Antonie van Leeuwenhoek (2026) 119:16
https://doi.org/10.1007/s10482-025-02214-5

ORIGINAL PAPER

Sphingobium salicis sp. nov., a novel endophyte with two subspecies from pioneer plants growing on primary substrates

Robert J. Tournay · John L. Freeman · Misha Levish · Douglas Baker · Andrea Firrincieli · Sharon L. Doty

Received: 19 June 2025 / Accepted: 17 November 2025 / Published online: 19 December 2025

Abstract Three endophytic *Sphingobium* strains, WW5^T, 11R-BB, and HT1-2^T, were isolated from *Salix sitchensis* stems, *Populus trichocarpa* roots, and *Pluchea carolinensis* shoots, respectively, growing on nutrient-limited rock substrates undergoing primary succession. Cells are Gram-negative, strictly aerobic, rod-shaped, catalase-positive, and motile by a single flagellum. Colonies formed within 2–3 days on nitrogen-limited combined carbon medium (NLCCM) and mannitol-glutamate/Luria (MGL) media. Growth occurred at 25–35 °C (optimum 30–32.5 °C), pH 5–8, and up to 2% NaCl for strains WW5^T and 11R-BB and 3% for strain HT1-2^T. Cellular fatty acid profiles were dominated by $C_{18:1}$ ω7c and $C_{16:1}$ ω7c. The predominant respiratory quinone was ubiquinone-10, and the principal polar lipids were phosphatidylethanolamine, sphingoglycolipids, and phosphatidylglycerol. Hybrid genome assemblies ranged from 5.49–5.72 Mb with 64.0–64.2 mol% GC, and 5404–5635 coding sequences. The 16S rRNA gene sequences of strains WW5^T, 11R-BB, and HT-12^T showed 99.5–99.9% identity to *S. yanoikuyae* JCM 7371^T, but formed a separate lineage in neighbor-joining, maximum-likelihood, and maximum-parsimony trees. Pairwise digital DNA–DNA hybridization (dDDH; 99.8%), average nucleotide identity based on BLAST (ANIb; 99.98%), and alignment fraction (AF; 99.5%) indicate that strains WW5^T and 11R-BB are conspecific and represent the same subspecies. Strain HT1-2^T showed lower relatedness (70% dDDH; 95.05–95.31% ANIb; 76–79% AF) but was more closely related to strains WW5^T and 11R-BB than to *S. yanoikuyae* JCM 7371^T, supporting its placement as a subspecies within the same species as strains WW5^T and 11R-BB. Based on phylogenomic, genomic, and phenotypic evidence, these strains constitute a novel species, *Sphingobium salicis* sp. nov., with the strain WW5^T (DSM 120182^T, NCCB 101075^T) designated as the type strain.

Keywords *Sphingobium* · Bacterial plant endophyte · Willow phyllosphere · Poplar rhizosphere · Salvia phyllosphere taxonomy · Metabolism

Robert J. Tournay and John L. Freeman has contributed equally to this work.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10482-025-02214-5.

R. J. Tournay (✉) · S. L. Doty
University of Washington, Seattle, WA, United States of America
e-mail: tournay@uw.edu

J. L. Freeman · M. Levish · D. Baker
Intrinsyx Bio, 1237 Midas Way, Sunnyvale, CA, United States of America

A. Firrincieli
University of Tuscia, Viterbo, Italy

Introduction

The genus *Sphingobium* belongs to the family *Sphingomonadaceae* (Phylum *Pseudomonadota*) and is characterized by glycosphingolipid-containing outer membranes. Established during the reorganization of *Sphingomonas* into four genera (*Sphingomonas* sensu stricto, *Sphingobium*, *Novosphingobium*, *Sphingopyxis*) (Takeuchi et al. 2001), it currently includes 48 validly published species (LPSN; Parte et al. 2020). The type species, *Sphingobium yanoikuyae*, was first isolated by Gibson et al. (1975) and initially assigned to *Beijerinckia*, later reclassified as *Sphingomonas yanoikuyae* (Yabuuchi et al. 1990), and transferred to *Sphingobium* (Takeuchi et al. 2001). Other well-studied members, such as *S. indicum*, *S. japonicum*, and *S. herbicidovorans,* are noted for their ability to degrade recalcitrant aromatics (Glaeser and Kämpfer 2015).

Sphingobium species are widespread in soils, freshwater, sediments, and hydrocarbon-rich sites, where they degrade biphenyl, naphthalene, and chlorinated aromatics (Gibson et al. 1975; Khan et al. 1996; Pinyakong et al. 2003; Seo et al. 2009; Stolz 2009; Táncsics et al. 2012; Mitra et al. 2020). Beyond contaminated habitats, they have been reported as plant-associated bacteria, including endophytes detected in maize, rice, tomato, and coastal dune grasses (Ali et al. 2018; Boss et al. 2022; Jou et al. 2022; Ma et al. 2025). Molecular studies further document *Sphingobium* in diverse plant hosts (Shaik and Thomas 2019; Arunthavasu et al. 2019), and *Sphingobium* strains have been identified in the rhizosphere of pioneer grasses colonizing newly deposited volcanic ash (Rincón-Molina et al. 2020). Despite these reports, the only validly published endophytic species is *S. endophyticum* from the roots of *Hylomecon japonica* (Zhu et al. 2015; Oren 2024).

Endophytes can improve host nutrient acquisition and stress tolerance through mechanisms such as nitrogen fixation and phosphorus solubilization (Reinhold-Hurek and Hurek 2011; Hardoim et al. 2015; Ercole et al. 2025). We recently demonstrated that strains WW5^T, 11R-BB, and HT1-2^T enhance nitrogenase activity in diazotrophic partners under nitrogen limitation, potentially through mitigation of nitrogenase oxygen sensitivity or diffusible compounds (Sher et al. 2024a). Strains WW5^T and 11R-BB were isolated from willow (*Salix sitchensis*) and poplar (*Populus trichocarpa*) on riparian banks in Washington State, while strain HT1-2^T was isolated from sourbush (*Pluchea carolinensis*) on volcanic substrates in Hawai'i. All strains were recovered from surface-sterilized plant tissues collected from nutrient-limited primary successional habitats.

Genome-based classification shows that taxonomic resolution within *Sphingobium* remains incomplete. For example, the Genome Taxonomy Database Toolkit (GTDB-Tk, release 226, accessedApril 2025) identifies 137 species clusters across 381 genomes, but only 35% (48/137) correspond to validly described species, leaving many lineages unresolved despite the genus's ecological and functional diversity. Phylogenetic analysis based on 16S rRNA gene sequences placed strains WW5^T, 11R-BB, and HT1-2^T closest to *S. yanoikuyae* JCM 7371^T. Although the 16S rRNA gene identities exceeded species thresholds (98.7%), genome-based analyses showed the strains form a distinct lineage. Phylogenomic and phenotypic data support recognition of *Sphingobium salicis* sp. nov., comprising subsp. *salicis* (type strain WW5^T) and subsp. *pluchea* (type strain HT1-2^T).

Methods

Isolation and culture conditions

Strains WW5^T, 11R-BB, and HT1-2^T were isolated from plant hosts growing on volcanic and alluvial substrates (Doty et al. 2009; Sher et al. 2024b). Strain HT1-2^T was isolated from sourbush shoots (*Pluchea carolinensis*, *Asteraceae*) collected from lava fields near Kona, Hawai'i (19.1426° N, 155.7521° W). The strains WW5^T and 11R-BB were isolated from plants growing in cobblestone-dominated riparian zones in Washington state. Strain WW5^T was isolated from willow stems (*Salix sitchensis*, *Salicaceae*) along the Snoqualmie River (47.5206° N, 121.7745° W), and strain 11R-BB was isolated from poplar roots (*Populus trichocarpa*, *Salicaceae*) along the Skykomish River (47.7806° N, 121.5007° W).

The sourbush tissues were surface-sterilized according to Sher et al. (2024b), and the poplar and willow tissues according to Doty et al. (2005). Sterilized plant tissues were homogenized in mortars with either water (sourbush) or nitrogen-free medium (poplar and willow). Homogenates (100 µL) were

plated on nitrogen-free combined carbon medium (NFCCM, Rennie 1981) for sourbush and poplar tissues, or yeast mannitol agar (YMA) for willow tissues. Distinct colonies were purified by successive streaking on mannitol glutamate/Luria (MGL) agar (Cangelosi et al. 1991) and incubated at 30 °C for 24–48 h. Pure isolates were preserved at −80 °C in 25% glycerol.

16S rRNA gene sequencing, identification, and phylogenetic analysis

The genomic DNA of the strains WW5^T, 11R-BB, and HT1-2^T was purified using phenol/chloroform extraction and ethanol precipitation (Doty et al. 2023). The 16S rRNA gene was amplified using the universal primers 8F and 1492R (Frank et al. 2008). PCR amplification conditions were: 95 °C for 10 min; 40 cycles of 95 °C for 1 min, 53 °C for 1 min, 72 °C for 1 min, and a final extension at 72 °C for 10 min, and the amplicons sequenced by Genewiz (Azenta Life Sciences, NJ). Phylogenetic analysis was conducted using 16S rRNA sequences of the three strains and 37 validly published reference strains obtained from EzBioCloud (Yoon et al. 2017; Chalita et al. 2024). Multiple sequence alignments were trimmed to 1393 bp and analyzed using neighbor-joining (NJ), maximum likelihood (ML), and maximum parsimony (MP) methods in *MEGA* 12 (Kumar et al. 2024). Evolutionary distances were calculated using the Kimura 2-parameter model with gamma-distributed rate variation (Kimura 1980). Bootstrap support (1000 replicates) values ≥ 80% were regarded as significant support. All trees were rooted using *Sphingomonas paucimobilis* NBRC 13935^T.

Whole genome sequencing, assembly, and annotation

Genomic DNA from strains WW5^T, 11R-BB, and HT1-2^{T}was extracted using a ZymoBIOMICS DNA Miniprep Kit (Zymo Research, Irvine, CA, USA), and sequenced using both Illumina and Oxford Nanopore Technologies (CD Genomics, New York, NY, USA). Genome assembly and annotation were performed using the Comprehensive Genome Analysis (CGA) pipeline through the Bacterial and Viral Bioinformatics Resource Center (BV-BRC) (Olson et al. 2023). Hybrid assemblies were generated with Unicycler v0.4.8 using SPAdes v3.13.0 for short-read processing (Bankevich et al. 2012; Wick et al. 2017), followed by consensus refinement with Racon v1.4.20 (Vaser et al. 2017). Contigs < 1000 bp in length or with < 10 × coverage were excluded. Assembly quality was assessed with CheckM v1.0.18 (Parks et al. 2015), and annotation was performed with RASTtk v1.073 (Brettin et al. 2015).

Whole-genome phylogeny and taxonomic placement

Genome-based taxonomic placement of strains WW5^T, 11R-BB, and HT1-2^T was assessed using three complementary whole-genome approaches: BLASTn-based average nucleotide identity (ANIb) (Richter and Rosselló-Móra 2009; Richter et al. 2016), digital DNA–DNA hybridization (dDDH) calculated with the Genome BLAST Distance Phylogeny (GBDP) method implemented in the Type Strain Genome Server (TYGS) (Meier-Kolthoff and Göker 2019), and phylogenomic reconstruction using the Genome Taxonomy Database Toolkit (GTDB-Tk) (Parks et al. 2022; Chaumeil et al. 2022). *Sphingomonas paucimobilis* NBRC 13935^T was used as the outgroup for TYGS and GTDB-Tk analyses.

The datasets for ANIb and TYGS comprised the same genome dataset. All validly published *Sphingobium* species were identified from the List of Prokaryotic names with Standing in Nomenclature (LPSN) (Parte et al. 2020, accessed October 2025). Type strains were verified using BacDive (https://bacdive.dsmz.de/), and corresponding genome assemblies were retrieved from NCBI RefSeq. Species lacking whole-genome assemblies were excluded. The final dataset included 43 *Sphingobium* type strains, the three study strains, and *S. paucimobilis* NBRC 13935^T (46 genomes in total).

Average nucleotide identity was calculated using the BLASTn-based ANIb method in JSpeciesWS (Richter et al. 2016), which employs 1020 bp sequence fragments for pairwise comparisons (Richter and Rosselló-Móra 2009). Both ANIb values and alignment fractions (AF)—the proportion of the smaller genome that aligns with the larger one in each comparison—were reported. Species-level relatedness was interpreted using the thresholds of ≥ 95–96% ANIb (Richter and Rosselló-Móra 2009) and ≥ 0.70 AF (Varghese et al. 2015).

The 46-genome dataset was submitted to the Type Strain Genome Server (TYGS) for

whole-genome–based taxonomic assignment. Pairwise digital DNA–DNA hybridization (dDDH) values were calculated using the Genome BLAST Distance Phylogeny (GBDP) (Meier-Kolthoff and Göker 2019). A whole-genome GBDP tree was inferred with FASTME 2.1.6.1, branch lengths representing GBDP distance formula d_5, and pseudo-bootstrap support values calculated from 100 replicates.

Phylogenomic analysis was conducted with GTDB-Tk (release R220), using the Bac120 marker gene set comprising 120 universally conserved bacterial single-copy genes (Parks et al. 2022; Chaumeil et al. 2022). Marker genes were extracted from the three study strains, concatenated, and aligned with corresponding sequences from *Sphingobium* species representatives in GTDB. Maximum-likelihood phylogeny was inferred in IQ-TREE (Nguyen et al. 2015) under the WAG substitution model with 1000 bootstrap replicates. Placement into the GTDB reference tree was performed with pplacer (Matsen et al. 2010). The tree was rooted with *S. paucimobilis* NBRC 13935^T, and node support was reported for values ≥ 80%.

Comparative genomics and functional analysis

Comparative genomic analyses were performed using the BV-BRC Comparative Systems service (Olson et al. 2023). The dataset comprised 46 genomes: the strains WW5^T, 11R-BB, HT1-2^T and 43 validly published *Sphingobium* type strains (as described in the TYGS methods). Orthologous genes were identified using PATRIC global protein families (PGfams), which are constructed across genera by the PATtyFams algorithm through k-mer similarity searches and Markov clustering (MCL), using thresholds of ≥ 50% amino acid identity over ≥ 80% alignment coverage (Davis et al. 2016). For comparison, genus-specific protein families (PLfams; ≥ 80% identity, ≥ 80% coverage, paralogs assigned separately) were also generated. Protein family presence/absence matrices were used to identify conserved (core) families present in all of the genomes, families unique to the study strains, and differences between the study strains and the most closely related type strain, *S. yanoikuyae* JCM 7371^T.

For the three strains and *S. yanoikuyae* JCM 7371^T, additional functional annotation was performed using KEGG KOfam (v2025-04-01, release 114.0; (Aramaki et al. 2020). Annotated contig FASTA files were submitted to the KofamKOALA webserver, and KEGG orthology assignments were integrated with BV-BRC family-based clustering to assess gene conservation and metabolic pathway completeness. Biosynthetic gene clusters (BGCs) were identified using antiSMASH v8.0, with the "relaxed" detection strictness setting. The pathogenic potential and virulence determinants of the strains were predicted in silico using PathogenFinder v1.1 (Cosentino et al. 2013) and VirulenceFinder v2.0 (Joensen et al. 2014). VirulenceFinder was run with default thresholds (≥ 90% nucleotide identity; ≥ 60% alignment length) against the *Staphylococcus aureus*, *Enterococcus*, *Listeria*, and *Escherichia coli* databases.

Phenotypic and chemotaxonomic characterizations

Cells of strains WW5^T, 11R-BB, and HT1-2^T were grown for 3 days at 25 °C in nitrogen-limited combined carbon medium (NLCCM, modified from Rennie 1981). Gram staining was performed with a commercial kit (Innovating Science, Avon, NY, USA), and cells were examined by brightfield microscopy (NuLife Sciences, Chattanooga, TN, USA) using oil immersion at 1000 × magnification. Cell shape and arrangement were recorded, and the presence and distribution of flagella were determined using the Ryu staining method (Heimbrook et al. 1989). Colony morphology was assessed on MGL and NLCCM agar after 2 and 3 days of incubation, respectively.

Growth of the strains WW5^T, 11R-BB, and HT1-2^T in nitrogen-free combined carbon media (NFCCM) was evaluated at 25 °C for up to 7 days with daily monitoring. Oxygen requirements were determined in fluid thioglycolate medium (Carolina, Burlington, NC, USA) after 2 days of incubation at 25 °C. Growth at 4, 25, 30, 35, and 37 °C was tested on MGL agar plates incubated for up to 7 days. Optimal growth was determined in NLCCM liquid medium at 25, 27.5, 30, 32.5, 35, and 37 °C by measuring the OD_{600} values after 2 days of incubation (SPECTRAmax Plus 384, Molecular Devices, CA, USA). Carbon source utilization, osmotic stress tolerance, and pH tolerance were assessed using Biolog MicroPlate™ PM1, PM9, and PM10 (Biolog, Hayward, CA, USA), respectively, according to the manufacturer's instructions.

Biomass for chemotaxonomic analyses was obtained from strains WW5^T, 11R-BB, and HT1-2^T, and *S. yanoikuyae* JCM 7371^T, which were cultivated

aerobically at 25 °C for 24 h on TSA (medium 535; DSMZ). Cellular fatty acid composition was determined by fatty acid methyl ester (FAME) analysis using the MIDI Microbial Identification System (Kunitsky et al. 2006). Respiratory quinones were extracted from fresh biomass with hexane, purified by silica-based solid phase extraction, and analyzed by HPLC–DAD on a reverse-phase column, with ubiquinones detected at 270 nm and menaquinones at 326 nm (Vieira et al. 2021). Polar lipids were extracted according to Bligh and Dyer (1959) and separated by two-dimensional thin-layer chromatography as described by Tindall et al. (2010), with lipid spots visualized using molybdophosphoric acid and group-specific spray reagents. All chemotaxonomic analyses were performed by DSMZ Services (Leibniz-Institute DSMZ, Braunschweig, Germany).

Results and discussion

Ecological context of the strains

Strains WW5^T, 11R-BB, and HT1-2^T were isolated from plants inhabiting nutrient-limited environments undergoing primary succession. Strains WW5^T and 11R-BB were obtained from *Salix sitchensis* stems and *Populus trichocarpa* roots, respectively, collected from cobble-dominated riparian banks along the Skykomish and Snoqualmie Rivers in Washington State, where frequent flooding and nutrient deficits occur during early successional stages (Doty et al. 2009). Strain HT1-2^T was isolated from shoots of *Pluchea carolinensis* growing on recently formed volcanic lava flows in Hawai'i, where minimal soil development constrains nutrient availability.

The strains were recovered from surface-sterilized plant tissues following validated protocols (Doty et al. 2005), and their endophytic lifestyle has been confirmed through independent functional and molecular studies. Internal colonization of poplar roots by strains WW5^T and 11R-BB was demonstrated using digital droplet PCR with strain-specific primers (Aufrecht et al. 2025), satisfying criteria for endophyte classification beyond surface sterilization. All three strains substantially enhanced nitrogenase activity of diazotrophic partners under nitrogen-limited conditions, as measured by acetylene reduction assays (Sher et al. 2024b). When applied in plant consortia, strain WW5^T was associated with improved water-use efficiency in drought-stressed rice (Rho et al. 2018) and enhanced fruit quality in apple (Rho et al. 2020). The isolation of these strains from nutrient-limited primary successional habitats, combined with documented plant-growth-promoting activities, motivated comparative genomic analysis to identify metabolic and functional features distinguishing these strains from other environmental *Sphingobium* species.

16S rRNA gene phylogeny

Strains WW5^T, 11R-BB, and HT1-2^T were placed within the genus *Sphingobium* based on their 16S rRNA gene sequences. Sequence identities to *S. yanoikuyae* JCM 7371^T ranged from 99.5 to 99.9%, and to *S. scionense* WP01^T from 99.0–99.4% (Table S1), all above the 98.7% threshold often used for species delineation (Yarza et al. 2014). NJ, ML, and MP trees of the 16S rRNA sequences consistently placed the three strains in a clade with *S. yanoikuyae* JCM 7371^T (Figs. S1-S3). Across all methods, strains WW5^T and 11R-BB formed a well-supported clade, with strain HT1-2^T clustering as their closest relative, and the resulting clade positioned adjacent to *S. yanoikuyae* JCM 7371^T. The 16S rRNA sequences are available in NCBI (PX282427, PX282425, PX282426).

Genome sequencing, assembly, and annotation

High-quality hybrid assemblies were generated for strains WW5^T, 11R-BB, and HT1-2^T from Illumina and Oxford Nanopore sequencing data. The assemblies comprised 8, 10, and 7 contigs, respectively, each with a single long contig capturing the majority of the genome; all had an L50 of 1 and N50 values of 5.06 Mb (strains WW5^T and 11R-BB) and 4.67 Mb (strain HT1-2^T). CheckM analyses estimated high completeness (100%) and low contamination (0.3–0.4%) for all three assemblies (Table S2).

Strains WW5^T and 11R-BB displayed nearly identical genome characteristics, with genome sizes of 5.72 Mb, GC content of 64.0%, and 5,628 and 5,635 coding sequences, respectively. Strain HT1-2^T had a slightly smaller genome (5.49 Mb), similar GC content (64.2%), and 5404 coding sequences. In all three assemblies, 38–44% of predicted proteins were annotated as hypothetical, consistent with values reported

for environmental sphingomonads (Aylward et al. 2013). These values are consistent with the genomic range observed in related *Sphingobium* type strains, while the larger genome size and greater number of predicted coding sequences relative to *S. yanoikuyae* JCM 7371^T highlight the distinct genomic profile of the new strains. The genome assemblies for strains WW5^T, 11R-BB, and HT1-2^T have been deposited in GenBank under the accession numbers GCF_049472635.1, GCF_049472495.1, and GCF_049472475.1.

Whole-genome phylogeny and taxonomic placement

Whole-genome relatedness among strains WW5^T, 11R-BB, and HT1-2^T was evaluated using ANIb, dDDH, and phylogenomics analyses. According to both the ANIb and dDDH results, the three strains were more closely related to each other than to *S. yanoikuyae* JCM 7371^T (Table X1). Within the group, strains WW5^T and 11R-BB were nearly identical, sharing 99.99% ANIb and 99.8% dDDH (formula d_4), while strain HT1-2^T was more divergent but remained more similar to strains WW5^T and 11R-BB than to *S. yanoikuyae* JCM 7371^T.

Similarly, the phylogenomic analyses conducted with TYGS (GBDP distance tree) and GTDB-Tk (Bac120 maximum-likelihood tree) placed strains WW5^T, 11R-BB, and HT1-2^T as a clade distinct from *S. yanoikuyae* JCM 7371^T (Figs. 1, 2). In the GBDP tree, the strains WW5^T and 11R-BB grouped together with 65% bootstrap support, strain HT1-2^T formed a sister branch with 65% support, and the clade was separated from *S. yanoikuyae* JCM 7371^T with 87% support. In the GTDB-Tk Bac120 maximum-likelihood tree, the three strains also formed a distinct lineage separated from *S. yanoikuyae* JCM 7371^T, with bootstrap support > 80%.

Taken together, the ANIb, dDDH, and phylogenomic analyses consistently resolved strains WW5^T, 11R-BB, and HT1-2^T as forming a lineage distinct from *S. yanoikuyae* JCM 7371^T. The congruence among these genome-scale methods supports recognition of a new species within the genus *Sphingobium*. Within this lineage, strains WW5^T and 11R-BB are genomically indistinguishable, whereas strain HT1-2^T shows more divergence but remains part of the same species (Table 1).

Comparative genomics and functional analysis

Comparative Systems analysis in BV-BRC identified 20,694 protein gene families (PGfams) across 46 genomes, comprising 43 validly published *Sphingobium* type strains and study strains WW5^T, 11R-BB, and HT1-2^T. The genus-wide core comprised 903 PGfams present across all 46 genomes (File S4). This core gene family set represents 19.7% of the average *Sphingobium* genome (4576 coding sequences), reflecting the high genomic diversity typical of sphingomonads (Aylward et al. 2013; Wang et al. 2018; Verma et al. 2020). Eighty-one PGfams conserved across strains WW5^T, 11R-BB, and HT1-2^T were absent from the 43 *Sphingobium* type strain dataset (Table S4); 71 comprised hypothetical protein families, while 10 encoded putative functions including mobile element-associated proteins and regulators. Multiple sequence alignment showed that 56 of the hypothetical families were identical at the amino acid level across the three genomes, suggesting functional constraint (data not shown). Notably, the lactate-responsive transcriptional regulator LldR (GntR family; PGfam_03726865) was present in the three strains but absent from the *Sphingobium* type strain dataset. In plant-associated niches, where lactate fluctuates in root exudates and accumulates under hypoxia (Jones 1998), this inducible regulation could enable rapid metabolic response to substrate availability, facilitating carbon utilization and metabolic integration with the host. Complementary PLfam analysis through BV-BRC Comparative Systems, which provides higher-resolution clustering than PGfam, identified divergent arylsulfatase (EC 3.1.6.1) families: seven conserved across all three strains and one unique to strain HT1-2^T (Table S5). Arylsulfatases hydrolyze sulfate esters, releasing sulfate and phenolic moieties (Slezack-Deschaumes et al. 2012; Toesch et al. 2014), and divergent forms may expand substrate use in plant-or soil-associated microorganisms (Knauff et al. 2003).

Plant endophytic adaptations

Comparative analysis of protein family (PGfam) complements revealed shared genomic features in strains WW5^T, 11R-BB, and HT1-2^T that were absent or divergent in the type strain *S. yanoikuyae* JCM 7371^T. Four major systems distinguished the three strains

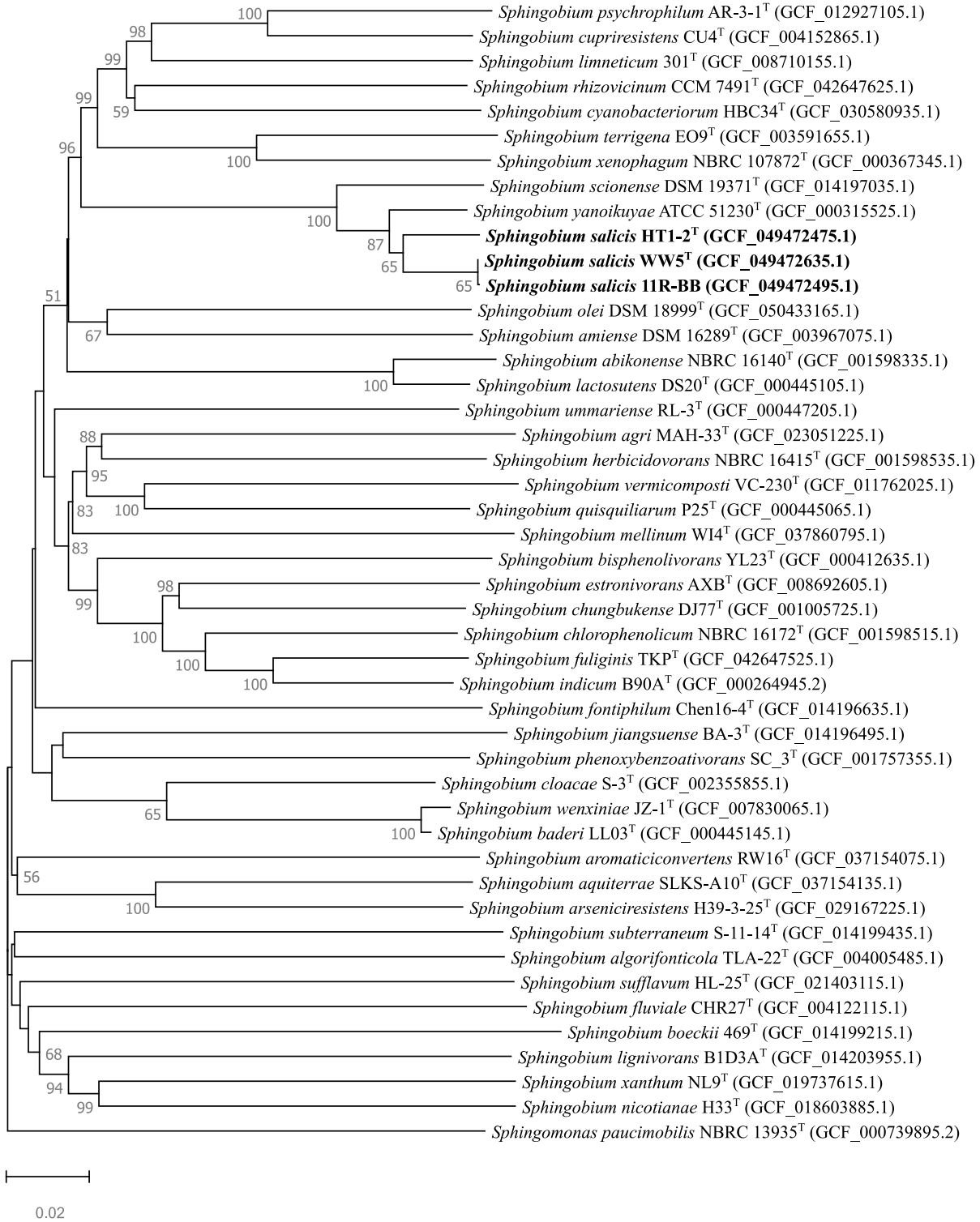

Fig. 1 Genome-based GBDP tree showing the position of strains WW5^T, 11R-BB, and HT1-2^T among type strains of *Sphingobium*. The tree was inferred in TYGS with FASTME 2.1.6.1 from GBDP distances of whole-genome sequences. Branch lengths represent GBDP distance formula d_5, and pseudo-bootstrap support values > 60% from 100 replicates are shown at nodes. The tree was rooted with *Sphingomonas paucimobilis* NBRC 13935^T (GCF_000739895.2)

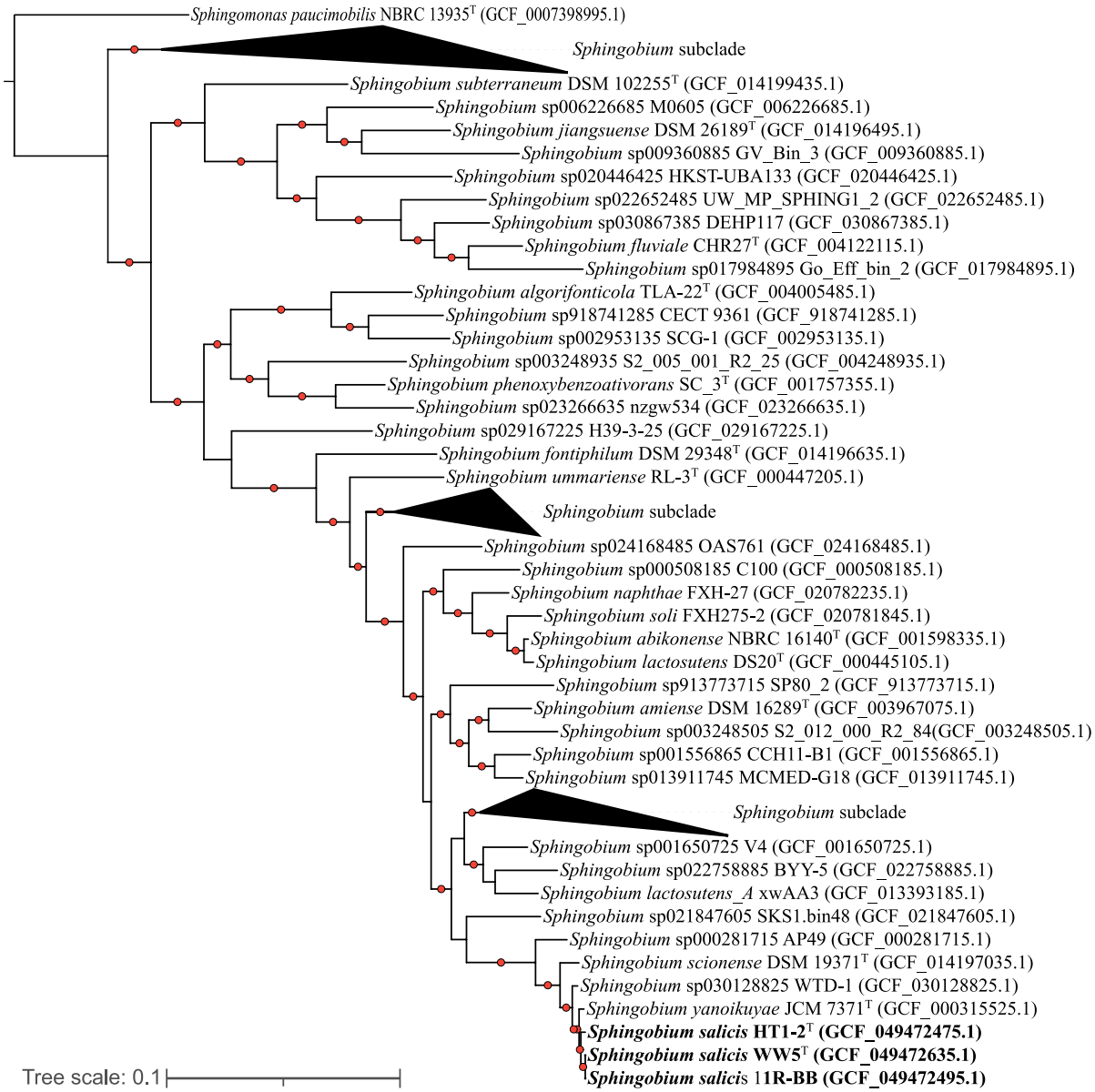

Fig. 2 Genome-wide phylogeny showing the position of strains WW5^T, 11R-BB, and HT1-2^T within the genus *Sphingobium*. The tree was inferred with GTDB-Tk (R220) from the Bac120 marker set using maximum-likelihood in IQ-TREE with 1,000 bootstrap replicates. The tree was rooted with *Sphingomonas paucimobilis* NBRC 13935^T (GCF_000739895.2)

from the type strain (Table S6): a complete beta-ketoadipate pathway for aromatic compound degradation, a lactate utilization system, a quorum sensing architecture lacking autoinducer synthesis, and the absence of soil-associated polysaccharide degradation systems.

The beta-ketoadipate pathway is a central aerobic catabolic system in some bacteria and fungi that links aromatic compound degradation to the tricarboxylic acid (TCA) cycle (Harwood and Parales 1996). Strains WW5^T, 11R-BB, and HT1-2^T possessed the complete pathway, oxidizing benzoate to catechol via benzoate 1,2-dioxygenase (EC 1.14.12.10) and

Table 1 Whole-genome relatedness 1a) among the strains WW5^T, 11R-BB, and HT1-2^T and 1b) between each strain and *S. yanoikuyae* JCM 7371^T based on average nucleotide identity BLAST-based (ANIb), alignment fraction (AF), and digital DNA–DNA hybridization (dDDH, formula d4) values

1a)

Query	Subject	ANIb (%)	AF	d4	CI d4	ΔGC%
Pairwise whole-genome comparisons among strains WW5^T, 11R-BB, and HT1-2^T showing ANIb (%), AF, d_4 values with 95% confidence intervals (CI d_4), and G + C content differences (ΔGC %)						
WW5^T	11R-BB	99.98	0.995	99.8	99.6–99.9	0.08
HT1-2^T	WW5^T	95.25	0.793	70.1	67.1–72.9	0.21
HT1-2^T	11R-BB	95.31	0.795	70.0	67.0–72.9	0.13

1b)

Query	ANIb (%)	AF	d4	CI d4	ΔGC%
Whole-genome comparisons between strains WW5^T, 11R-BB, and HT1-2^T and S. yanoikuyae JCM 7371^T showing ANIb (%), AF, d_4 values with 95% confidence intervals (CI d_4), and G + C content differences (ΔGC %)					
WW5^T	94.07	0.747	66.2	63.2–69.0	0.39
11R-BB	94.03	0.747	66.2	63.3–69.1	0.31
HT1-2^T	95.00	0.759	68.1	65.1–71.0	0.18

dihydrodiol dehydrogenase (EC 1.3.1.25), and further metabolizing catechol through catechol 1,2-dioxygenase (EC 1.13.11.1), muconate cycloisomerase (EC 5.5.1.1), muconolactone isomerase (EC 5.3.3.4), and dienelactone hydrolase (EC 3.1.1.45) to yield β-ketoadipate, which enters central metabolism as succinyl-CoA and acetyl-CoA. In contrast, *S. yanoikuyae* JCM 7371^T retained only a remnant 2-keto-4-pentenoate hydratase (EC 4.2.1.80).

Genes encoding L-lactate dehydrogenase (EC 1.1.2.3), lactate permease, and the regulator LldR were identified in strains WW5^T, 11R-BB, and HT1-2^T, enabling utilization of lactate generated under root hypoxia (Raman et al. 2022), but were not detected in *S. yanoikuyae* JCM 7371^T. The ability to metabolize both aromatic compounds and lactate broadens the metabolic versatility of these strains. By consuming lactate exuded under oxygen limitation, they may help moderate local lactate accumulation, a process that could reduce lactate-induced cellular stress in roots. Such activity has the potential to support membrane integrity and root function under flooding or hypoxic conditions, although plant-level benefits remain to be demonstrated (Raman et al. 2022).

Differences were also identified in signaling and polymer metabolism. Strains WW5^T, 11R-BB, and HT1-2^T lacked LuxI synthases but retained LuxR-type receptors, suggesting reliance on detecting external quorum signals, whereas *S. yanoikuyae* JCM 7371^T encoded both LuxI and LuxR, supporting autonomous quorum control. The type strain also carried genes for soil-associated polysaccharide and sulfolipid metabolism, including xylanase (EC 3.2.1.8), xylosidase (EC 3.2.1.37), glucarate dehydratase (EC 4.2.1.40), and UDP-sulfoquinovose synthase (EC 3.13.1.1), functions associated with hemicellulose and sulfolipid metabolism (Juturu and Wu 2014; Bhardwaj et al. 2019). Their absence in strains WW5^T, 11R-BB, and HT1-2^T suggests a shift away from soil polymer degradation toward reliance on soluble substrates.

Biosynthetic gene cluster (BGC) analysis with antiSMASH v8.0 identified 8 BGCs in strains WW5^T and 11R-BB, 6 in strain HT1-2^T, and 11 in *S. yanoikuyae* JCM 7371^T (Files S5-S8). Ectoine biosynthesis was conserved in all strains, where it may contribute osmoprotection by stabilizing proteins and membranes (Richter et al. 2019). In plant-associated niches, ectoine likely supports bacterial survival in fluctuating osmotic conditions, which may indirectly benefit host plants (Empadinhas and da Costa 2008). *S. yanoikuyae* JCM 7371^T displayed greater BGC diversity with unique Hserlactone (HSL), RiPP-like clusters, and lassopeptides. Conversely, an RRE-containing cluster, associated with peptide-modifying enzyme recognition (Kloosterman et al. 2020), was found only in strains WW5^T and 11R-BB.

Plant growth promotion genomic features

Genomic analysis of strains WW5^T, 11R-BB, and HT1-2^T using KEGG KofamKOALA (ver. 2025-09-01, KEGG release 115.0) revealed shared plant growth–promoting traits centered on osmolyte biosynthesis, phosphate mobilization, and antioxidant protection (Table S6). All three strains encoded the trehalose pathway via α,α-trehalose-phosphate synthase (EC 2.4.1.15; K00697) and trehalose-6-phosphate phosphatase (EC 3.1.3.12; K01087) within starch and sucrose metabolism. Trehalose acts as a compatible solute and has been implicated in microbial and plant tolerance to drought, salinity, and oxidative stress (Iturriaga et al. 2009).

A broad phosphatase repertoire was identified in the three genomes, including alkaline phosphatase (EC 3.1.3.1; K01077), acid phosphatase (EC 3.1.3.2; K01078), phosphoserine phosphatase (EC 3.1.3.3; K01079), 5′-nucleotidase (EC 3.1.3.5; K01081), fructose-1,6-bisphosphatase (EC 3.1.3.11; K03841), phosphoglycolate phosphatase (EC 3.1.3.18; K01091), inositol-1-monophosphatase (EC 3.1.3.25; K01092), and pyridoxal-5′-phosphate phosphatase (EC 3.1.3.74; K06215), distributed across diverse metabolic modules. This enzymatic diversity is consistent with the ability to hydrolyze organic phosphorus esters and liberate inorganic phosphate for potential host uptake (Varga et al. 2020).

All three strains carried catalase KatE (EC 1.11.1.6; K03781), supporting detoxification of hydrogen peroxide and reduction of oxidative stress during host colonization. Antioxidant protection of this type may help buffer both microbes and roots from reactive oxygen species generated under stress conditions (Imlay 2013).

A strain-specific distinction was observed in strain HT1-2^T, which uniquely encoded levanase (EC 3.2.1.65; K07046), a beta-glucan hydrolase in starch and sucrose metabolism. The ability to utilize levan, a beta-2, 6-fructan not broadly metabolized among root-associated microbes (Hernandez et al. 1995), differentiates strain HT1-2^T from strains WW5^T and 11R-BB and may extend its substrate range in the rhizosphere or endosphere.

Collectively, these genomic features indicate that the three endophytic strains possess complementary mechanisms for nutrient acquisition and stress mitigation. Such traits support their potential roles as bioinoculants in sustainable agriculture. The in silico assessment predicted all strains to be non-pathogenic, with PathogenFinder v1.1 classifying them as non-pathogens and VirulenceFinder v2.0 detecting no virulence factors (Table S7). These tools provide

Table 2 Physiological and biochemical characteristics of strains WW5^T, 11R-BB, HT1-2^T, and *S. yanoikuyae* JCM 7371^T. Growth parameters (temperature, pH, NaCl tolerance) and differential carbon source utilization patterns from 18 substrates. Symbols: "+" = growth; "–" = no growth

2a)

Characteristic	WW5^T	11R-BB	HT1-2^T	JCM 7371^T
Growth parameters of the strains WW5^T, 11R-BB, HT1-2^T and S. yanoikuyae JCM 7371^T (temperature, pH, NaCl tolerance)				
Temperature Range (°C)	25–35	25–35	25–35	25–30
pH range	5–8	5–8	5–8	5–8
NaCl tolerance (% w/v)	1–2%	1–2%	1–3%	1–2%
Aerotolerance	Obligate aerobe	Obligate aerobe	Obligate aerobe	Obligate aerobe

2b)

Carbon Source Utilization	WW5^T	11R-BB	HT1-2^T	JCM 7371^T
Differential carbon source utilization of the strains WW5^T, 11R-BB, HT1-2^T and S. yanoikuyae JCM 7371^T				
D-Galactose	+	+	+	–
D-Gluconic Acid	+	+	+	–
D,L-Malic Acid	+	+	+	–
D-Fructose	+	+	+	–
L-Malic Acid	+	+	+	–

computational predictions, which are indicative but not a substitute for experimental confirmation.

Morphological, physiological, and biochemical analyses

Cells are Gram-negative, rod-shaped, monotrichous, catalase positive, and strictly aerobic. Growth occurs on NLCCM and MGL media, with colonies visible after 2–3 days of incubation. The optimum growth range is 30–32.5 °C, with growth observed between 25 and 35 °C and no growth at 37 °C. The three strains tolerate pH 5–8 (optimum pH 7) and NaCl concentrations up to 2% (strains WW5^T and 11R-BB) or 3% (strain HT1-2^T). The strains displayed identical carbon utilization profiles, metabolizing all tested substrates except D-ribose and L-lyxose (Table 2). In contrast, *S. yanoikuyae* JCM 7371^T did not grow on D-galactose, D-gluconic acid, D,L-malate, L-malate, or D-fructose, compounds common in plant tissues and root exudates (Badri and Vivanco 2009). The strains encoded ABC-and PTS-mediated uptake systems and complete central metabolic pathways, consistent with capacity for carbon acquisition under nutrient limitation (Deutscher et al. 2006; Badri and Vivanco 2009). Enzymes such as xylanases and pectin acetylesterases indicate potential for modification of plant cell wall components, consistent with traits reported in beneficial endophytes (Compant et al. 2005; Piromyou et al. 2015). The concordance of genotype and phenotype suggests these functions are conserved and important for endophytic colonization.

Chemotaxonomic analysis

The cellular fatty acid profiles of strains WW5^T, 11R-BB, and HT1-2^T were dominated by vaccenic acid ($C_{18:1}$ ω7c, 57–60% of total FAME) with palmitoleic acid ($C_{16:1}$ ω7c, 15–16%) as the secondary component (Table S8). Compared to *S. yanoikuyae* JCM 7371^T, the three strains contained higher proportions of palmitic acid ($C_{16:0}$, + 6–7%) and lower levels of $C_{16:1}$ ω7c, $C_{16:1}$ ω5c, and $C_{14:0}$ 2–OH. These consistent differences represent chemotaxonomic differences distinguishing the strains from *S. yanoikuyae* JCM 7371^T, while the *S. yanoikuyae* JCM 7371^T, whose profile corresponded to the published reference range (Yabuuchi et al. 1990; Khan et al. 1996). The predominant respiratory quinone in strain WW5^T was ubiquinone with ten isoprene units (Q-10, 89%), accompanied by Q-9 (9.7%). Strains 11R-BB and HT1-2^T showed similar profiles, with Q-10 at 88.3% and 87.8% and Q-9 at 10.6% and 11.3%, respectively. The polar lipid patterns of all three strains were similar, consisting of sphingoglycolipids, phosphatidyldimethylethanolamine, phosphatidylethanolamine, diphosphatidylglycerol,

phosphatidylmonomethylethanolamine, lysophosphatidylethanolamine, and phosphatidylglycerol (Figure S4).

Taxonomic conclusions

Phylogenomic, genomic, and phenotypic analyses demonstrate that strains WW5^T, 11R-BB, and HT1-2^T represent a novel species within the genus *Sphingobium*. Comparative genome analyses confirmed their separation from *S. yanoikuyae* and other described species of the genus, and phylogenomic reconstruction consistently resolved them as a distinct lineage. The strains share the characteristic chemotaxonomic and physiological features of *Sphingobium*, with minor genomic and ecological divergence among them. These results support the proposal of *Sphingobium salicis* sp. nov., comprising *S. salicis* subsp. *salicis* subsp. nov. (strains WW5^T and 11R-BB) and *S. salicis* subsp. *plucheae* subsp. nov. (strain HT1-2^T).

Description of *Sphingobium salicis* sp. nov

Sphingobium salicis (sa'li.cis. L. fem. n. *salix*, willow; L. gen. fem. n. *salicis*, of the willow, referring to the primary host of isolation).

Cells are Gram-negative, rod-shaped, monotrichous, catalase positive, and strictly aerobic. Colonies form within 2–3 days on NLCCM and MGL agar. Growth occurs between 25 and 35 °C with an (optimum at 30–32.5 °C), but not at 37 °C. Growth occurs at pH 5–8 (optimum 7). Strains tolerate up to 2% NaCl (strains WW5^T and 11R-BB) or up to 3% NaCl (strain HT1-2^T).

Growth occurs on L-arabinose, D-galactose, D-trehalose, D-mannose, D-gluconic acid, D-xylose, D, L-malate, L-rhamnose, D-fructose, α-D-glucose, maltose, D-melibiose, α-D-lactose, sucrose, β-methyl-D-glucoside, D-cellobiose, and L-malate, but not D-ribose or L-lyxose.

The major fatty acids are $C_{18:1}$ ω7c (57–60%) and $C_{16:1}$ ω7c (15–16%). The predominant respiratory quinone is ubiquinone Q-10 with minor Q-9. Polar lipids include sphingoglycolipids, phosphatidyldimethylethanolamine, diphosphatidylglycerol, lysophosphatidylethanolamine, phosphatidylethanolamine, phosphatidylglycerol, and phosphatidylmonomethylethanolamine. The DNA G + C content is 63.9–64.3%. Genome sizes range from 5.39–5.72 Mb.

Strains $WW5^T$ and 11R-BB were isolated as endophytes from stems of *Salix sitchensis* and roots of *Populus trichocarpa* in Washington, USA, and strain HT1-2^T from shoots of *Pluchea carolinensis* in Hawai'i, USA.

Description of *Sphingobium salicis subsp. salicis subsp. nov*

Sphingobium salicis subsp. *salicis* (sa'li.cis. L. fem. n. *salix*, willow; L. gen. fem. n. *salicis*, of the willow, referring to the host of isolation).

Genomic comparison of the type strain $WW5^T$ with *Sphingobium salicis* subsp. *plucheae* HT1-2^T shows 95.1% average nucleotide identity (ANIb) and 70.1% digital DNA–DNA hybridization (dDDH), values consistent with subspecies-level divergence within *S. salicis*.

The type strain is $WW5^T$ (= DSM 120182^T = NCCB 101075^T = NRRL B-68079^T), isolated from stems of *Salix sitchensis* collected along the Snoqualmie River, Washington, USA. The GenBank accession numbers for the 16S rRNA gene and whole-genome sequences of strain $WW5^T$ are PX282427 and JAYRCS000000000, respectively.

Description of *Sphingobium salicis subsp. pluchea subsp. nov*

Sphingobium salicis subsp. plucheae (plu'che.a. N.L. fem. n. *plucheae*, referring to the host genus *Pluchea from which the strain was isolated).*).

The type strain HT1-2^T differs from *S. salicis* subsp. *salicis* $WW5^T$ by 95.1% ANIb and 70.1% dDDH, supporting recognition as a distinct subspecies within *S. salicis*.

The type strain is HT1-2^T (= DSM 35353^T = NRRL B-68485^T), isolated from shoots of *Pluchea carolinensis* collected on volcanic substrates near Kona, Hawai'i, USA. The GenBank accession numbers for the 16S rRNA gene and whole-genome sequences of strain HT1-2 ^{T}are PX282426 and JBCEBS000000000, respectively.

Acknowledgements We appreciate the technical assistance provided by Phuong Thuy Phan, Intrinsyx Bio, and Jean-Yves Mugnier, Syngenta

Author Contributions JF conceived the experiments with help from AF. AF and RT conducted the genomic and bioinformatics analyses. JF, ML, and DB conducted chemotaxonomic and phenotypic analyses. SLD provided the strains. All authors contributed to writing the manuscript.

Funding United States Department of Energy Office of Science, Office of Biological and Environmental Research (BER), grant no.DE-SC0021137.

Data availability The genome assemblies for strains WW5T, 11R-BB, and HT1-2 T have been deposited in GenBank under the accession numbers GCF_049472635.1, GCF_049472495.1, and GCF_049472475.1, respectively, and their corresponding 16S rRNA gene sequences under PX282427, PX282425, and PX282426. Cultures of strain WW5\u207f are available in the Agricultural Research Culture Collection (NRRL B-68079 T), Netherlands Culture Collection of Bacteria (NCCB 101075 T), and German Collection of Microorganisms and Cell Cultures (DSM 120182 T). Strain 11R-BB is available as DSM 35354 and NRRL B-68483, and strain HT1-2 T as DSM 35353 T and NRRL B-68485 T.

Repositories German Collection of Microorganisms and Cell Cultures (DSMZ) codes WW5T (DSM 120182 T), 11R-BB (DSM 35354), HT1-2 T (DSM 35353 T).

WI-KNAW culture collection code; WW5T (NCCB 101075 T). Agricultural Research Culture Collection (NRRL) codes; WW5T (B-68079 T), 11R-BB (B-68483), HT1-2 T (B-68485 T).

Declarations None.

Conflict of interest Intrinsyx Bio is developing agricultural products based on the microbial strains described in this study.

Ethical approval Not required.

References

Ali S, Isaacson J, Kroner Y, Saldias S, Kandasamy S, Lazarovits G (2018) Corn sap bacterial endophytes and their potential in plant growth-promotion. Environ Sustain 1:341–355. https://doi.org/10.1007/s42398-018-00030-4

Aramaki T, Blanc-Mathieu R, Endo H, Ohkubo K, Kanehisa M, Goto S, Ogata H (2020) KofamKOALA: KEGG ortholog assignment based on profile HMM and adaptive

score threshold. Bioinformatics 36:2251–2252. https://doi.org/10.1093/bioinformatics/btz859

Arunthavasu R, Thangavel K, Uthandi S (2019) Impact of drought-tolerant rice apoplastic fluid endophyte (*Sphingobium yanoikuyae* MH394206) on the morphological and physiological characteristics of rice (CO51) grown in moisture deficit condition. Madras Agric J 106:1–5. https://doi.org/10.29321/MAJ.2019.000249

Aufrecht J, Velickovic D, Tournay R, Couvillion S, Balasubramanian V, Winkler T, Herrera D, Stanley R, Doty SL, Ahkami AH (2025) Endophytes induce systemic spatial reprogramming of metabolism in poplar roots under drought. bioRxiv. https://doi.org/10.1101/2025.06.02.657501

Aylward FO, McDonald BR, Adams SM, Valenzuela A, Schmidt RA, Goodwin LA, Woyke T, Currie CR, Suen G, Poulsen M (2013) Comparison of 26 *Sphingomonad* genomes reveals diverse environmental adaptations and biodegradative capabilities. Appl Environ Microbiol 79:3724–3733. https://doi.org/10.1128/AEM.00518-13

Badri DV, Vivanco JM (2009) Regulation and function of root exudates. Plant Cell Environ 32:666–681. https://doi.org/10.1111/j.1365-3040.2009.01926.x

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477. https://doi.org/10.1089/cmb.2012.0021

Bhardwaj N, Kumar B, Verma P (2019) A detailed overview of xylanases: an emerging biomolecule for current and future prospective. Bioresour Bioprocess 6:40. https://doi.org/10.1186/s40643-019-0276-2

Bligh EG, Dyer WJ (1959) A rapid method of total lipid extraction and purification. Can J Biochem Physiol 17:911–917. https://doi.org/10.1139/o59-099

Boss BL, Wanees AE, Zaslow SJ, Normile TG, Izquierdo JA (2022) Comparative genomics of the plant-growth-promoting bacterium *Sphingobium* sp. strain AEW4 isolated from the rhizosphere of the beachgrass *Ammophila breviligulata*. BMC Genomics 23:508. https://doi.org/10.1186/s12864-022-08738-8

Brettin T, Davis JJ, Disz T, Edwards RA, Gerdes S, Olsen GJ, Olson R, Overbeek R, Parrello B, Pusch GD, Shukla M, Thomason JA, Stevens R, Vonstein V, Wattam AR, Xia F (2015) RASTtk: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. Sci Rep 5:8365. https://doi.org/10.1038/srep08365

Cangelosi GA, Abest E, Martinetti G, Nester EW (1991) Genetic analysis of *Agrobacterium*. Methods Enzymol 204:384–397. https://doi.org/10.1016/0076-6879(91)04019-8

Chalita M, Kim YO, Park S, Oh H-S, Cho JH, Moon J, Baek N, Moon C, Lee K, Yang J, Nam GG, Jung Y, Na S-I, Bailey MJ, Chun J (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

Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH (2022) GTDB-Tk v2: memory friendly classification with the genome taxonomy database. Bioinformatics 38:5315–5316. https://doi.org/10.1093/bioinformatics/btac672

Compant S, Reiter B, Sessitsch A, Nowak J, Clément C, Ait Barka E (2005) Endophytic colonization of *Vitis vinifera* L. by plant growth-promoting bacterium *Burkholderia* sp. strain PsJN. Appl Environ Microbiol 71:1685–1693. https://doi.org/10.1128/AEM.71.4.1685-1693.2005

Cosentino S, Larsen MV, Aarestrup FM, Lund O (2013) PathogenFinder: distinguishing friend from foe using bacterial whole genome sequence data. PLoS ONE 8:e77302. https://doi.org/10.1371/journal.pone.0077302

Davis JJ, Gerdes S, Olsen GJ, Olson R, Pusch GD, Shukla M, Vonstein V, Wattam AR, Yoo H (2016) PATtyFams: protein families for the microbial genomes in the PATRIC database. Front Microbiol 7:118. https://doi.org/10.3389/fmicb.2016.00118

Deutscher J, Francke C, Postma PW (2006) How phosphotransferase system-related protein phosphorylation regulates carbohydrate metabolism in bacteria. Microbiol Mol Biol Rev 70:939–1031. https://doi.org/10.1128/MMBR.00024-06

Doty SL, Dosher MR, Singleton GL, Moore AL, Van Aken B, Stettler RF, Strand SE, Gordon MP (2005) Identification of an endophytic *Rhizobium* in stems of *Populus*. Symbiosis 39:27–35

Doty SL, Oakley B, Xin G, Kang JW, Singleton G, Khan Z, Vajzovic A, Staley JT (2009) Diazotrophic endophytes of native black cottonwood and willow. Symbiosis 47:23–33. https://doi.org/10.1007/BF03179967

Doty SL, Joubert PM, Firrincieli A, Sher AW, Tournay R, Kill C, Parikh SS, Okubara P (2023) Potential biocontrol activities of *Populus* endophytes against several plant pathogens using different inhibitory mechanisms. Pathogens 12:13. https://doi.org/10.3390/pathogens12010013

Empadinhas N, da Costa MS (2008) Osmoadaptation mechanisms in prokaryotes: distribution of compatible solutes. Int Microbiol 11:151–161. https://doi.org/10.2436/20.1501.01.55

Ercole TG, Bonotto DR, Hungria M, Kava VM, Galli LV (2025) The role of endophytic bacteria in enhancing plant growth and health for sustainable agriculture. Antonie Van Leeuwenhoek 118:88. https://doi.org/10.1007/s10482-025-02100-0

Frank JA, Reich CI, Sharma S, Weisbaum JS, Wilson BA, Olsen GJ (2008) Critical evaluation of two primers commonly used for amplification of bacterial 16S rRNA genes. Appl Environ Microbiol 74:2461–2470. https://doi.org/10.1128/AEM.02272-07

Gibson DT, Mahadevan V, Jerina DM, Yagi H, Yeh HJC (1975) Oxidation of the carcinogens benzo[a]pyrene and benzo[a]anthracene to dihydrodiols by a bacterium. Science 189:295–297. https://doi.org/10.1126/science.1098377

Glaeser SP, Kämpfer P (2015) Multilocus sequence analysis (MLSA) in prokaryotic taxonomy. Syst Appl Microbiol 38:237–245. https://doi.org/10.1016/j.syapm.2015.03.007

Hardoim PR, van Overbeek LS, Berg G, Pirttilä AM, Compant S, Campisano A, Döring M, Sessitsch A (2015) The

hidden world within plants: ecological and evolutionary considerations for defining functioning of microbial endophytes. Microbiol Mol Biol Rev 79:293–320. https://doi.org/10.1128/MMBR.00050-14

Harwood CS, Parales RE (1996) The beta-ketoadipate pathway and the biology of self-identity. Annu Rev Microbiol 50:553–590. https://doi.org/10.1146/annurev.micro.50.1.553

Heimbrook ME, Wang WL, Campbell G (1989) Staining bacterial flagella easily. J Clin Microbiol 27:2612–2615. https://doi.org/10.1128/jcm.27.11.2612-2615.1989

Hernandez L, Arrieta J, Menendez C, Vazquez R, Coego A, Suarez V, Selman G, Petit-Glatron MF, Chambert R (1995) Isolation and enzymic properties of levansucrase secreted by *Acetobacter diazotrophicus* SRT4, a bacterium associated with sugar cane. Biochem J 309(Pt 1):113–118. https://doi.org/10.1042/bj3090113

Imlay JA (2013) The molecular mechanisms and physiological consequences of oxidative stress: lessons from a model bacterium. Nat Rev Microbiol 11:443–454. https://doi.org/10.1038/nrmicro3032

Iturriaga G, Suárez R, Nova-Franco B (2009) Trehalose metabolism: from osmoprotection to signaling. Int J Mol Sci 10:3793–3810. https://doi.org/10.3390/ijms10093793

Joensen KG, Scheutz F, Lund O, Hasman H, Kaas RS, Nielsen EM, Aarestrup FM (2014) Real-time whole-genome sequencing for routine typing, surveillance, and outbreak detection of verotoxigenic *Escherichia coli*. J Clin Microbiol 52:1501–1510. https://doi.org/10.1128/JCM.03617-13

Jones DL (1998) Organic acids in the rhizosphere – a critical review. Plant Soil 205:25–44. https://doi.org/10.1023/A:1004356007312

Jou Y-T, Tarigan EJ, Prayogo C, Kobua CK, Weng Y-T, Wang Y-M (2022) Effects of *Sphingobium yanoikuyae* SJTF8 on rice (*Oryza sativa*) seed germination and root development. Agriculture 12:1890. https://doi.org/10.3390/agriculture12111890

Juturu V, Wu JC (2014) Microbial exo-xylanases: a mini review. Appl Biochem Biotechnol 174:81–92. https://doi.org/10.1007/s12010-014-1042-8

Khan AA, Wang RF, Cao WW, Franklin W, Cerniglia CE (1996) Reclassification of a polycyclic aromatic hydrocarbon-metabolizing bacterium, *Beijerinckia* sp. strain B1, as *Sphingomonas yanoikuyae* by fatty acid analysis, protein pattern analysis, DNA-DNA hybridization, and 16S ribosomal DNA sequencing. Int J Syst Evol Microbiol 46:466–469. https://doi.org/10.1099/00207713-46-2-466

Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 16:111–120. https://doi.org/10.1007/BF01731581

Kloosterman AM, Shelton KE, van Wezel GP, Medema MH, Mitchell DA (2020) RRE-finder: a genome-mining tool for class-independent RiPP discovery. mSystems 5:e00267-20. https://doi.org/10.1128/mSystems.00267-20

Knauff U, Schulz M, Scherer HW (2003) Arylsulfatase activity in the rhizosphere and roots of different crop species. Eur J Agron 19:215–223. https://doi.org/10.1016/S1161-0301(02)00035-7

Kumar S, Stecher G, Suleski M, Sanderford M, Sharma S, Tamura K (2024) MEGA12: molecular evolutionary genetic analysis version 12 for adaptive and green computing. Mol Biol Evol 41:msae263. https://doi.org/10.1093/molbev/msae263

Kunitsky C, Osterhout G, Sasser M (2006) Identification of microorganisms using fatty acid methyl ester (FAME) analysis and the MIDI Sherlock microbial identification system. MIDI Tech Note 3:1–18

Ma J, Peng Q, Chen S, Liu Z, Zhang W, Zhang C, Du X, Sun S, Peng W, Lei Z, Zhang L, Su P, Zhang D, Liu Y (2025) Microbiome migration from soil to leaves in maize and rice. Microorganisms 13:947. https://doi.org/10.3390/microorganisms13040947

Matsen FA, Kodner RB, Armbrust EV (2010) pplacer: linear time maximum-likelihood and Bayesian phylogenetic placement of sequences onto a fixed reference tree. BMC Bioinformatics 11:538. https://doi.org/10.1186/1471-2105-11-538

Meier-Kolthoff JP, Göker M (2019) TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy. Nat Commun 10:2182. https://doi.org/10.1038/s41467-019-10210-3

Mitra M, Nguyen KM-A-K, Box TW, Gilpin JS, Hamby SR, Berry TL, Duckett EH (2020) Isolation and characterization of a novel *Sphingobium yanoikuyae* strain variant that uses biohazardous saturated hydrocarbons and aromatic compounds as sole carbon sources. F1000Res 9:767. https://doi.org/10.12688/f1000research.25284.1

Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ (2015) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol 32:268–274. https://doi.org/10.1093/molbev/msu300

Olson RD, Assaf R, Brettin T, Conrad N, Cucinell C, Davis JJ, Dempsey DM, Dickerman A, Dietrich EM, Kenyon RW, Kuscuoglu M, Lefkowitz EJ, Lu J, Machi D, Macken C, Mao C, Niewiadomska A, Nguyen M, Olsen GJ, Overbeek JC, Parrello B, Parrello V, Porter JS, Pusch GD, Shukla M, Singh I, Stewart L, Tan G, Thomas C, VanOeffelen M, Vonstein V, Wallace ZS, Warren AS, Wattam AR, Xia F, Yoo H, Zhang Y, Zmasek CM, Scheuermann RH, Stevens RL (2023) Introducing the bacterial and viral bioinformatics resource center (BV-BRC): a resource combining PATRIC, IRD and ViPR. Nucleic Acids Res 51:D678–D689. https://doi.org/10.1093/nar/gkac1003

Oren A (2024) On validly published names, correct names, and changes in the nomenclature of phyla and genera of prokaryotes: a guide for the perplexed. NPJ Biofilms Microbiomes 10:4. https://doi.org/10.1038/s41522-024-00494-9

Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW (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

Parks DH, Chuvochina M, Rinke C, Mussig AJ, Chaumeil P-A, Hugenholtz P (2022) GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. Nucleic Acids Res 50:D785–D794. https://doi.org/10.1093/nar/gkab776

Parte AC, Sardà Carbasse J, Meier-Kolthoff JP, Reimer LC, Göker M (2020) List of prokaryotic names with standing in nomenclature (LPSN) moves to the DSMZ. Int J Syst Evol Microbiol 70:5607–5612. https://doi.org/10.1099/ijsem.0.004332

Pinyakong O, Habe H, Omori T (2003) The unique aromatic catabolic genes in sphingomonads degrading polycyclic aromatic hydrocarbons (PAHs). J Gen Appl Microbiol 49:1–19. https://doi.org/10.2323/jgam.49.1

Piromyou P, Greetatorn T, Teamtisong K, Okubo T, Shinoda R, Nuntakij A, Tittabutr P, Boonkerd N, Minamisawa K, Teaumroong N (2015) Preferential association of endophytic *Bradyrhizobia* with different rice cultivars and its implications for rice endophyte evolution. Appl Environ Microbiol 81:3049–3061. https://doi.org/10.1128/AEM.04253-14

Raman J, Kim J-S, Choi KR, Eun H, Yang D, Ko Y-J, Kim S-J (2022) Application of lactic acid bacteria (LAB) in sustainable agriculture: advantages and limitations. Int J Mol Sci 23:7784. https://doi.org/10.3390/ijms23147784

Reinhold-Hurek B, Hurek T (2011) Living inside plants: bacterial endophytes. Curr Opin Plant Biol 14:435–443. https://doi.org/10.1016/j.pbi.2011.04.004

Rennie RJ (1981) A single medium for the isolation of acetylene-reducing (dinitrogen-fixing) bacteria from soils. Can J Microbiol 27:8–14. https://doi.org/10.1139/m81-002

Rho H, Van Epps V, Wegley N, Doty SL, Kim S-H (2018) Salicaceae endophytes modulate stomatal behavior and increase water use efficiency in rice. Front Plant Sci 9:188. https://doi.org/10.3389/fpls.2018.00188

Rho H, Van Epps V, Kim S-H, Doty SL (2020) Endophytes increased fruit quality with higher soluble sugar production in Honeycrisp apple (*Malus pumila*). Microorganisms 8:699. https://doi.org/10.3390/microorganisms8050699

Richter M, Rosselló-Móra R (2009) Shifting the genomic gold standard for the prokaryotic species definition. Proc Natl Acad Sci U S A 106:19126–19131. https://doi.org/10.1073/pnas.0906412106

Richter M, Rosselló-Móra R, Glöckner FO, 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

Richter AA, Mais C-N, Czech L, Geyer K, Hoeppner A, Smits SHJ, Erb TJ, Bange G, Bremer E (2019) Biosynthesis of the stress-protectant and chemical chaperone ectoine: biochemistry of the transaminase EctB. Front Microbiol 10:2811. https://doi.org/10.3389/fmicb.2019.02811

Rincón-Molina CI, Martínez-Romero E, Ruiz-Valdiviezo VM, Velázquez E, Ruiz-Lau N, Rogel-Hernández MA, Villalobos-Maldonado JJ, Rincón-Rosales R (2020) Plant growth-promoting potential of bacteria associated with pioneer plants from an active volcanic site of Chiapas (Mexico). Appl Soil Ecol 146:103390. https://doi.org/10.1016/j.apsoil.2019.103390

Seo J-S, Keum Y-S, Li QX (2009) Bacterial degradation of aromatic compounds. Int J Environ Res Public Health 6:278–309. https://doi.org/10.3390/ijerph6010278

Shaik SP, Thomas P (2019) In vitro activation of seed-transmitted cultivation-recalcitrant endophytic bacteria in tomato and host–endophyte mutualism. Microorganisms 7:132. https://doi.org/10.3390/microorganisms7050132

Sher AW, Aufrecht JA, Herrera D, Zimmerman AE, Kim Y-M, Munoz N, Trejo JB, Paurus VL, Cliff JB, Hu D, Chrisler WB, Tournay RJ, Gomez-Rivas E, Orr G, Ahkami AH, Doty SL (2024a) Dynamic nitrogen fixation in an aerobic endophyte of *Populus*. ISME J 18:wrad012. https://doi.org/10.1093/ismejo/wrad012

Sher AW, Tournay RJ, Gomez-Rivas E, Doty SL (2024b) Bacterial synergies amplify nitrogenase activity in diverse systems. ISME Commun 4:ycae158. https://doi.org/10.1093/ismeco/ycae158

Slezack-Deschaumes S, Piutti S, Vong P-C, Benizri E (2012) Dynamics of cultivable arylsulfatase-producing bacterial and fungal communities along the phenology of field-grown rape. Eur J Soil Biol 48:66–72. https://doi.org/10.1016/j.ejsobi.2011.07.014

Stolz A (2009) Molecular characteristics of xenobiotic-degrading sphingomonads. Appl Microbiol Biotechnol 81:793–811. https://doi.org/10.1007/s00253-008-1752-3

Takeuchi M, Hamana K, Hiraishi A (2001) Proposal of the genus *Sphingomonas* sensu stricto and three new genera, *Sphingobium*, *Novosphingobium* and *Sphingopyxis*, on the basis of phylogenetic and chemotaxonomic analyses. Int J Syst Evol Microbiol 51:1405–1417. https://doi.org/10.1099/00207713-51-4-1405

Táncsics A, Szoboszlay S, Szabó I, Farkas M, Kovács B, Kukolya J, Mayer Z, Kriszt B (2012) Quantification of subfamily I.2.C catechol 2,3-dioxygenase mRNA transcripts in groundwater samples of an oxygen-limited BTEX-contaminated site. Environ Sci Technol 46:232–240. https://doi.org/10.1021/es201842h

Tindall BJ, Rosselló-Móra R, Busse H-J, Ludwig W, Kämpfer P (2010) Notes on the characterization of prokaryote strains for taxonomic purposes. Int J Syst Evol Microbiol 60:249–266. https://doi.org/10.1099/ijs.0.016949-0

Toesch M, Schober M, Faber K (2014) Microbial alkyl- and aryl-sulfatases: mechanism, occurrence, screening and stereoselectivities. Appl Microbiol Biotechnol 98:1485–1496. https://doi.org/10.1007/s00253-013-5438-0

Varga T, Hixson KK, Ahkami AH, Sher AW, Barnes ME, Chu RK, Battu AK, Nicora CD, Winkler TE, Reno LR, Fakra SC, Antipova O, Parkinson DY, Hall JR, Doty SL (2020) Endophyte-promoted phosphorus solubilization in *Populus*. Front Plant Sci 11:567918. https://doi.org/10.3389/fpls.2020.567918

Varghese NJ, Mukherjee S, Ivanova N, Konstantinidis KT, Mavrommatis K, Kyrpides NC, Pati A (2015) Microbial species delineation using whole genome sequences. Nucleic Acids Res 43:6761–6771. https://doi.org/10.1093/nar/gkv657

Vaser R, Sović I, Nagarajan N, Šikić M (2017) Fast and accurate de novo genome assembly from long uncorrected reads. Genome Res 27:737–746. https://doi.org/10.1101/gr.214270.116

Verma H, Dhingra GG, Sharma M, Gupta V, Negi RK, Singh Y, Lal R (2020) Comparative genomics of *Sphingopyxis* spp. unravelled functional attributes. Genomics 112:1956–1969. https://doi.org/10.1016/j.ygeno.2019.11.008

Vieira S, Huber KJ, Neumann-Schaal M, Geppert A, Luckner M, Wanner G, Overmann J (2021) *Usitatibacter rugosus* gen. nov., sp. nov. and *Usitatibacter palustris* sp. nov., novel members of *Usitatibacteraceae* fam. nov. within the order *Nitrosomonadales* isolated from soil. Int J Syst Evol Microbiol 71:004631. https://doi.org/10.1099/ijsem.0.004631

Wang J, Wang C, Li J, Bai P, Li Q, Shen M, Li R, Li T, Zhao J (2018) Comparative genomics of degradative *Novosphingobium* strains with special reference to microcystin-degrading *Novosphingobium* sp. THN1. Front Microbiol 9:2238. https://doi.org/10.3389/fmicb.2018.02238

Wick RR, Judd LM, Gorrie CL, Holt KE (2017) Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput Biol 13:e1005595. https://doi.org/10.1371/journal.pcbi.1005595

Yabuuchi E, Yano I, Oyaizu H, Hashimoto Y, Ezaki T, Yamamoto H (1990) Proposals of *Sphingomonas paucimobilis* gen. nov. and comb. nov., *Sphingomonas parapaucimobilis* sp. nov., *Sphingomonas yanoikuyae* sp. nov., *Sphingomonas adhaesiva* sp. nov., *Sphingomonas capsulata* comb. nov., and two genospecies of the genus *Sphingomonas*. Microbiol Immunol 34:99–119. https://doi.org/10.1111/j.1348-0421.1990.tb00996.x

Yarza P, Yilmaz P, Pruesse E, Glöckner FO, Ludwig W, Schleifer K-H, Whitman WB, Euzéby J, Amann R, Rosselló-Móra R (2014) Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. Nat Rev Microbiol 12:635–645. https://doi.org/10.1038/nrmicro3330

Yoon S-H, Ha S-M, Kwon S, Lim J, Kim Y, Seo H, Chun J (2017) Introducing EzBioCloud: a taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies. Int J Syst Evol Microbiol 67:1613–1617. https://doi.org/10.1099/ijsem.0.001755

Zhu L, Xin K, Chen C, Li C, Si M, Zhao L, Shi X, Zhang L, Shen X (2015) *Sphingobium endophyticus* sp. nov., isolated from the root of *Hylomecon japonica*. Antonie Van Leeuwenhoek 107:1001–1008. https://doi.org/10.1007/s10482-015-0392-8