently used in conjunction in genomic taxonomy studies, since ANI provides a robust measure of overall similarity between nucleotide sequences, while dDDH incorporates additional aspects related to genomic conservation and the alignment of homologous regions (Riesco and Trujillo 2024). In the present study, the consistently low ANI and dDDH values observed for isolate B718/18 relative to the reference strains analyzed reinforce its taxonomic distinction, supporting its proposal as a potential new species within the genus *Brachybacterium*. The results of the genomic taxonomy analyses conducted consistently indicate that the delineation and differentiation of species within the genus *Brachybacterium* require a polyphasic approach that integrates genotypic, phenotypic, physiological, and chemotaxonomic data.

Finally, the pangenome analysis, which included representatives of both valid and invalid species of the genus *Brachybacterium* (Table 1), revealed an open pangenome (Fig. 7a). It was also observed that the number of genes classified as persistent decreases progressively as new genomes are incorporated into the analysis. Pangenomic studies for *Brachybacterium* are still limited; however, a previous investigation based on seven genomes also reported an open pangenome, with the core genome predominantly associated with essential functions, such as amino acid metabolism, growth, and cell death, among others (Farooq et al. 2023). In the present study, approximately 107,565 genes were grouped into 53,358 gene families, of which only 584 were classified as persistent. The frequency distribution of these families (Fig. 7a) reveals a high proportion of rare genes (cloud), a pattern also corroborated by the presence-absence matrix (Fig. 7b), in addition to a relatively small set of persistent genes. In addition, 1,276 regions of genomic plasticity were identified, distributed across 129 insertion sites. Modularity analysis revealed 1,082 gene modules containing 6,599 gene families, composed predominantly of genes from the cloud fraction (97.98%), while only 2.02% belonged to the shell fraction; no persistent genes associated with modules were observed. This profile suggests high genomic plasticity and the continuous occurrence of gene gain and loss events, possibly driven by processes of environmental adaptation and horizontal gene transfer. Consistent with this, Farooq et al. (2023) demonstrated that the *Brachybacterium* accessory genome exhibits significant representation of functions related to

carbohydrate metabolism, cellular community organization, terpenoid and polyketide metabolism, drug resistance, as well as biodegradation and metabolism of xenobiotics.

Description of *Brachybacterium manguinensis* sp. nov.

Brachybacterium manguinensis (man.gui.nen'sis N.L. masc./fem adj. *manguinensis*), referring to Manguinhos neighborhood, located in the city of Rio de Janeiro, Brazil, the geographical origin of the type strain. The bacterial strain B718/18, isolated from pharmaceutical production environment, formed yellow, raised colonies with smooth margins, measuring approximately 1.8–2.5 mm in diameter on TSA medium. The cells were found to be Gram-positive, non-motile coccobacilli, oxidase-negative, catalase-positive, negative for hydrolysis of gelatin, production of indole, presence of urease and arginine dihydrolase, but positive for nitrate reduction, glucose fermentation and aesculin. Furthermore, they assimilated xylose, galactose, fructose, cellobiose, melibiose, sucrose, gentiobiose, turanose, L-fucose, and maltotriose as carbohydrate sources. The presence of the enzymes, leucine arylamidase, β -galactosidase, L-pyrrolidonyl arylamidase, tyrosine arylamidase, β -glucosidase, β -xylosidase, L-proline arylamidase, α -glucosidase, α -galactosidase, β -D-fucosidase, β -galactopyranosidase, β -glucuronidase, alanine arylamidase, and phenylalanine arylamidase was observed. However, the enzymes, lipase, α -mannosidase, glycine arylamidase, phosphatidylinositol phospholipase C, arginine dihydrolase 1, L-aspartate arylamidase, phosphatase, ornithine decarboxylase, and arginine dihydrolase 2, were absent. The enzyme Ala-Phe-Pro arylamidase was considered variable, since its presence was detected using the CBC card but not with the GP card, both employing the bioMérieux VITEK 2 system methodology. Using the VITEK 2 GP card, no growth was observed in 6.5% NaCl. Furthermore, the bacterial isolate showed resistance to optochin, but sensitivity to bacitracin and novobiocin using this card.

The strain *Brachybacterium manguinensis* sp. nov. B718/18^T (= CBAS 1562^T = P 7488^T) was isolated from viable airborne particle counting in a pharmaceutical facility producing immunobiologicals, in Brazil. The isolate B718/18 was designated as the type strain of the species and deposited in the Collection of Bacteria from Environmental and Health (CBAS) and Collection of Pathogenic Bacteria (CBP) of the Oswaldo Cruz Foundation (Fiocruz), under the numbers CBAS 1562^T and P 7488^T, respectively. Both collections are affiliated with the World Federation for Culture Collections, under the registers WDCM 958 and WDCM 1312, respectively. Accession numbers for the 16S *rRNA* gene and the whole genome sequence in GenBank/EMBL/DDBJ are PV425877 and NZ_JBVDDO000000000, respectively. The version described in this article is version NZ_JBVDDO000000000.1. The raw sequencing reads were deposited in the NCBI Sequence Read Archive (SRA) under accession number PRJNA1420883, BioSample SAMN55279352.

Declarations

Funding: This work was supported by Plataforma Genômica - Sequenciamento de Nova Geração - RPT01J - Rede de Plataformas Tecnológicas FIOCRUZ; Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ) (Grant No. E-26/200.546/2025), Conselho Nacional de

Desenvolvimento Científico e Tecnológico (CNPq) (grant number 407747/2021-4) and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001. Marcelo Luiz Lima Brandão is grateful for the support from CNPq in the form of a Productivity in Research Fellowship (PQ-C) (Process: 304877/2024-7).

Competing Interests: Author Stephen James Forsythe is the maintainer of Foodmicrobe.com Ltd. The remaining 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.

Author contributions: Conceptualization: N.G.S.C., L.V.d.C., M.L.L.B. and V.Z; Data curation: N.G.S.C., L.V.d.C. and M.L.S.d.S. Formal analysis and investigation: N.G.S.C., J.N.R, M.L.S.d.S. B.d.J.S; Methodology: N.G.S.C., M.L.S.d.S., B.d.J.S., R.V.d.S.L.d.M., T.B.V., B.V., V.M., L.V.d.C. and M.L.L.B.; Project administration: N.G.S.C., V.Z. and M.L.L.B.; Resources: V.Z. and M.L.L.B; Software: J.N.R; Supervision: V.Z. and M.L.L.B; Validation: S.J.F., M.L.S.d.S. Visualization: N.G.S.C. and J.N.R; Writing – original draft preparation: N.G.S.C and J.N.R.; Writing – review and editing: M.L.S.d.S., B.d.J.S., B.V., V.M., L.V.d.C., S.J.F., V.Z. and M.L.L.B.; Funding acquisition: V.Z. and M.L.L.B.; All authors have read and approved the final version of the manuscript.

Data availability: The strain has been deposited in the Collection of Bacteria from Environmental and Health (CBAS) and Collection of Pathogenic Bacteria (CBP) of the Fiocruz, under the numbers CBAS 1562^T and P 7488^T, respectively. Accession numbers for the *16S rRNA* gene and the whole genome sequence in GenBank/EMBL/DDBJ are PV425877 and NZ_JBVDDO000000000, respectively. The version described in this article is version NZ_JBVDDO000000000.1. The raw sequencing reads were deposited in the NCBI Sequence Read Archive (SRA) under accession number PRJNA1420883, BioSample SAMN55279352.

Acknowledgments: The authors gratefully acknowledge Bio-Manguinhos/Fiocruz, the Graduate Program in Biodiversity and Health at IOC/Fiocruz, the National Institute of Quality Control in Health/Fiocruz, the funding agencies CNPq and Faperj/RJ, and the Fiocruz Technological Platforms Network for their support.

References

1. Alcock BP, Huynh W, Chalil R, 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. <https://doi.org/10.1093/nar/gkac920>
2. Andrews S (2010) FastQC - A quality control tool for high throughput sequence data. In: <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
3. Ashfaq MY, Da'na DA, Al-Ghouti MA (2022) Application of MALDI-TOF MS for identification of environmental bacteria: A review. *J Environ Manage* 305:. <https://doi.org/10.1016/j.jenvman.2021.114359>

4. Ashraf N, Anas A, Sukumaran V, et al (2024) Biofilm-forming bacteria associated with corals secrete melanin with UV-absorption properties. *World J Microbiol Biotechnol* 40:. <https://doi.org/10.1007/S11274-024-04120-W>
5. Aziz RK, Bartels D, Best A, et al (2008) The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* 9:. <https://doi.org/10.1186/1471-2164-9-75>
6. Blin K, Shaw S, Augustijn HE, et al (2023) antiSMASH 7.0: new and improved predictions for detection, regulation, chemical structures and visualisation. *Nucleic Acids Res* 51:W46–W50. <https://doi.org/10.1093/nar/gkad344>
7. Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120. <https://doi.org/10.1093/BIOINFORMATICS/BTU170>
8. Bonin N, Doster E, Worley H, et al (2023) MEGARes and AMR++, v3.0: an updated comprehensive database of antimicrobial resistance determinants and an improved software pipeline for classification using high-throughput sequencing. *Nucleic Acids Res* 51:D744–D752. <https://doi.org/10.1093/NAR/GKAC1047>
9. Bortolaia V, Kaas RS, Ruppe E, et al (2020) ResFinder 4.0 for predictions of phenotypes from genotypes. *J Antimicrob Chemother* 75(12):3491–3500. <https://doi.org/10.1093/jac/dkaa345>
10. Brummaier T, Hinfiothong P, Soe NL, et al (2020) *Brachybacterium nesterenkovi* isolated from a human blood culture—a first report. *New Microbes New Infect* 36:100699. <https://doi.org/10.1016/J.NMNI.2020.100699>
11. Bruna P, Núñez-Montero K, Contreras MJ, et al (2024) Biosynthetic gene clusters with biotechnological applications in novel Antarctic isolates from Actinomycetota. *Appl Microbiol Biotechnol* 108(1):325. <https://doi.org/10.1007/S00253-024-13154-X>
12. Caldeira NGS, de Souza MLS, de Miranda RV da SL, et al (2024) Characterization by MALDI-TOF MS and 16S rRNA Gene Sequencing of Aerobic Endospore-Forming Bacteria Isolated from Pharmaceutical Facility in Rio de Janeiro, Brazil. *Microorganisms* 12:724. <https://doi.org/10.3390/MICROORGANISMS12040724>
13. Calderaro A, Chezzi C (2024) MALDI-TOF MS: A Reliable Tool in the Real Life of the Clinical Microbiology Laboratory. *Microorganisms* 12:232. <https://doi.org/10.3390/MICROORGANISMS12020322>
14. Campbell RE, Mosimann SC, Tanner ME, Strynadka NCJ (2000) The Structure of UDP-N-Acetylglucosamine 2-Epimerase Reveals Homology to Phosphoglycosyl Transferases†. *Biochemistry* 39:14993–15001. <https://doi.org/10.1021/BI001627X>
15. Carattoli A, Zankari E, García-Fernández A, et al (2014) In Silico Detection and Typing of Plasmids using PlasmidFinder and Plasmid Multilocus Sequence Typing. *Antimicrob Agents Chemother* 58:3895. <https://doi.org/10.1128/AAC.02412-14>
16. Chalita M, Kim YO, Park S, et al (2024) EzBioCloud: a genome-driven database and platform for microbiome identification and discovery. *Int J Syst Evol Microbiol* 74:006421. <https://doi.org/10.1099/IJSEM.0.006421>

17. Chaumeil PA, Mussig AJ, Hugenholtz P, Parks DH (2020) GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* 36:1925–1927. <https://doi.org/10.1093/BIOINFORMATICS/BTZ848>
18. Chen S (2025) fastp 1.0: An ultra-fast all-round tool for FASTQ data quality control and preprocessing. *iMeta* 4:e70078. <https://doi.org/10.1002/IMT2.70078>;WEBSITE:WEBSITE:PERICLES;ISSUE:ISSUE:DOI
19. Collins MD, Brown J, Jones D (1988) *Brachybacterium faecium* gen. nov., sp. nov., a coryneform bacterium from poultry deep litter. *Int J Syst Bacteriol* 38:45–48. <https://doi.org/10.1099/00207713-38-1-45/CITE/REFWORKS>
20. Costa LV da, Miranda RV da SL de, Fonseca EL da, et al (2022a) Assessment of VITEK® 2, MALDI-TOF MS and full gene 16S rRNA sequencing for aerobic endospore-forming bacteria isolated from a pharmaceutical facility. *J Microbiol Methods* 194:106419. <https://doi.org/10.1016/J.MIMET.2022.106419>
21. Costa LV da, Miranda RV da SL de, Reis CMF dos, et al (2022b) MALDI-TOF MS database expansion for identification of *Bacillus* and related genera isolated from a pharmaceutical facility. *J Microbiol Methods* 203:106625. <https://doi.org/10.1016/J.MIMET.2022.106625>
22. Couvin D, Bernheim A, Toffano-Nioche C, et al (2018) CRISPRCasFinder, an update of CRISRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. *Nucleic Acids Res* 46:W246–W251. <https://doi.org/10.1093/NAR/GKY425>
23. de Castro I, Ribeiro S, Oliveira V, et al (2023) *Brachybacterium atlanticum* sp. nov., a novel marine bacterium isolated from the Atlantic Ocean. *Int J Syst Evol Microbiol* 73(8): <https://doi.org/10.1099/IJSEM.0.005959>
24. Emms DM, Kelly S (2019) OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol* 20(1):238-. <https://doi.org/10.1186/S13059-019-1832-Y>
25. Farooq A, Lee M, Han S, et al (2023a) Kinetic, genomic, and physiological analysis reveals diversity in the ecological adaptation and metabolic potential of *Brachybacterium equifaecis* sp. nov. isolated from horse feces. *Microbiol Spectr* 11:. <https://doi.org/10.1128/SPECTRUM.05048-22;WGROU:STRING:PUBLICATION>
26. Farooq A, Lee M, Han S, et al (2023b) Kinetic, genomic, and physiological analysis reveals diversity in the ecological adaptation and metabolic potential of *Brachybacterium equifaecis* sp. nov. isolated from horse feces. *Microbiol Spectr* 11(5):e05048-22. <https://doi.org/10.1128/SPECTRUM.05048-22>
27. Feldgarden M, Brover V, Haft DH, et al (2019) Validating the AMRFinder Tool and Resistance Gene Database by Using Antimicrobial Resistance Genotype-Phenotype Correlations in a Collection of Isolates. *Antimicrob Agents Chemother* 63:e00483-19. <https://doi.org/10.1128/AAC.00483-19>
28. Ferreira LS, Showemimo T, Juliano LB, et al (2025) Use of MALDI-TOF to identify cryopreserved mastitis pathogens collected from 2003 to 2011 that were originally identified using conventional microbiological methods. *JDS communications* 6:422–426. <https://doi.org/10.3168/JDSC.2024-0711>

29. Freese HM, Meier-Kolthoff JP, Sardà Carbasse J, et al (2025) TYGS and LPSN in 2025: a Global Core Biodata Resource for genome-based classification and nomenclature of prokaryotes within DSMZ Digital Diversity. *Nucleic Acids Res* 54(D1):D884–D991. <https://doi.org/10.1093/NAR/GKAF1110>
30. Gautreau G, Bazin A, Gachet M, et al (2020) PPanGGOLiN: Depicting microbial diversity via a partitioned pangenome graph. *PLoS Comput Biol* 16(3):e1007732. <https://doi.org/10.1371/journal.pcbi.1007732>
31. Gren T, Whitford CM, Mohite OS, et al (2021) Characterization and engineering of *Streptomyces griseofuscus* DSM 40191 as a potential host for heterologous expression of biosynthetic gene clusters. *Scientific Reports* 2021 11:1 11:18301-. <https://doi.org/10.1038/s41598-021-97571-2>
32. Gupta SK, Padmanabhan BR, Diene SM, et al (2014) ARG-ANNOT, a new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes. *Antimicrob Agents Chemother* 58:212–220. <https://doi.org/10.1128/AAC.01310-13>
33. Gvozdyak OR, Nogina TM, Schumann P (1992) Taxonomic study of the genus *Brachybacterium*: *Brachybacterium nesterenkovi* sp. nov. *Int J Syst Bacteriol* 42:74–78. <https://doi.org/10.1099/00207713-42-1-74>
34. Hannigan GD, Prihoda D, Palicka A, et al (2019) A deep learning genome-mining strategy for biosynthetic gene cluster prediction. *Nucleic Acids Res* 47:e110–e110. <https://doi.org/10.1093/NAR/GKZ654>
35. Hlaing PPT, Junqueira ACM, Uchida A, et al (2019) Complete Genome Sequence of *Brachybacterium* sp. Strain SGAir0954, Isolated from Singapore Air. *Microbiology Resour Announc* 8(32):e00619-19. <https://doi.org/10.1128/MRA.00619-19>
36. Hussain MA, Zafar M, Khan YS, et al (2025) Molecular identification, antibiotic susceptibility, and biofilm formation of airborne bacteria. *AMB Express* 15(1):74. <https://doi.org/10.1186/S13568-025-01861-1>
37. Iungin O, Prekrasna-Kviatkovska Y, Kalinichenko O, et al (2024) Endophytic Bacterial Biofilm-Formers Associated with Antarctic Vascular Plants. *Microorganisms* 12(10):1938. <https://doi.org/10.3390/MICROORGANISMS12101938>
38. Jolley KA, Bliss CM, Bennett JS, et al (2012) Ribosomal multilocus sequence typing: universal characterization of bacteria from domain to strain. *Microbiology (Reading)* 158:1005–1015. <https://doi.org/10.1099/MIC.0.055459-0>
39. Kalyaanamoorthy S, Minh BQ, Wong TKF, et al (2017) ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods* 2017 14:6 14:587–589. <https://doi.org/10.1038/nmeth.4285>
40. Katoh K, Standley DM (2013) MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol Biol Evol* 30:772–780. <https://doi.org/10.1093/MOLBEV/MST010>
41. Lawson DM, Stevenson CEM (2012) Structural and functional dissection of aminocoumarin antibiotic biosynthesis: a review. *J Struct Funct Genomics* 13:125–133.

<https://doi.org/10.1007/S10969-012-9138-2>

42. Lee I, Chalita M, Ha SM, et al (2017) ContEst16S: An algorithm that identifies contaminated prokaryotic genomes using 16S RNA gene sequences. *Int J Syst Evol Microbiol* 67:2053–2057. <https://doi.org/10.1099/IJSEM.0.001872>,
43. Letunic I, Bork P (2024) Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. *Nucleic Acids Res* 52:W78–W82. <https://doi.org/10.1093/nar/gkae268>
44. Li L, MacIntyre LW, Brady SF (2021) Refactoring biosynthetic gene clusters for heterologous production of microbial natural products. *Curr Opin Biotechnol* 69:145–152. <https://doi.org/10.1016/J.COPBIO.2020.12.011>
45. Liu B, Zheng D, Zhou S, et al (2022) VFDB 2022: a general classification scheme for bacterial virulence factors. *Nucleic Acids Res* 50:D912–D917. <https://doi.org/10.1093/NAR/GKAB1107>
46. Madej T, Lanczycki CJ, Zhang D, et al (2014) MMDB and VAST+: tracking structural similarities between macromolecular complexes. *Nucleic Acids Res* 42:D297-303. <https://doi.org/10.1093/NAR/GKT1208>
47. Makarova KS, Wolf YI, Koonin E V. (2018) Classification and Nomenclature of CRISPR-Cas Systems: Where from Here? *CRISPR J* 1:325–336. <https://doi.org/10.1089/CRISPR.2018.0033>
48. Meier-Kolthoff JP, Carbasse JS, Peinado-Olarte RL, 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. <https://doi.org/10.1093/NAR/GKAB902>
49. Mikheenko A, Prjibelski A, Saveliev V, et al (2018) Versatile genome assembly evaluation with QUAST-LG. *Bioinformatics* 34:i142–i150. <https://doi.org/10.1093/BIOINFORMATICS/BTY266>
50. Miranda RVLS, Costa LV, Albuquerque LS, et al (2023) Identification of *Sutcliffiella horikoshii* strains in an immunobiological pharmaceutical industry facility. *Lett Appl Microbiol* 76:. <https://doi.org/10.1093/LAMBIO/OVAD056>
51. Moreira FM, Pereira P de A, Miranda RV da SL de, et al (2023) Evaluation of MALDI-TOF MS, sequencing of D2 LSU rRNA and internal transcribed spacer regions (ITS) for the identification of filamentous fungi isolated from a pharmaceutical facility. *J Pharm Biomed Anal* 234:115531. <https://doi.org/10.1016/J.JPBA.2023.115531>
52. O’Leary NA, Cox E, Holmes JB, et al (2024) Exploring and retrieving sequence and metadata for species across the tree of life with NCBI Datasets. *Sci Data* 11(1):732-. <https://doi.org/10.1038/s41597-024-03571-y>
53. Orakov A, Fullam A, Coelho LP, et al (2021) GUNC: detection of chimerism and contamination in prokaryotic genomes. *Genome Biol* 22(1):178-. <https://doi.org/10.1186/S13059-021-02393-0>
54. Parental Drug Association (2015) Technical Report No. 69: Bioburden and Biofilm Management in Pharmaceutical Manufacturing Operations | PDA. <https://www.pda.org/bookstore/product-detail/2789-tr-69-bioburden-biofilm>. Accessed 14 Jun 2026
55. Parks DH, Imelfort M, Skennerton CT, et al (2015) CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* 25:1043–1055.

<https://doi.org/10.1101/GR.186072.114>

56. Price MN, Dehal PS, Arkin AP (2010) FastTree 2—approximately maximum-likelihood trees for large alignments. PLoS One 5(3):e94905. <https://doi.org/10.1371/JOURNAL.PONE.0009490>
57. Prjibelski A, Antipov D, Meleshko D, et al (2020) Using SPAdes De Novo Assembler. Curr Protoc Bioinformatics 70:e102. <https://doi.org/10.1002/CPBI.102>
58. Ramos JN, Baio PVP, Veras JFC, et al (2023) Novel configurations of type I-E CRISPR-Cas system in *Corynebacterium striatum* clinical isolates. Braz J Microbiol 54:69–80. <https://doi.org/10.1007/S42770-022-00881-4>
59. Ramos JN, Veloso Da Costa L, Vieira VV, et al (2025) Challenges in the Identification of Environmental Bacterial Isolates from a Pharmaceutical Industry Facility by 16S rRNA Gene Sequences. DNA 2025, Vol 5, Page 33 5:33. <https://doi.org/10.3390/DNA5030033>
60. Richter M, Rosselló-Móra R, Oliver Glöckner F, Peplies J (2016) JSpeciesWS: a web server for prokaryotic species circumscription based on pairwise genome comparison. Bioinformatics 32:929–931. <https://doi.org/10.1093/BIOINFORMATICS/BTV681>
61. Riesco R, Trujillo ME (2024) Update on the proposed minimal standards for the use of genome data for the taxonomy of prokaryotes. Int J Syst Evol Microbiol 74(3):006300. <https://doi.org/10.1099/IJSEM.0.006300>
62. Ronquist F, Teslenko M, Van Der Mark P, et al (2012) MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst Biol 61:539–542. <https://doi.org/10.1093/SYSBIO/SYS029>
63. Schoch CL, Ciufo S, Domrachev M, et al (2020) NCBI Taxonomy: a comprehensive update on curation, resources and tools. Database (Oxford) 2020:baaa062. <https://doi.org/10.1093/database/baaa062>
64. Seemann T (2014) Prokka: rapid prokaryotic genome annotation. Bioinformatics 30(14):2068–2069. <https://doi.org/10.1093/BIOINFORMATICS/BTU153>
65. Sharma S, Mohler J, Mahajan SD, et al (2023) Microbial Biofilm: A Review on Formation, Infection, Antibiotic Resistance, Control Measures, and Innovative Treatment. Microorganisms 11:.. <https://doi.org/10.3390/MICROORGANISMS11061614>
66. Simão FA, Waterhouse RM, Ioannidis P, et al (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31(19):3210–3212. <https://doi.org/10.1093/BIOINFORMATICS/BTV351>
67. Song M, Li Q, Liu C, et al (2024) A comprehensive technology strategy for microbial identification and contamination investigation in the sterile drug manufacturing facility—a case study. Front Microbiol 15:1327175. <https://doi.org/10.3389/fmicb.2024.1327175>
68. Takeuchi M, Fang CX, Yokota A (1995) Taxonomic study of the genus *Brachybacterium*: Proposal of *Brachybacterium conglomeratum* sp. nov., nora. rev., *Brachybacterium paraconglomeratum* sp. nov., and *Brachybacterium rhamnosum* sp. nov. Int J Syst Bacteriol 45:160–168. <https://doi.org/10.1099/00207713-45-1-160/CITE/REFWORKS>

69. Tamura K, Stecher G, Kumar S (2021) MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Mol Biol Evol* 38:3022–3027. <https://doi.org/10.1093/MOLBEV/MSAB120>
70. Tatusova T, Dicuccio M, Badretdin A, et al (2016) NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res* 44:6614–6624. <https://doi.org/10.1093/NAR/GKW569>
71. Tegenfeldt F, Kuznetsov D, Manni M, et al (2025) OrthoDB and BUSCO update: annotation of orthologs with wider sampling of genomes. *Nucleic Acids Res* 53:D516–D522. <https://doi.org/10.1093/NAR/GKAE987>
72. Thompson CC, Chimetto L, Edwards RA, et al (2013) Microbial genomic taxonomy. *BMC Genomics* 14:913. <https://doi.org/10.1186/1471-2164-14-913>
73. Vasconcellos L, Silva SV da, Costa LV da, et al (2026) Phenotypical and Molecular Characterization of *Pseudomonas* spp. Isolated from a Pharmaceutical Facility. *Processes* 14:231. <https://doi.org/10.3390/PR14020231>
74. Viana MVC, Figueiredo H, Ramos R, et al (2017) Comparative genomic analysis between *Corynebacterium pseudotuberculosis* strains isolated from buffalo. *PLoS One* 12:e0176347. <https://doi.org/10.1371/JOURNAL.PONE.0176347>
75. Walker BJ, Abeel T, Shea T, et al (2014) Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9(11):e112963. <https://doi.org/10.1371/JOURNAL.PONE.0112963>
76. Wong TKF, Ly-Trong N, Ren H, et al (2025) IQ-TREE 3: Phylogenomic Inference Software using Complex Evolutionary Models. 43(5):msag117. <https://doi.org/10.32942/X2P62N>
77. Yamada Y, Kuzuyama T, Komatsu M, et al (2015) Terpene synthases are widely distributed in bacteria. *Proc Natl Acad Sci U S A* 112:857–862. <https://doi.org/10.1073/PNAS.1422108112;PAGE:STRING:ARTICLE/CHAPTER>
78. Zhang D, Gao F, Jakovlić I, et al (2020) PhyloSuite: An integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. *Mol Ecol Resour* 20:348–355. <https://doi.org/10.1111/1755-0998.13096>
79. Zhang Y, Ma Y, Xia X, et al (2025) *sodA* modulates in vitro and in vivo virulence of *Yersinia enterocolitica*. *Front Microbiol* 16:1643172. <https://doi.org/10.3389/FMICB.2025.1643172/BIBTEX>

Figures

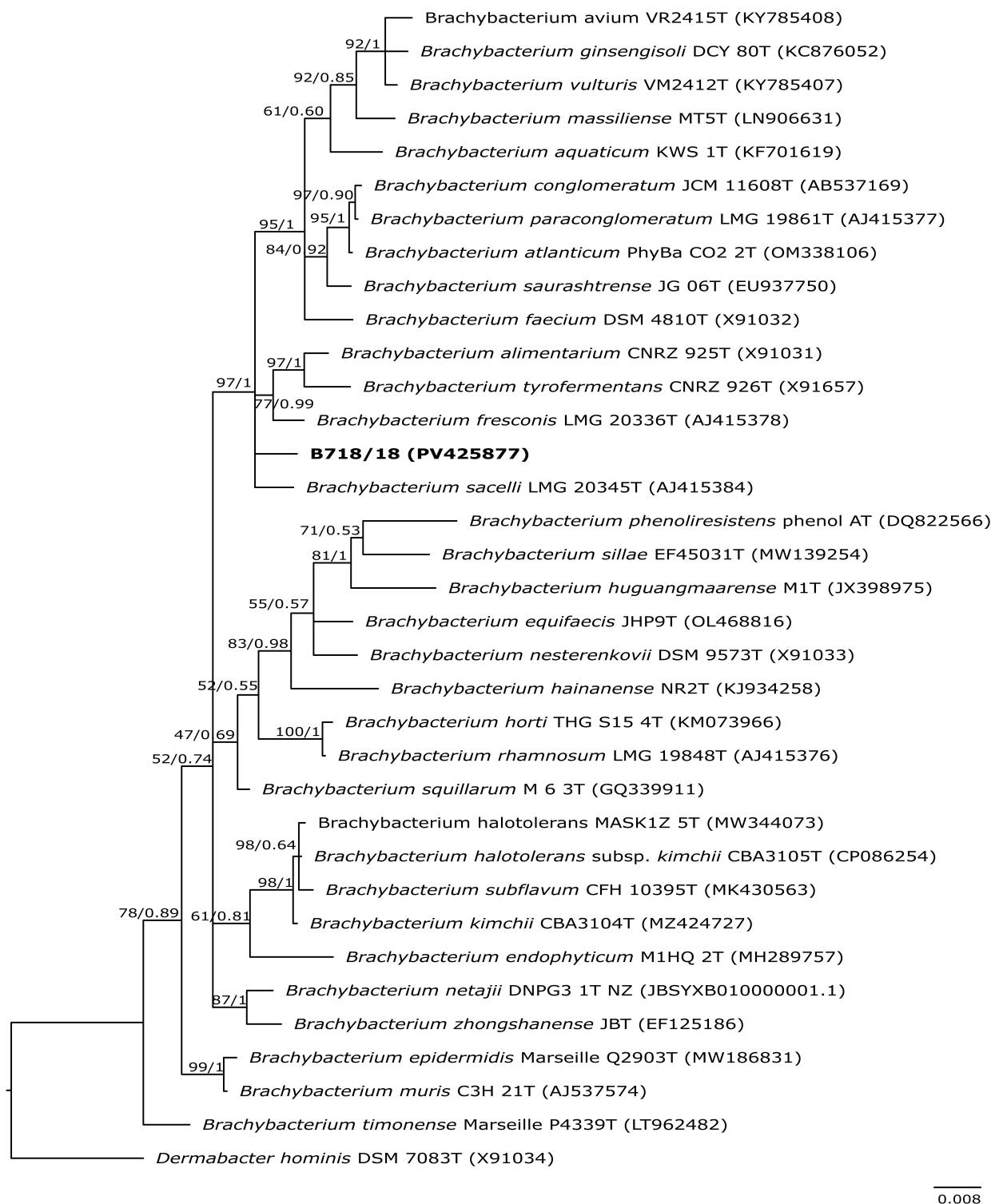


Figure 1

Phylogenetic tree based on 16S rRNA gene sequences of *Brachybacterium* species. The consensus tree, inferred from 1,000 replicates, was reconstructed using maximum likelihood and Bayesian inference. Bootstrap values and posterior probabilities are shown next to the branches. The evolutionary model used was General Time Reversible, with rate heterogeneity G+ and I+ category 4. The type strain of *Dermabacter hominis* DMS7083^T was used as an outgroup.

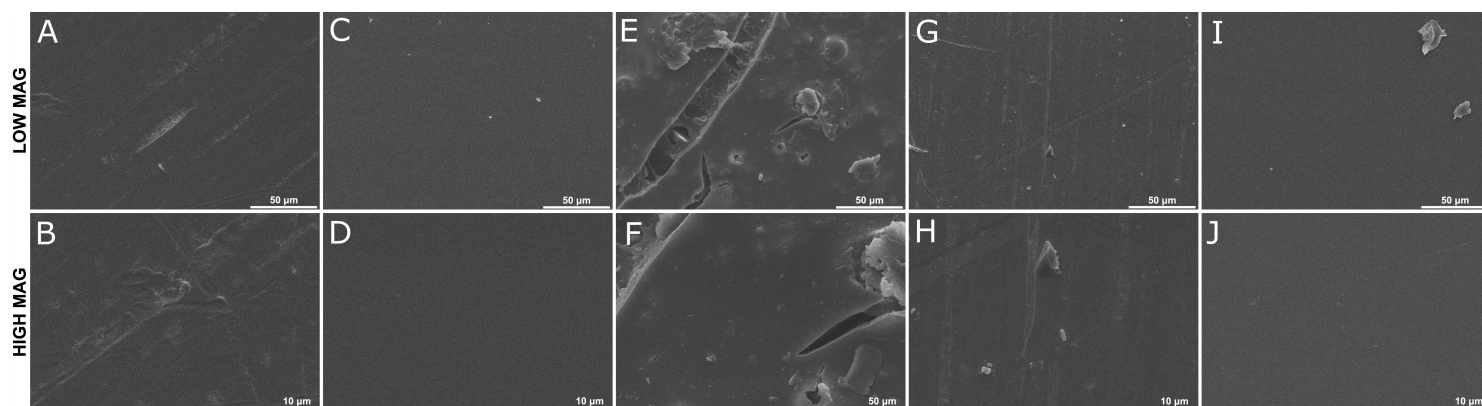


Figure 2

Negative biofilm formation for bacterial isolate B718/18 on different surfaces analyzed by scanning electron microscopy. (A and B) - Negative control (brain-heart infusion broth in stainless steel); (C and D) - Glass; (E and F) - Vinyl coating; (G and H) - Stainless steel; (I and J) - Acrylic; (H) - Bacterial cell characteristic of isolate B178/18, with coccobacillus morphology.

Distribution of functional categories - RAST Annotation
Total number of annotated genes: 827

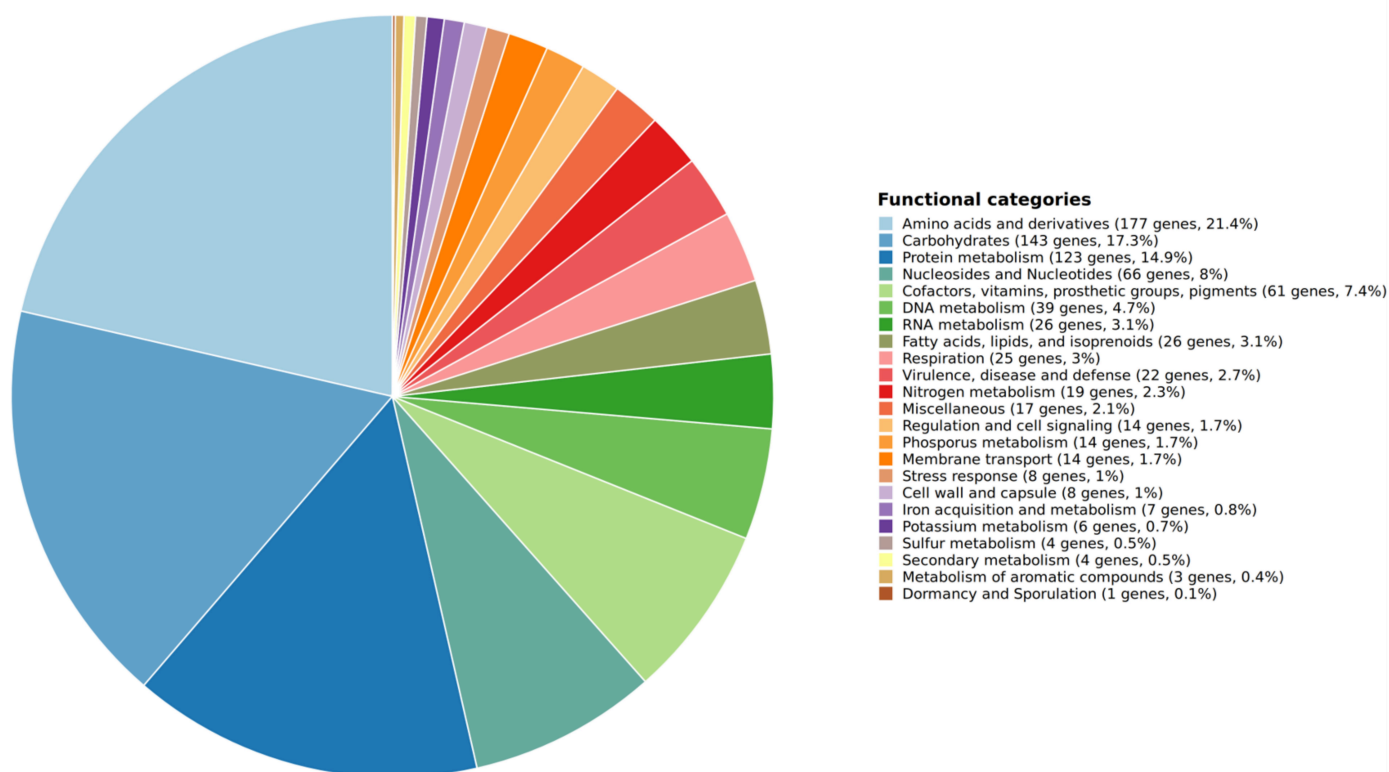


Figure 3

Distribution of functional categories of genes annotated by RAST (n = 827). The most represented functions include amino acid and derivative metabolism (21.4%), carbohydrate metabolism (17.3%), and protein metabolism (14.9%), followed by nucleotide, cofactor, and DNA metabolism, while the remaining categories account for smaller proportions of the genome.

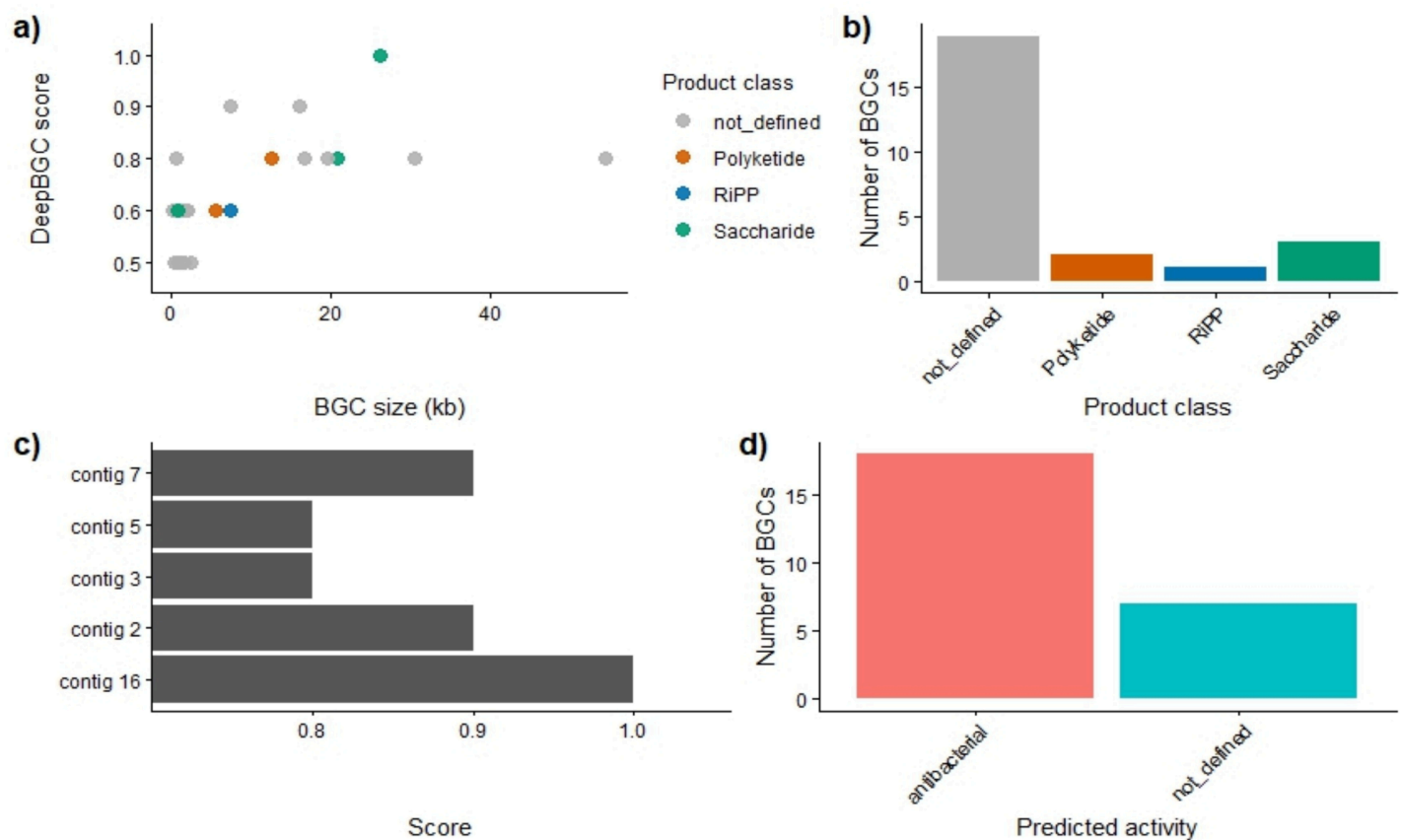


Figure 4

BGCs predicted by the DeepBGC program. Distribution of BGC sizes (kb) by product class and score (a). Number of BGCs classified by product class (b). Highest-scoring BGCs and their locations on the contigs (c). Number of BGCs classified by product activity (d).

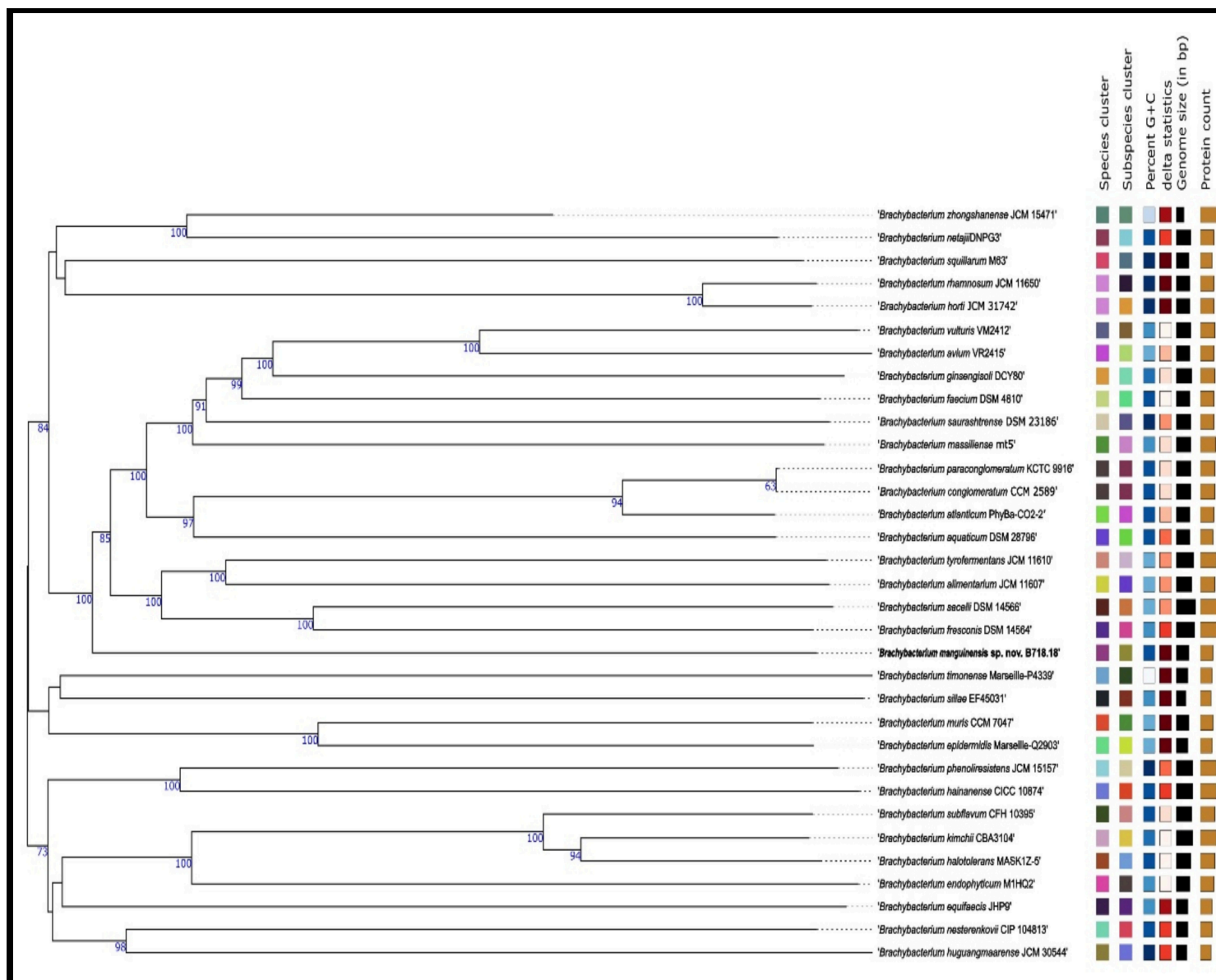


Figure 5

Phylogenomic tree showing the relationship between the isolate B718/18 and the other *Brachybacterium* species, based on whole-genome sequences of type strains described in the Table 1. Genomic sequences were analyzed by TYGS using standard parameters. Tree inferred with FastME 2.1.6.1 based on GBDP distances calculated from the genomic sequences. Branch lengths are scaled according to the GBDP distance formula D5. The numbers above the branches are GBDP pseudo-bootstrap support values (>60%) from 100 replicates.

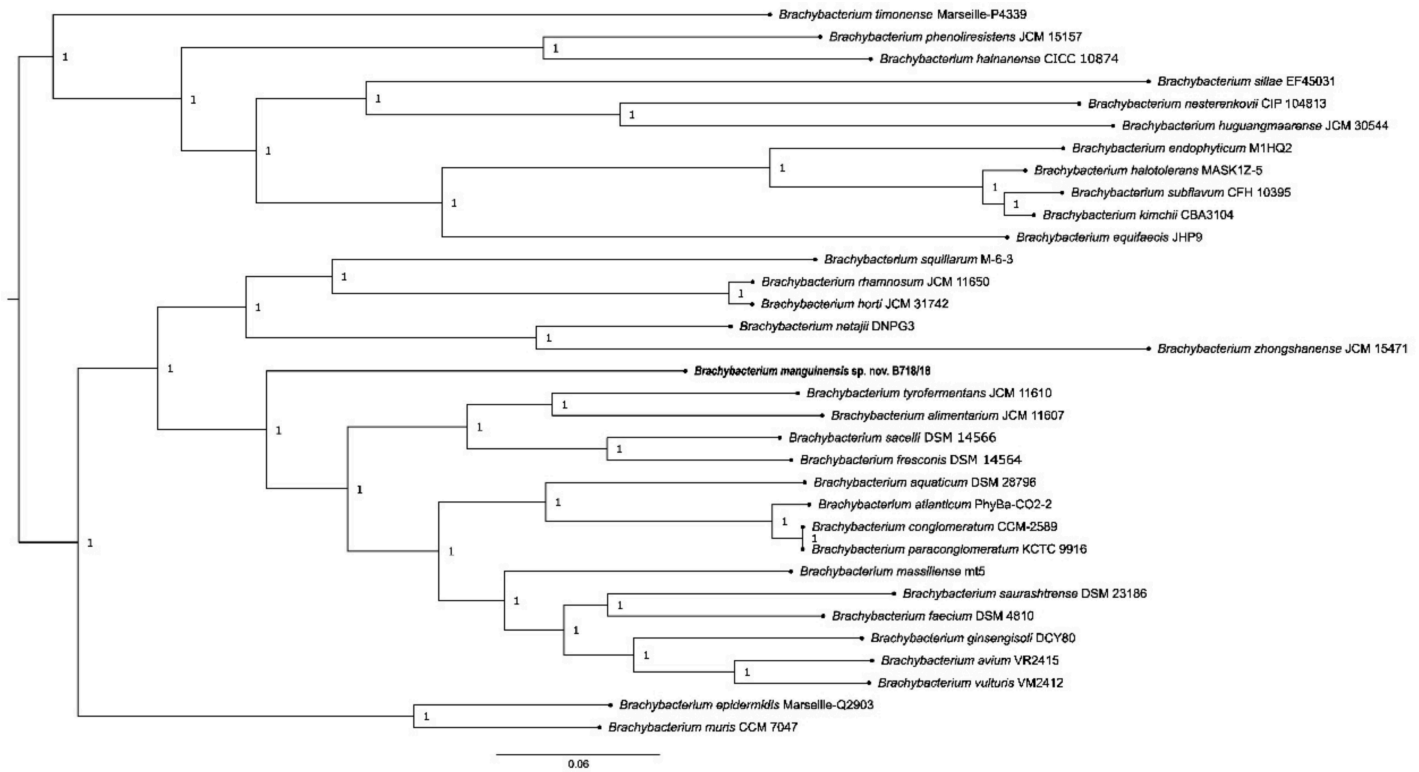


Figure 6

Phylogenomic tree of the genus *Brachybacterium* inferred from 357 shared orthologous proteins, constructed based on the orthologous protein groups identified by OrthoFinder tool. The values at the nodes indicate the statistical support of the tree. The scale bar represents the number of substitutions per site. The accession numbers of type strains genomes are described in Table 1.

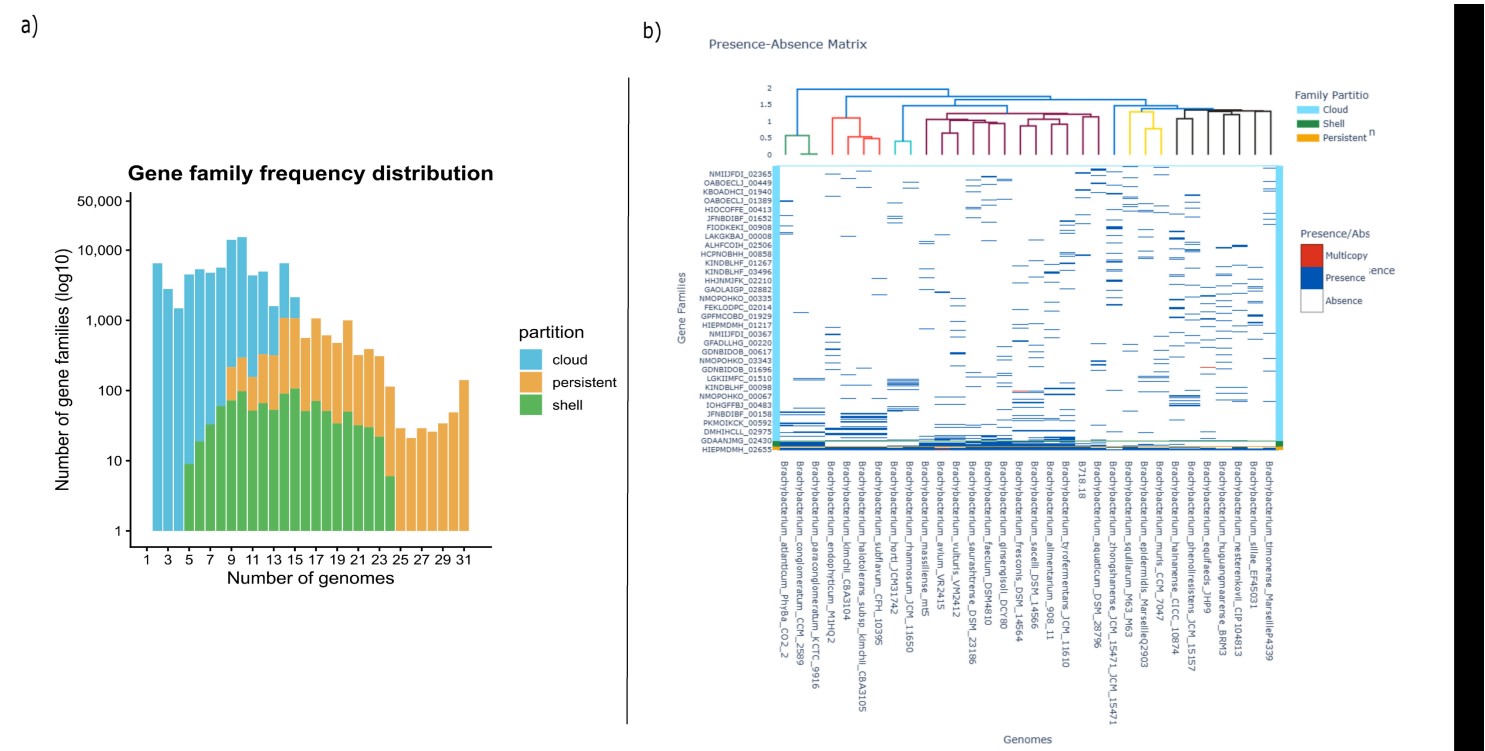


Figure 7

Pangenome accumulation curve (a). Frequency distribution of gene families in the cloud, shell, and persistent genomes (b). Presence-absence matrix for genes in the cloud, shell, and persistent genomes.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [TableS1.docx](#)
- [TableS2.docx](#)