# *Sphingobium salicis* sp. nov., is a plant endophyte isolated from pioneer species of willow, poplar, and salvia growing on primary rock substrate

**Robert J. Tournay**

tournay@uw.ed

University of Washington

**John L. Freeman**

Intrinsyx Bio Inc

**Misha Levish**

Intrinsyx Bio Inc

**Douglas Baker**

Intrinsyx Bio Inc

**Andrea Firrincieli**

Tuscia University

**Sharon L. Doty**

University of Washington

**Research Article**

**Keywords:** Sphingobium, Bacterial Plant Endophyte, Willow Phyllosphere, Poplar Rhizosphere, Salvia Phyllosphere Taxonomy, Metabolism

**Posted Date:** June 27th, 2025

**DOI:** https://doi.org/10.21203/rs.3.rs-6933765/v1



**Additional Declarations:** Competing interest reported. Intrinsyx Bio is developing agricultural products based on the microbial strains described in this study.

**Version of Record:** A version of this preprint was published at Antonie van Leeuwenhoek on December 19th, 2025. See the published version at https://doi.org/10.1007/s10482-025-02214-5.

# Abstract

Three Gram-negative *Sphingobium* endophyte strains, designated WW5$^T$, 11R-BB, and HT1-2, were isolated from stems of *Salix sitchensis*, roots of *Populus trichocarpa*, and shoots of *Pluchea carolinensis*, respectively, growing on nutrient limited rock substrate undergoing primary succession. Phylogenetic analysis based on 16S rRNA gene sequences placed all three strains within the genus *Sphingobium*, with highest sequence identity to *S. yanoikuyae* JCM 7371$^T$. Digital DNA-DNA hybridization analysis revealed values of 66.2-68.2% to *S. yanoikuyae* JCM 7371$^T$, falling below the 70% species delineation threshold, while 99.9% identity between WW5$^T$ and 11R-BB and 70.1% between WW5$^T$ and HT1-2 confirmed their assignment to a single species. Average WGS nucleotide identity values to *S. yanoikuyae* JCM 7371$^T$ were 95.68% (WW5$^T$), 95.56% (11R-BB), and 96.65% (HT1-2), while phylogenomic WGS analysis using GTDB-Tk classified the strains as distinct from the *S. yanoikuyae* lineage, placing them under the placeholder *Sphingobium yanoikuyae_A*. Hybrid Illumina-Nanopore genome sequencing generated high-quality assemblies of 5.39-5.72 Mb with approximately 64% GC content. Comparative genomic analysis revealed 96 gene families conserved across all three strains but absent from other *Sphingobium* genomes, with 80 classified as highly conserved hypothetical proteins and 16 functionally annotated genes including arylsulfatases, heavy metal efflux systems, and ion transporters that may support their plant-associated niche. Chemotaxonomic analysis demonstrated predominant fatty acids distinguishing them from *S. yanoikuyae* JCM 7371$^T$, while physiological tests showed the strains metabolized plant-derived carbon compounds not utilized by *S. yanoikuyae* JCM 7371$^T$. Based on phylogenomic, genomic, and phenotypic analyses, these strains represent a novel species, for which the name *Sphingobium salicis* sp. nov. was proposed, with WW5$^T$ (DSM 120182, NCCB 101075) designated as the type strain.


# INTRODUCTION

The genus *Sphingobium* is a member of the family Sphingomonadaceae (Alphaproteobacteria), a diverse group of Gram-negative, rod-shaped bacteria characterized by glycosphingolipid-containing outer membranes. *Sphingobium* was established by (Takeuchi et al. 2001) during a phylogenetic reorganization that split *Sphingomonas* into four distinct genera: *Sphingomonas* (sensu stricto), *Sphingobium*, *Novosphingobium*, and *Sphingopyxis*. The earliest described member of the genus, *Sphingobium yanoikuyae*, was originally isolated by (Gibson et al. 1975) and initially assigned to *Beijerinckia*, later reclassified as *Sphingomonas yanoikuyae* (Yabuuchi et al. 1990), and finally transferred to *Sphingobium* (Takeuchi et al. 2001). Currently, the genus comprises 48 validly published species according to the List of Prokaryotic Names with Standing in Nomenclature (LPSN, accessed April 2025). Well-characterized members include *S. yanoikuyae*, *S. indicum*, *S. japonicum*, and *S. herbicidovorans*, which have been isolated from diverse terrestrial and aquatic environments and are noted for their metabolic versatility in degrading recalcitrant aromatic compounds (Glaeser and Kämpfer 2015).

*Sphingobium* species are widely distributed environmental bacteria commonly found in soils, freshwater systems, groundwater, and sediments. Members of the genus are frequently isolated from anthropogenically impacted environments, including petroleum-contaminated groundwater, PAH-impacted sediments, and other hydrocarbon-rich sites, where they demonstrate aerobic degradation of complex aromatic compounds, including 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). In addition to contaminated environments, *Sphingobium* species have been frequently isolated from inside of plants, including as true endophytes of maize, rice, and from coastal dune grasses and other wild plants (Ali et al. 2018; Boss et al. 2022; Jou et al. 2022; Ma et al. 2025).

Endophytes, bacteria and other microorganisms colonizing the internal tissues of plants, contribute to improved host plant tolerance to environmental stresses via increased nutrient acquisition, phytohormone production, stress response, and increased resistance to biotic and abiotic stresses (Reinhold-Hurek and Hurek 2011; Hardoim et al. 2015). Among these benefits, biological nitrogen fixation by diazotrophic endophytes increased phosphorus solubility and root uptake from soil can improve both nitrogen and phosphorus status of crop plants growing under nutrient limited conditions (Ercole et al. 2025). Recently, we have shown that three endophytic *Sphingobium* strains WW5$^{T}$, 11R-BB, and HT1-2 enhance nitrogenase activity in diazotrophic endophyte partners when co-cultured in nitrogen-limited conditions (Sher et al. 2024). These synergistic effects, which may involve mitigation of nitrogenase oxygen sensitivity or as yet unknown secreted diffusible compounds, were observed in different diazotrophic partners (*Azorhizobium* sp. HT1-9, *Rahnella aceris* WP5, and *Azospirillum* sp. 11R-A) and were not restricted to specific hosts or geographic locations.

Genome-based classification reveals greater phylogenetic diversity within the genus *Sphingobium* than is represented in current formal taxonomy. The Genome Taxonomy Database (GTDB release R226, accessed April 2025) identifies 137 species clusters across 381 *Sphingobium* genomes, yet only 48 clusters (35%) correspond to formally described species. Despite the ecological and functional diversity represented within the genus, taxonomic resolution remains incomplete. In this report, we describe the three *Sphingobium* strains, WW5$^{T}$, 11R-BB, and HT1-2, isolated from plants growing on nutrient-limited primary substrates, including Sitka willow (*Salix sitchensis*), black cottonwood (*Populus trichocarpa*), and salvia (*Pluchea carolinensis*), respectively. These strains are phylogenetically close to *S. yanoikuyae* yet display distinct genomic and phenotypic features. Based on phylogenomic and phenotypic analyses, we propose that they represent a novel species, *Sphingobium salicis* sp. nov., with WW5 designated as the type strain.

# METHODS

## Isolation and culture conditions

The three strains were isolated from diverse plant hosts growing on primary rock substrate (previously described in Doty et al. 2009; Sher et al. 2024). Briefly, strain HT1-2 was obtained from *Pluchea*

*carolinensis* shoots on lava fields near Kona, Hawai'i (19.142559° N, 155.752072° W), while the strains 11R-BB and WW5$^T$ were isolated from riparian zones in Washington state: 11R-BB from *Populus trichocarpa* roots along the Skykomish River (47.780563° N, 121.500670° W) and WW5$^T$ from *Salix sitchensis* stems along the Snoqualmie River (47.5206° N, 121.7745° W).

Hawaiian samples were surface-sterilized according to (Sher et al. 2024) while poplar and willow tissues used the method described by (Doty et al. 2005). Sterilized tissues were homogenized in mortars with either water (Hawaiian samples) or nitrogen-free medium (NFM; poplar and willow). Homogenates (100 µL) were plated on nitrogen-free combined carbon medium (NF-CCM, Rennie 1981) for Hawaiian and poplar root samples or yeast mannitol agar (YMA) for poplar and willow stems. Distinct colonies were purified through successive streak-plates on mannitol glutamate/Luria (MGL; (Cangelosi et al. 1991) agar, after incubation (30 ℃, 24–58 hr), and preserved at -80°C in 25% glycerol.

## DNA extraction and 16S rRNA gene sequence phylogenetic analysis

Genomic DNA was purified using standard protocols with phenol/chloroform extraction and ethanol precipitation as previously described (Doty et al. 2023). Genus level identification utilized PCR amplification of the 16S rRNA gene with universal primers 8F (5'-AGAGTTTGATCCTGGCTCAG-3') and 1492R (5'-GGTTACCTTGTTACGACTT-3'). PCR amplification used the following program: 95°C for 10 min; 40 cycles of 95°C for 1 min, 53°C for 1 min, 72°C for 1 min; final extension at 72°C for 10 min. Amplicons were sequenced by Genewiz (Azenta Life Sciences, NJ), and the resulting sequences were analyzed using BLAST against the NCBI database with a ≥ 98% sequence identity threshold for *Sphingobium* genus-level identification.

## Genome sequencing, assembly, and annotation

Whole-genome sequencing was completed for WW5, HT1-2, and 11R-BB using a hybrid approach combining Illumina and Oxford Nanopore technologies (CD Genomics, NY). Illumina libraries (280 bp insert size) were sequenced on a NovaSeq X Plus system (2×151 bp paired-end reads). Demultiplexing and adapter trimming were performed using bcl-convert (v4.2.4). Nanopore libraries were prepared using the PCR-free Ligation Sequencing kit and sequenced on MinION Mk1B or GridION systems with R10.4.1 flow cells. Basecalling, demultiplexing, and adapter removal were conducted with Guppy (v6.5.7) in super-accurate mode.

Raw sequencing data were deposited in the Bacterial and Viral Bioinformatics Resource Center (BV-BRC) (Olson et al. 2023). High-quality hybrid assemblies were constructed using Unicycler v0.4.8, combining SPAdes v3.13.0 short-read assembly (Bankevich et al. 2012) with long-read graph resolution (Wick et al. 2017). Assemblies were refined through four iterations of PILON and RACON polishing, with contigs less than1000 bp or greater than 10× coverage excluded. Assembly completeness was assessed via CheckM (Parks et al. 2015). Genome annotation was performed using the RASTtk pipeline v1.073 (Brettin et al.

2015). Annotated assemblies and raw sequence data were deposited in BV-BRC and GenBank (Table S1).

## Whole-genome phylogeny and taxonomic placement

Taxonomic placements were assessed using two whole-genome methods. First, digital DNA-DNA hybridization (dDDH, formula d4) was calculated between each study genome and all *Sphingobium* type-strain assemblies via the Type (Strain) Genome Server (TYGS; (Meier-Kolthoff and Göker 2019). Second, Average Nucleotide Identity (ANI) values were calculated between study genomes and type strain assemblies using FastANI v1.34 with a 95–96% species boundary threshold (Jain et al. 2018). Finally, the Genome Taxonomy Database Toolkit (GTDB-Tk) (Parks et al. 2022) was used to extract, align, and concatenate 120 bacterial single-copy marker proteins (bac120) from strains WW5, 11R-BB, and HT1-2. These were analyzed together with the species-representative *Sphingobium* genomes in the GTDB-Tk release 220. The resulting alignment file was analyzed in IQ-TREE (Nguyen et al. 2015) to construct a maximum likelihood phylogenetic tree under the Whelan and Goldman (WAG) model, with 1,000 pseudo-bootstrap replicates. Phylogenetic placement of individual strains was conducted using pplacer (Matsen et al. 2010), which assigns query genomes to positions within a fixed reference tree using maximum likelihood criteria. The phylogenetic tree was rooted using Sphingobium paucimobilis NBRC 13935. Statistical support for tree topology was assessed using 1,000 pseudo-bootstrap replicates with an 80.0% node support threshold. The phylogenetic tree was rooted using *Sphingobium paucimobilis* NBRC 13935.

## Pathogenicity Analysis

Pathogenicity Analysis
Finally, the pathogenic potential and virulence determinants of the strains were assessed with PathogenFinder v1.1 (Cosentino et al. 2013) and VirulenceFinder v2.0 (Joensen et al. 2014). Draft genome assemblies of the three strains were uploaded to the respective web servers using the default parameters. VirulenceFinder was run with default thresholds (≥ 90% nucleotide identity; ≥60% alignment length) against the *Staphylococcus aureus*, *Enterococcus*, *Listeria*, and *Escherichia coli* databases.

# Comparative genomics and functional analysis

The BV-BRC Comparative Systems pipeline was used to conduct a comparative genome analysis to identify genes involved in the glycolytic pathway, TCA cycle, oxidative phosphorylation, anaplerotic reactions, and nutrient uptake via ABC-transporters. Functional annotation was performed using KEGG KOfam (v2025-04-01, release 114.0) as the reference database (Aramaki et al. 2020). Gene conservation was determined through comparison with a data set of 61 *Sphingobium* type strain assemblies obtained from the NCBI ReSeq database (Table S1). Gene datasets were filtered using PATRIC genus-specific protein families (PLfams), and core genome and strain-specific genes were identified. The BV-BRC

Protein Family Sorter was used to analyze conserved and unique genomic features and protein family distributions to classify core and accessory gene sets.

## Morphological, physiological, and biochemical analyses

The strains WW5$^T$, 11R-BB, and HT1-2 were grown for 3 days at 25.5°C in nitrogen-limited combined carbon medium (NL-CCM; modified from Rennie 1981). Cells were Gram stained with a kit (Innovating Science, Avon, NY, USA) and visualized with a brightfield microscope (NuLife Sciences, Chattanooga, TN, USA) using an oil immersion lens at 1000x total magnification. Colony morphology was characterized on both mannitol glutamate/Luria (MGL; (Cangelosi et al. 1991) and NLCCM after 2- and 3-days incubation, respectively. The ability of each strain to grow in modified nitrogen-free NLCCM (without yeast extract) was assessed. Cultures were incubated at 25.5°C for up to 7 days with daily monitoring for growth. Oxygen preference was determined using fluid thioglycolate medium (Carolina, Burlington, NC, USA) after 2 days incubation at 25.5°C. Growth at various temperatures (4, 25.5, 30, 35, and 37°C) was tested on MGL agar plates incubated for up to 7 days with daily monitoring. Carbon source utilization, osmotic stress tolerance, and pH tolerance were determined using Biolog MicroPlates™ PM1, PM9, and PM10 (Biolog, Hayward, CA, USA), respectively. Each plate was prepared according to the manufacturer's protocol.

## Chemotaxonomic analysis

The Fatty Acid Methyl Ester (FAME) analysis was performed following the MIDI automated Microbial Identification System guidelines (Kunitsky et al. nd). For biomass production the four strains were cultivated aerobically at 25°C for 24 hr on trypticase soy broth agar (M535, https://mediadive.dsmz.de/medium/535). The gas chromatography, flame ionization detector (GC-FID) and mass spectrometry (MS) analyses of the cellular fatty acids were carried out by DSMZ Services (Leibniz-Institute DSMZ, Braunschweig, Germany).

## RESULTS AND DISCUSSION

# Ecological context of strain isolation

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

provided a framework for investigating the role of endophytes in the nitrogen dynamics of non-leguminous plants.

# 16S rRNA gene phylogeny

Phylogenetic analysis based on 16S rRNA gene sequences placed strains WW5$^T$, 11R-BB, and HT1-2 within the genus *Sphingobium*, with sequence identities of 99.59%, 99.52%, and 99.45% to *S. yanoikuyae* JCM 7371, respectively. Although these values exceed the traditional 98.7% threshold for species delineation (Yarza et al. 2014), additional genomic analyses were conducted because the 16S rRNA gene sequence similarity provides insufficient resolution for species assignment (Rosselló-Móra and Amann 2015).

# Genome sequencing, assembly, and annotation

Hybrid Illumina-Nanopore sequencing generated high-quality assemblies for *S. salicis* strains WW5, 11R-BB, and HT1-2. Genome sizes ranged from 5.39–5.72 Mb with ~ 64% GC content and ≤ 0.4% contamination as assessed with CheckM (Table S1). Final assemblies comprised 8 (WW5$^T$), 10 (11R-BB), and 7 (HT1-2) contigs with N50 values of 5.06 Mb (WW5 and 11R-BB) and 4.67 Mb (HT1-2). Annotations revealed 5,404-5,635 coding sequences per genome, with hypothetical proteins constituting 38–44% of predicted proteins and assignments to PATRIC genus-specific protein families ranging from 4,872-5,095. The genome assemblies were deposited in the NCBI database BioProject PRJNA1059640 for strain WW5$^T$ (BioSample SAMN39212602), and BioProject PRJNA1058978 for strains 11R-BB (BioSample SAMN39201427) and HT1-2 (BioSample SAMN39201428).

# Whole-genome phylogeny and taxonomic placement

Whole-genome taxonomic classification using TYGS (dDDH) and GTDB-Tk provided genome-scale resolution. TYGS phylogenomic analysis placed all three strains in a monophyletic clade with 100% bootstrap support (Fig. 1), with dDDH values of 99.9% between WW5$^T$ and 11R-BB, while HT1-2 shared 70.1% and 70.0% similarity with WW5$^T$ and 11R-BB, respectively. The dDDH values relative to *S. yanoikuyae* JCM 7371 (66.2–68.2%) fell below the 70% species delineation threshold for all three strains, indicating distinct species status (Table 1).

Table 1

TYGS derived digital DNA–DNA hybridization ($d_4$) values, confidence intervals, and G + C content differences used to assess genomic similarity among a) Sphingobium salicis strains, and b) between WW5 and type strains of related Sphingobium and Sphingomonas species. Original isolation sources and literature provided for all strains. Only the most 10 closely matches strains shown.

**Table 1a. TYGS pairwise comparisons among *S. salicis* strains WW5$^T$, 11R-BB, and HT1-2.**

| Query | Subject | $d_4$ | C.I. $d_4$ | Diff. GC% |
|---|---|---|---|---|
| WW5$^T$ | *S. salicis* 11R-BB | 100.0 | 99.9–100.0 | 0.08 |
| 11R-BB | *S. salicis* HT1-2 | 70.0 | 67.0-72.9 | 0.13 |
| HT1-2 | *S. salicis* WW5$^T$ | 70.1 | 67.1–72.9 | 0.21 |

Table 1b. TYGS comparisons between *S. salicis* strain WW5$^T$ and type strains of related species.

| Subject | $d_4$ | C.I. $d_4$ | Diff. GC% |
|---|---|---|---|
| *S. yanoikuyae* JCM 7371 | 66.2 | 63.2–69.0 | 0.39 |
| *S. scionense* DSM 19371 | 51.6 | 48.9–54.2 | 0.11 |
| *S. rhizovicinum* CCM 7491 | 26.3 | 23.9–28.8 | 0.69 |
| *S. paulinellae* KACC 19283 | 26.2 | 23.8–28.6 | 0.64 |
| *S. algicolab* KACC 19283 | 26.2 | 23.8–28.6 | 0.68 |
| *S. limneticum* DSM25076 | 26.1 | 23.7–28.6 | 0.74 |
| *S. lucknowense* CCM 7544 | 24.9 | 22.5–27.3 | 0.35 |
| *S. cloacae* NBRC 102517 | 24.7 | 22.4–27.2 | 0.56 |
| *Sphingomonas bisphenolicum* AO1 | 24.5 | 22.2–27.0 | 0.06 |
| *S. francense* DSM 26779 | 24.4 | 22.1–26.9 | 1.45 |

GTDB-Tk classified the three strains as *Sphingobium yanoikuyae_A* (Fig. 2), a provisional designation for genomically divergent lineages within named taxa that lack formal taxonomic resolution (Table S3; Parks et al. 2022). ANI values relative to *S. yanoikuyae* JCM 7371 ranged from 95.56–96.65%, placing all strains within or slightly above the 95–96% boundary zone for species delineation (Richter and Rosselló-Móra 2009). This placement reflects the disparity between genome-based species delineation and formal taxonomy, as *S. yanoikuyae* currently comprises two GTDB-recognized lineages that highlight unresolved taxonomic boundaries (Parks et al. 2020; Wu et al. 2025).

# Pathogenicity Analysis

Pathogenicity analysis using PathogenFinder v1.1 classified all strains as non-pathogenic (Table S4). VirulenceFinder v2.0 detected no virulence factors against *S. aureus*, *Listeria*, *E. coli*, or *Enterococcus* databases.

# Comparative genomics and functional analysis

Comparative genomic analysis compared *S. salicis* strains WW5□, 11R-BB, and HT1-2 with 61 *Sphingobium* type strains, identifying 570 orthologs conserved across the 64 *Sphingobium* genomes (Table S5). This relatively small core genome is characteristic of *Sphingobium* and other sphingomonads, reflecting the high genomic diversity that enables their broad ecological distribution (Aylward et al. 2013; Wang et al. 2018; Verma et al. 2020).

Species-specific analysis identified 96 orthologs conserved in *S. salicis* but absent from *Sphingobium* type strains, comprising 16 functionally annotated and 80 hypothetical proteins (Table S6). These species-specific genes represent a core accessory genome that distinguishes *S. salicis* from other described *Sphingobium* species. The hypothetical proteins displayed remarkably high conservation, with 72 of 80 showing ≥ 95% sequence homology at the amino acid level, contrasting with typical accessory gene variability (Tambong et al. 2023; Terra et al. 2025). This conservation suggests functional constraint consistent with niche-specific adaptation, indicating these genes likely represent uncharacterized proteins with conserved function under selective pressure for specialized cellular processes in plant-associated environments (Rinke et al. 2013; Lidbury et al. 2021).

Individual strain analysis revealed additional accessory genomic content, with 328, 326, and 282 strain-specific accessory genes in WW5□, 11R-BB, and HT1-2, respectively. Of these, 52 ($WW5^T$), 51 (11R-BB), and 43 (HT1-2) carried functional annotations (File S4), with strain HT1-2 uniquely encoding 27 additional functional families including copper-export ATPase and phosphonate transporters that may reflect adaptation to the volcanic substrate from which it was isolated.

*Plant-association genomic features.*

The 16 functionally annotated gene families conserved across all three *S. salicis* strains but absent from other *Sphingobium* species align with established categories for plant-microbe interactions (Hacquard et al. 2015; Levy et al. 2018). These genes can be organized into three functional categories, direct interaction, stress resistance, and nutrient acquisition, which could support endophytic colonization and plant-beneficial interactions.

Direct interaction mechanisms included the CzcABC RND system (EC 3.6.3.16) for metal efflux (Nies 1995; Nies 2003), $Na^+/H^+$ antiporter (EC 3.6.3.26) for pH homeostasis during apoplastic colonization (Habibian et al. 2005)d lactate permease for organic acid utilization from root exudates (Wang et al. 2019). Four arylsulfatases (EC 3.1.6.1) enable utilization of plant sulfate compounds (Toesch et al. 2014), while molybdate ATPase (EC 7.3.2.5) supports enhanced nutrient acquisition (Cheng et al. 2016). Regulatory circuits facilitate metabolite detection and response coordination (Maddocks and Oyston 2008). Stress resistance systems comprised β-lactamase (EC 3.5.2.6) and additional metal efflux

components that facilitate persistence under interspecies antagonism and elevated metal concentrations encountered in plant tissues (Martínez 2008). These systems complement the core stress response pathways and provide enhanced tolerance to the chemical challenges of endophytic environments. Nutrient acquisition systems included high-affinity phosphate transport components and alkaline phosphatase (EC 3.1.3.1), supporting uptake of both inorganic and organic phosphate sources (Chen et al. 2024). Iron acquisition was supported by TonB-dependent siderophore receptors and partial non-ribosomal peptide synthetase gene clusters, suggesting capacity for both siderophore uptake and biosynthesis. These systems enable effective nutrient scavenging in the nutrient-limited environments where *S. salicis* strains were isolated. The conservation of these functional categories across all three strains, combined with their absence in the other *Sphingobium* species, may indicate adaptations for the endophytic lifestyle that help distinguish *S. salicis* from its environmental relatives.

# Metabolic distinctions from S. yanoikuyae JCM 7371$^T$

Metabolic reconstruction revealed key differences between *S. salicis* and *S. yanoikuyae* JCM 7371$^T$ that distinguish the species. Both species possessed complete central carbon metabolism pathways but differed at critical branch points affecting carbon source utilization and redox metabolism.

Neither species encoded hexokinase (EC 2.7.1.1), suggesting both relied on alternative routes for glucose phosphorylation. However, their pentose phosphate pathways differed in that *S. yanoikuyae* possessed 6-phosphogluconate dehydrogenase (EC 1.1.1.44) for NADPH generation through the oxidative branch, while *S. salicis* lacked this enzyme, potentially limiting its NADPH production capacity. In contrast, *S. salicis* uniquely encoded mannate dehydratase (EC 4.2.1.47), enabling mannose metabolism through sugar acid intermediates. These metabolic differences suggest distinct redox strategies that may reflect adaptation to endophytic versus environmental lifestyles (Taghavi et al. 2010; Chavarría et al. 2013).

Sugar utilization capabilities further distinguished the species. *S. salicis* genomes encoded complete enzymatic pathways for galactose utilization, including galactokinase (EC 2.7.1.6), galactose-1-phosphate uridylyltransferase (EC 2.7.7.12), and UDP-glucose 4-epimerase (EC 5.1.3.2), supporting efficient conversion of plant-derived galactose sources. In contrast, canonical fructose phosphorylation via fructokinase (EC 2.7.1.5) was absent in all *S. salicis* strains, indicating dependence on alternative fructose utilization pathways. These differences correlated with the observed phenotypic capacity of *S. salicis* to grow on D-galactose and D-fructose, substrates that failed to support growth of *S. yanoikuyae* JCM 7371$^T$.

Organic acid metabolism showed both conserved and divergent features. While both species encoded gluconokinase (EC 2.7.1.12) for gluconate phosphorylation, only *S. yanoikuyae* possessed the complete oxidative gluconate pathway through 6-phosphogluconate dehydrogenase, potentially limiting *S. salicis* access to this sugar acid catabolism route. Among *S. salicis* strains, HT1-2 uniquely encoded L-idonate 5-dehydrogenase (EC 1.1.1.264), potentially expanding access to plant-derived aldarate intermediates unavailable to the other strains. These metabolic distinctions support the species-level separation of *S.*

*salicis* from *S. yanoikuyae* and reflect adaptation to plant-derived carbon sources characteristic of endophytic environments.

# Nitrogen metabolism

Nitrogen acquisition and assimilation pathways were conserved across *S. salicis* strains. All genomes encoded assimilatory nitrate reductase (EC 1.7.1.4) and nitrite reductase (EC 1.7.99.4), enabling reduction of nitrate to ammonium, with downstream incorporation supported by glutamate dehydrogenase (EC 1.4.1.2), glutamine synthetase (EC 6.3.1.2), and glutamate synthase (EC 1.4.1.13). HT1-2 uniquely encoded nitric oxide-forming nitrite reductase (EC 1.7.99.1), potentially supporting microaerophilic denitrification in plant tissues (Asaf et al. 2018). These nitrogen acquisition capabilities were comparable between *S. salicis* and *S. yanoikuyae*, suggesting conserved strategies for nitrogen utilization.

# Stress response systems

Stress response systems in *S. salicis* included catalases (EC 1.11.1.6), superoxide dismutase (EC 1.15.1.1), and peroxiredoxin (EC 1.11.1.15) for oxidative stress management, complementing the metal efflux and β-lactamase systems described above. *S. salicis* further encoded glycine betaine biosynthesis enzymes (EC 1.1.99.1, EC 1.2.1.8) for osmoprotection and molecular chaperones for protein stability. These expanded stress tolerance mechanisms, particularly for oxidative and osmotic challenges, may facilitate survival in the variable conditions of plant tissues.

# Plant growth promotion traits

*S. salicis* strains encoded multiple mechanisms for plant growth promotion absent from *S. yanoikuyae*. For phytohormone production, a nitrilase/cyanide hydratase domain protein was predicted to catalyze conversion of indole-3-acetonitrile to indole-3-acetic acid (IAA), while indole-3-acetamide hydrolase (EC 3.5.1.4) may enable partial IAM-to-IAA conversion, though the upstream IAM-generating enzyme was absent (Sun et al. 2022; Tang et al. 2023). Nutrient mobilization capabilities included the high-affinity phosphate transport system and alkaline phosphatase described above, complemented by gluconate-generating enzymes that may facilitate phosphate solubilization through organic acid production. Iron acquisition was supported by TonB-dependent siderophore receptors and partial NRPS gene clusters, suggesting both uptake and biosynthesis capabilities. Additionally, *S. salicis* uniquely possessed dTDP-rhamnose biosynthesis genes (EC 4.2.1.46), potentially enhancing secondary metabolite glycosylation important for plant interactions (Wagstaff et al. 2021). These integrated nutrient acquisition and phytohormone production capabilities were absent from *S. yanoikuyae* JCM 7371$^{T}$.

# Morphological, physiological, and biochemical analyses

The *S. salicis* strains displayed identical carbon-utilization profiles, metabolizing all tested substrates except D-ribose and L-lyxose (Table 2). In contrast, *S. yanoikuyae* JCM 7371 failed to grow on D-galactose, D-gluconic acid, D,L-malate, L-malate, and D-fructose; carbon compounds commonly found in plant tissues and root exudates (Badri and Vivanco 2009). The *S. salicis* strains possessed conserved

metabolic pathways for monosaccharides, disaccharides, and organic acids commonly released by plant tissues. Their ABC- and PTS-mediated uptake systems and complete central metabolic pathways support efficient carbon acquisition under nutrient limitation (Deutscher et al. 2006; Badri and Vivanco 2009). Cell wall-modifying enzymes, including xylanases and pectin acetylesterases, suggest capacity for controlled plant tissue modification that may facilitate endophytic colonization and nutrient acquisition, consistent with beneficial plant-microbe interactions where such enzymes enable tissue penetration and movement while avoiding pathogenic responses through regulated expression (Compant et al. 2005; Piromyou et al. 2015). The consistent genotype-phenotype concordance indicates strong selective pressure to maintain these functions for successful plant tissue colonization.

# Chemotaxonomic analysis

The cellular fatty-acid profiles of strains WW5□, 11R-BB, and HT1-2 were dominated by vaccenic acid (C18:1 ω7c, 57–60% of total FAME) with palmitoleic acid (C16:1 ω7c, 15–16%) as the secondary component (Table S7). Relative to *Sphingobium yanoikuyae* JCM 7371□, the three strains contained higher proportions of palmitic acid (C16:0, + 6–7%) and lower levels of C16:1 ω7c, C16:1 ω5c, and 14:0 2-OH. These consistent quantitative differences provide additional chemotaxonomic evidence that, together with genomic and phenotypic data, supports recognition of the *S. salicis* strains as a distinct *Sphingobium* species, while the *S. yanoikuyae* profile matched the published reference range (Yabuuchi et al., 1990; Khan et al., 1996).

Table 2

Physiological and biochemical characteristics of S. salicis WW5T, 11R-BB, HT1-2, and Sphingobium yanoikuyae JCM 7371T. Growth parameters (temperature: 4–37°C, pH, NaCl tolerance) and carbon utilization profiles were assessed using Biolog MicroPlates™ PM1, PM9, and PM10. Temperature ranges were determined on MGL agar plates, and oxygen preference in fluid thioglycolate medium after 48 h at 25.5°C. "+" indicates growth; "–" indicates no growth. Strains: 1, S. salicis WW5□; 2, S. salicis 11R-BB; 3, S. salicis HT1-2; 4, S. yanoikuyae JCM 7371□. ND, no data.

| **Characteristic** | **1** | **2** | **3** | **4** |
|---|---|---|---|---|
| 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 |
| **Carbon Source Utilization** | | | | |
| L-Arabinose | + | + | + | + |
| D-Galactose | + | + | + | - |
| D-Trehalose | + | + | + | + |
| D-Mannose | + | + | + | + |
| D-Gluconic Acid | + | + | + | - |
| D-Xylose | + | + | + | + |
| D,L-Malic Acid | + | + | + | - |
| D-Ribose | - | - | - | - |
| L-Rhamnose | + | + | + | + |
| D-Fructose | + | + | + | - |
| α-D-Glucose | + | + | + | + |
| Maltose | + | + | + | + |
| D-Melibiose | + | + | + | + |
| α-D-Lactose | + | + | + | + |
| Sucrose | + | + | + | + |
| ß-Methyl-D- Glucoside | + | + | + | + |
| D-Cellobiose | + | + | + | + |
| L-Malic Acid | + | + | + | - |
| L-Lyxose | - | - | - | - |

# Taxonomic conclusions

The combined genomic, metabolic, and phenotypic evidence establishes *Sphingobium salicis* as a distinct species specialized for plant association. Genome-based delineation (dDDH 66.2–68.2% to *S. yanoikuyae*), phylogenomic placement within the *Sphingobium yanoikuyae_A* clade, and conserved plant-interaction genes support the recognition of *S. salicis* sp. nov.. The specialized metabolic capabilities for plant-derived carbon utilization, nitrogen acquisition systems, and expanded stress response systems distinguish *S. salicis* from its closest relative and establish it as an endophytic bacterium adapted to nutrient-limited plant environments.

### 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 source of isolation). Cells are Gram-negative, rod-shaped, catalase positive and strictly aerobic. Growth occurs on NLCCM and MGL media, with colonies visible after 2–3 days incubation. The optimum temperature range for growth is 25–35°C, with reduced growth at 37°C. Tolerates pH 5–8 (optimum pH 7) and NaCl concentrations up to 2% (WW5 and 11R-BB) or 3% (HT1-2). Positive growth was observed for D-galactose, D-gluconic acid, D,L-malate, L-malate, and D-fructose, but not D-ribose or L-lyxose. Possesses complete Embden-Meyerhof-Parnas, Entner-Doudoroff, and pentose phosphate pathways. Contains machinery for galactose, gluconic acid, fructose, and malate metabolism, supporting utilization of plant-derived carbon compounds. Major cellular fatty acids are vaccenic acid (18:1 ω7c, ~ 58%) and palmitoleic acid (16:1 ω7c, 15–21%). Contains higher concentrations of methyl palmitate (16:0) compared to *S. yanoikuyae* JCM 7371, with distinctive differences in 16:0, 16:1 ω7c, 16:1 ω5c, and 14:0 2OH fatty acids. The DNA G + C content is approximately 63.9–64.3%. Genome sizes range from 5.39 to 5.72 Mb. Isolated from willow (*Salix* spp.) tissues as an endophytic bacterium. The type strain is WW5. The genome assemblies have been deposited in the NCBI database under BioProject PRJNA1058978 for strains 11R-BB (BioSample SAMN39201427) and HT1-2 (BioSample SAMN39201428), and under BioProject PRJNA1059640 for strain WW5 (BioSample SAMN39212602).

# Declarations

### Supplementary Information

The online version contains supplementary material available at https----------.

### Author Contribution

JF conceived the experiments with help from AF. AF and RT conducted the genomic and bioinformatics analyses. JF, ML, and DB conducted chemotaxonomic 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 have been deposited in the NCBI database under BioProjects PRJNA1058978 and PRJNA1059640. Strain WW5$^T$ culture are available in the Agricultural Research Culture Collection (NRRL), Netherlands Culture Collection of Bacteria (NCCB), and German Collection of Microorganisms and Cell Cultures (DSMZ) under the accession codes DSM 120182, NCCB 101075, and NRRL B-68079, respectively.

### Conflicts of Interest

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

### Ethical Approval

Not required

## References

1. Ali S, Isaacson J, Kroner Y, Saldias S, Kandasamy S, Lazarovits G. 2018. Corn sap bacterial endophytes and their potential in plant growth-promotion. Environmental Sustainability. 1(4):341–355. doi:10.1007/s42398-018-00030-4. [accessed 2025 May 29]. https://link.springer.com/article/10.1007/s42398-018-00030-4.
2. Angelosi GA, Abest E, Martinetti G, Nester EW. 1991. [18] Genetic Analysis of *Agrobacterium*. In: Methods in Enzymology. Vol. 204. Academic Press. (Bacterial Genetic Systems). p. 384–397. [accessed 2025 May 2]. https://www.sciencedirect.com/science/article/pii/007668799104020O.
3. 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(7):2251–2252. doi:10.1093/bioinformatics/btz859. [accessed 2025 May 2]. https://doi.org/10.1093/bioinformatics/btz859.
4. 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. Applied and Environmental Microbiology. 79(12):3724–3733. doi:10.1128/AEM.00518-13. [accessed 2025 May 11]. https://journals.asm.org/doi/full/10.1128/aem.00518-13.
5. Badri DV, Vivanco JM. 2009. Regulation and function of root exudates. Plant, Cell & Environment. 32(6):666–681. doi:10.1111/j.1365-3040.2009.01926.x. [accessed 2025 Jun 19]. https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-3040.2009.01926.x.

6. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, et al. 2012. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. Journal of Computational Biology. 19(5):455–477. doi:10.1089/cmb.2012.0021. [accessed 2024 Feb 28]. https://www.liebertpub.com/doi/abs/10.1089%2Fcmb.2012.0021.
7. 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(1):508. doi:10.1186/s12864-022-08738-8.
8. Brettin T, Davis JJ, Disz T, Edwards RA, Gerdes S, Olsen GJ, Olson R, Overbeek R, Parrello B, Pusch GD, et al. 2015. RASTtk: A modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. Sci Rep. 5(1):8365. doi:10.1038/srep08365. [accessed 2024 Mar 4]. https://www.nature.com/articles/srep08365.
9. Chavarría M, Nikel PI, Pérez-Pantoja D, de Lorenzo V. 2013. The Entner-Doudoroff pathway empowers Pseudomonas putida KT2440 with a high tolerance to oxidative stress. Environ Microbiol. 15(6):1772–1785. doi:10.1111/1462-2920.12069.
10. Chen B, Zhang M, Lin D, Ye J, Tang K. 2024. Roseihalotalea indica gen. nov., sp. nov., a halophilic Bacteroidetes from mesopelagic Southwest Indian Ocean with higher carbohydrate metabolic potential. Antonie van Leeuwenhoek. 117(1):66. doi:10.1007/s10482-024-01965-x. [accessed 2025 Jun 9]. https://doi.org/10.1007/s10482-024-01965-x.
11. Cheng G, Karunakaran R, East AK, Poole PS. 2016. Multiplicity of Sulfate and Molybdate Transporters and Their Role in Nitrogen Fixation in Rhizobium leguminosarum bv. viciae Rlv3841. MPMI. 29(2):143–152. doi:10.1094/MPMI-09-15-0215-R. [accessed 2025 May 31]. https://apsjournals.apsnet.org/doi/10.1094/MPMI-09-15-0215-R.
12. 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. Applied and Environmental Microbiology. 71(4):1685–1693. doi:10.1128/AEM.71.4.1685-1693.2005. [accessed 2025 Jun 19]. https://journals.asm.org/doi/10.1128/aem.71.4.1685-1693.2005.
13. Cosentino S, Larsen MV, Aarestrup FM, Lund O. 2013. PathogenFinder - Distinguishing Friend from Foe Using Bacterial Whole Genome Sequence Data. PLOS ONE. 8(10):e77302. doi:10.1371/journal.pone.0077302. [accessed 2024 Apr 9]. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0077302.
14. Deutscher J, Francke C, Postma PW. 2006. How Phosphotransferase System-Related Protein Phosphorylation Regulates Carbohydrate Metabolism in Bacteria. Microbiol Mol Biol Rev. 70(4):939–1031. doi:10.1128/MMBR.00024-06. [accessed 2025 Jun 19]. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1698508/.
15. Doty SL, Dosher MR, Singleton GL, Moore AL, Aken BV, Stettler RF, Strand SE, Gordon MP. 2005. Identification of an endophytic Rhizobium in stems of Populus. Synbiosis. 39:27–35.

16. 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(1):13. doi:10.3390/pathogens12010013. [accessed 2024 Oct 21]. https://www.mdpi.com/2076-0817/12/1/13.
17. 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(1):23–33. doi:10.1007/BF03179967. [accessed 2025 Mar 24]. https://link.springer.com/10.1007/BF03179967.
18. 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(7):1–22. doi:10.1007/s10482-025-02100-0. [accessed 2025 Jun 10]. https://link-springer-com.offcampus.lib.washington.edu/article/10.1007/s10482-025-02100-0.
19. 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, New Series. 189(4199):295–297. http://www.jstor.org/stable/1740147.
20. Glaeser SP, Kämpfer P. 2015. Multilocus sequence analysis (MLSA) in prokaryotic taxonomy. Systematic and Applied Microbiology. doi:10.1016/j.syapm.2015.03.007.
21. Habibian R, Dzioba J, Barrett J, Galperin MY, Loewen PC, Dibrov P. 2005. Functional Analysis of Conserved Polar Residues in Vc-NhaD, Na+/H+ Antiporter of *Vibrio cholerae**. Journal of Biological Chemistry. 280(47):39637–39643. doi:10.1074/jbc.M509328200. [accessed 2025 May 31]. https://www.sciencedirect.com/science/article/pii/S0021925820591771.
22. Hacquard S, Garrido-Oter R, González A, Spaepen S, Ackermann G, Lebeis S, McHardy AC, Dangl JL, Knight R, Ley R, et al. 2015. Microbiota and Host Nutrition across Plant and Animal Kingdoms. Cell Host & Microbe. 17(5):603–616. doi:10.1016/j.chom.2015.04.009. [accessed 2025 Jun 19]. https://www.cell.com/cell-host-microbe/abstract/S1931-3128(15)00167-5.
23. 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. Microbiology and Molecular Biology Reviews. 79(3):293–320. doi:10.1128/mmbr.00050-14. [accessed 2024 Mar 30]. https://journals.asm.org/doi/10.1128/mmbr.00050-14.
24. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. 2018. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat Commun. 9(1):5114. doi:10.1038/s41467-018-07641-9. [accessed 2024 Apr 6]. https://www.nature.com/articles/s41467-018-07641-9.
25. 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(5):1501–1510. doi:10.1128/JCM.03617-13.
26. 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(11):1890. doi:10.3390/agriculture12111890. [accessed 2025 May 2]. https://www.mdpi.com/2077-0472/12/11/1890.

27. 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 Bacteriol. 46(2):466–469. doi:10.1099/00207713-46-2-466.
28. Kunitsky C, Osterhout G, Sasser M. nd. IDENTIFICATION OF MICROORGANISMS USING FATTY ACID METHYL ESTER (FAME) ANALYSIS AND THE MIDI SHERLOCK® MICROBIAL IDENTIFICATION SYSTEM.
29. Levy A, Salas Gonzalez I, Mittelviefhaus M, Clingenpeel S, Herrera Paredes S, Miao J, Wang K, Devescovi G, Stillman K, Monteiro F, et al. 2018. Genomic features of bacterial adaptation to plants. Nat Genet. 50(1):138–150. doi:10.1038/s41588-017-0012-9. [accessed 2024 Apr 9]. https://www.nature.com/articles/s41588-017-0012-9.
30. Lidbury IDEA, Borsetto C, Murphy ARJ, Bottrill A, Jones AME, Bending GD, Hammond JP, Chen Y, Wellington EMH, Scanlan DJ. 2021. Niche-adaptation in plant-associated Bacteroidetes favours specialisation in organic phosphorus mineralisation. ISME J. 15(4):1040–1055. doi:10.1038/s41396-020-00829-2. [accessed 2025 Jun 1]. https://www.nature.com/articles/s41396-020-00829-2.
31. Ma J, Peng Q, Chen S, Liu Z, Zhang W, Zhang C, Du X, Sun S, Peng W, Lei Z, et al. 2025. Microbiome Migration from Soil to Leaves in Maize and Rice. Microorganisms. 13(4):947. doi:10.3390/microorganisms13040947. [accessed 2025 May 29]. https://www.mdpi.com/2076-2607/13/4/947.
32. Maddocks SE, Oyston PCF. 2008. Structure and function of the LysR-type transcriptional regulator (LTTR) family proteins. Microbiology. 154(12):3609–3623. doi:10.1099/mic.0.2008/022772-0. [accessed 2025 May 31]. https://www.microbiologyresearch.org/content/journal/micro/10.1099/mic.0.2008/022772-0.
33. Martínez JL. 2008. Antibiotics and Antibiotic Resistance Genes in Natural Environments. Science. 321(5887):365–367. doi:10.1126/science.1159483. [accessed 2025 May 31]. https://www.science.org/doi/10.1126/science.1159483.
34. 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(1):538. doi:10.1186/1471-2105-11-538. [accessed 2025 Jun 13]. https://doi.org/10.1186/1471-2105-11-538.
35. Meier-Kolthoff JP, Göker M. 2019. TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy. Nat Commun. 10(1):2182. doi:10.1038/s41467-019-10210-3. [accessed 2024 Feb 28]. https://www.nature.com/articles/s41467-019-10210-3.
36. 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. doi:10.12688/f1000research.25284.1. [accessed 2025 May 2]. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7477647/.

37. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. Molecular Biology and Evolution. 32(1):268–274. doi:10.1093/molbev/msu300. [accessed 2025 May 2]. https://doi.org/10.1093/molbev/msu300.
38. Olson RD, Assaf R, Brettin T, Conrad N, Cucinell C, Davis JJ, Dempsey DM, Dickerman A, Dietrich EM, Kenyon RW, et al. 2023. Introducing the Bacterial and Viral Bioinformatics Resource Center (BV-BRC): a resource combining PATRIC, IRD and ViPR. Nucleic Acids Res. 51(D1):D678–D689. doi:10.1093/nar/gkac1003.
39. Parks DH, Chuvochina M, Chaumeil P-A, Rinke C, Mussig AJ, Hugenholtz P. 2020. A complete domain-to-species taxonomy for Bacteria and Archaea. Nat Biotechnol. 38(9):1079–1086. doi:10.1038/s41587-020-0501-8. [accessed 2025 Feb 20]. https://www.nature.com/articles/s41587-020-0501-8.
40. 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 Research. 50(D1):D785–D794. doi:10.1093/nar/gkab776. [accessed 2024 Jul 21]. https://doi.org/10.1093/nar/gkab776.
41. 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(7):1043–1055. doi:10.1101/gr.186072.114. [accessed 2024 Feb 28]. https://genome.cshlp.org/content/25/7/1043.
42. Pinyakong O, Habe H, Omori T. 2003. The unique aromatic catabolic genes in sphingomonads degrading polycyclic aromatic hydrocarbons (PAHs). The Journal of General and Applied Microbiology. 49(1):1–19. doi:10.2323/jgam.49.1.
43. 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(9):3049–3061. doi:10.1128/AEM.04253-14. [accessed 2025 Jun 19]. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4393458/.
44. Reinhold-Hurek B, Hurek T. 2011. Living inside plants: bacterial endophytes. Current Opinion in Plant Biology. 14(4):435–443. doi:10.1016/j.pbi.2011.04.004. [accessed 2024 Mar 30]. https://www.sciencedirect.com/science/article/pii/S1369526611000549.
45. Rennie RJ. 1981. A single medium for the isolation of acetylene-reducing (dinitrogen-fixing) bacteria from soils. Can J Microbiol. 27(1):8–14. doi:10.1139/m81-002.
46. Richter M, Rosselló-Móra R. 2009. Shifting the genomic gold standard for the prokaryotic species definition. Proceedings of the National Academy of Sciences of the United States of America.

doi:10.1073/pnas.0906412106.

47. Rinke C, Schwientek P, Sczyrba A, Ivanova NN, Anderson IJ, Cheng J-F, Darling A, Malfatti S, Swan BK, Gies EA, et al. 2013. Insights into the phylogeny and coding potential of microbial dark matter. Nature. 499(7459):431–437. doi:10.1038/nature12352. [accessed 2025 May 2]. https://www.nature.com/articles/nature12352.
48. Rosselló-Móra R, Amann R. 2015. Past and future species definitions for Bacteria and Archaea. Systematic and Applied Microbiology. doi:10.1016/j.syapm.2015.02.001.
49. Seo J-S, Keum Y-S, Li QX. 2009. Bacterial Degradation of Aromatic Compounds. International Journal of Environmental Research and Public Health. 6(1):278–309. doi:10.3390/ijerph6010278. [accessed 2025 May 2]. https://www.mdpi.com/1660-4601/6/1/278.
50. Sher AW, Tournay RJ, Gomez-Rivas E, Doty SL. 2024. Bacterial synergies amplify nitrogenase activity in diverse systems. ISME Communications. 4(1):ycae158. doi:10.1093/ismeco/ycae158. [accessed 2025 Jan 6]. https://doi.org/10.1093/ismeco/ycae158.
51. Stolz A. 2009. Molecular characteristics of xenobiotic-degrading sphingomonads. Appl Microbiol Biotechnol. 81(5):793–811. doi:10.1007/s00253-008-1752-3.
52. Sun H, Zhang J, Liu W, E W, Wang X, Li H, Cui Y, Zhao D, Liu K, Du B, et al. 2022. Identification and combinatorial engineering of indole-3-acetic acid synthetic pathways in Paenibacillus polymyxa. Biotechnology for Biofuels and Bioproducts. 15(1):81. doi:10.1186/s13068-022-02181-3. [accessed 2025 Jun 3]. https://doi.org/10.1186/s13068-022-02181-3.
53. Taghavi S, van der Lelie D, Hoffman A, Zhang Y-B, Walla MD, Vangronsveld J, Newman L, Monchy S. 2010. Genome Sequence of the Plant Growth Promoting Endophytic Bacterium Enterobacter sp. 638. PLoS Genet. 6(5):e1000943. doi:10.1371/journal.pgen.1000943. [accessed 2025 Jun 3]. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2869309/.
54. 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(Pt 4):1405–1417. doi:10.1099/00207713-51-4-1405. [accessed 2024 Apr 7]. https://doi.org/10.1099/00207713-51-4-1405.
55. Tambong JT, Bilodeau GJ, Fall ML. 2023. Editorial: Insights on plant-associated microorganisms: diversity, systematics and genomics. Front Microbiol. 14:1323823. doi:10.3389/fmicb.2023.1323823. [accessed 2025 Jun 1]. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10644803/.
56. 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(1):232–240. doi:10.1021/es201842h. [accessed 2025 May 2]. https://doi.org/10.1021/es201842h.
57. Tang J, Li Y, Zhang L, Mu J, Jiang Y, Fu H, Zhang Y, Cui H, Yu X, Ye Z. 2023. Biosynthetic Pathways and Functions of Indole-3-Acetic Acid in Microorganisms. Microorganisms. 11(8):2077.

doi:10.3390/microorganisms11082077. [accessed 2025 Jun 3]. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10459833/.

58. Terra LA, Klepa MS, Nogueira MA, Hungria M. 2025. Pangenome analysis indicates evolutionary origins and genetic diversity: emphasis on the role of nodulation in symbiotic Bradyrhizobium. Front Plant Sci. 16:1539151. doi:10.3389/fpls.2025.1539151. [accessed 2025 Jun 1]. https://www.frontiersin.org/journals/plant-science/articles/10.3389/fpls.2025.1539151/full.
59. Toesch M, Schober M, Faber K. 2014. Microbial alkyl- and aryl-sulfatases: mechanism, occurrence, screening and stereoselectivities. Appl Microbiol Biotechnol. 98(4):1485–1496. doi:10.1007/s00253-013-5438-0. [accessed 2025 May 31]. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3920027/.
60. 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(2):1956–1969. doi:10.1016/j.ygeno.2019.11.008. [accessed 2025 May 11]. https://www.sciencedirect.com/science/article/pii/S0888754319303714.
61. Wagstaff BA, Zorzoli A, Dorfmueller HC. 2021. NDP-rhamnose biosynthesis and rhamnosyltransferases: building diverse glycoconjugates in nature. Biochemical Journal. 478(4):685–701. doi:10.1042/BCJ20200505. [accessed 2025 Jun 3]. https://doi.org/10.1042/BCJ20200505.
62. 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. doi:10.3389/fmicb.2018.02238. [accessed 2025 Jun 13]. https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2018.02238/full.
63. Wang Y, Zhang C, Liu G, Ju J, Yu B, Wang L. 2019. Elucidating the Role and Regulation of a Lactate Permease as Lactate Transporter in Bacillus coagulans DSM1. Appl Environ Microbiol. 85(14):e00672-19. doi:10.1128/AEM.00672-19. [accessed 2025 May 31]. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6606885/.
64. Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. PLOS Computational Biology. 13(6):e1005595. doi:10.1371/journal.pcbi.1005595. [accessed 2024 Feb 28]. https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1005595.
65. Wu D, Seshadri R, Kyrpides NC, Ivanova NN. 2025. A metagenomic perspective on the microbial prokaryotic genome census. Science Advances. 11(3):eadq2166. doi:10.1126/sciadv.adq2166. [accessed 2025 May 29]. https://www.science.org/doi/10.1126/sciadv.adq2166.
66. 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. Microbiology and Immunology. 34(2):99–119.

doi:10.1111/j.1348-0421.1990.tb00996.x. [accessed 2024 Apr 7]. https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1348-0421.1990.tb00996.x.

67. 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(9):635–645. doi:10.1038/nrmicro3330. [accessed 2025 May 31]. https://www.nature.com/articles/nrmicro3330.

# Figures

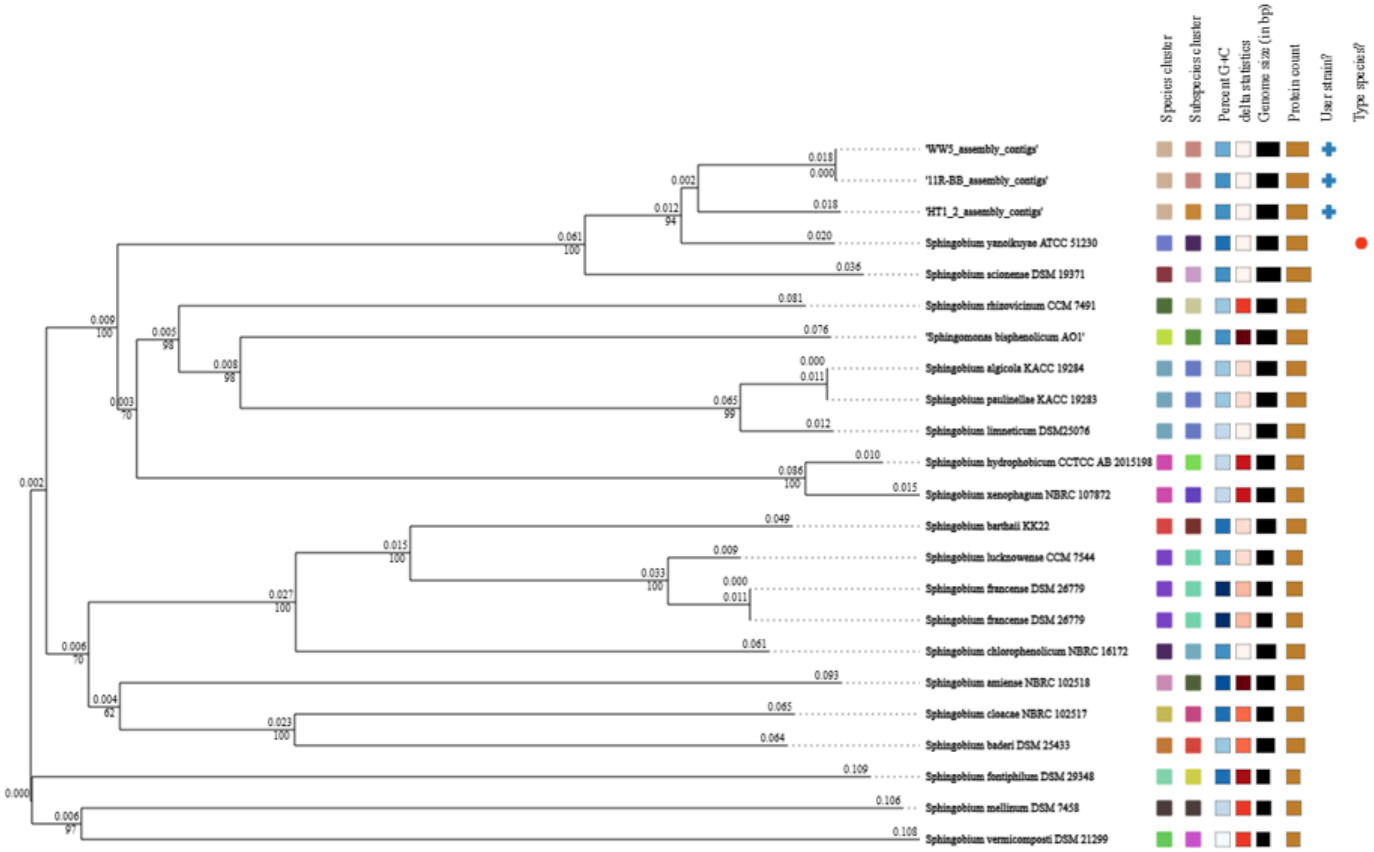


**Figure 1**

Genome-based phylogenetic delineation of Sphingobium salicis sp. nov. within the genus Sphingobium. GBDP-based phylogeny (TYGS platform, formula d5, 100 bootstraps) showing evolutionary relationships among Sphingobium strains. WW5, 11R-BB, and HT1-2 form a distinct clade (bootstrap support >94%) separate from S. yanoikuyae ATCC 51230. Colored bars indicate taxonomic, genomic, and genetic features. Blue plus signs: analyzed strains; red dot: S. yanoikuyae type strain.

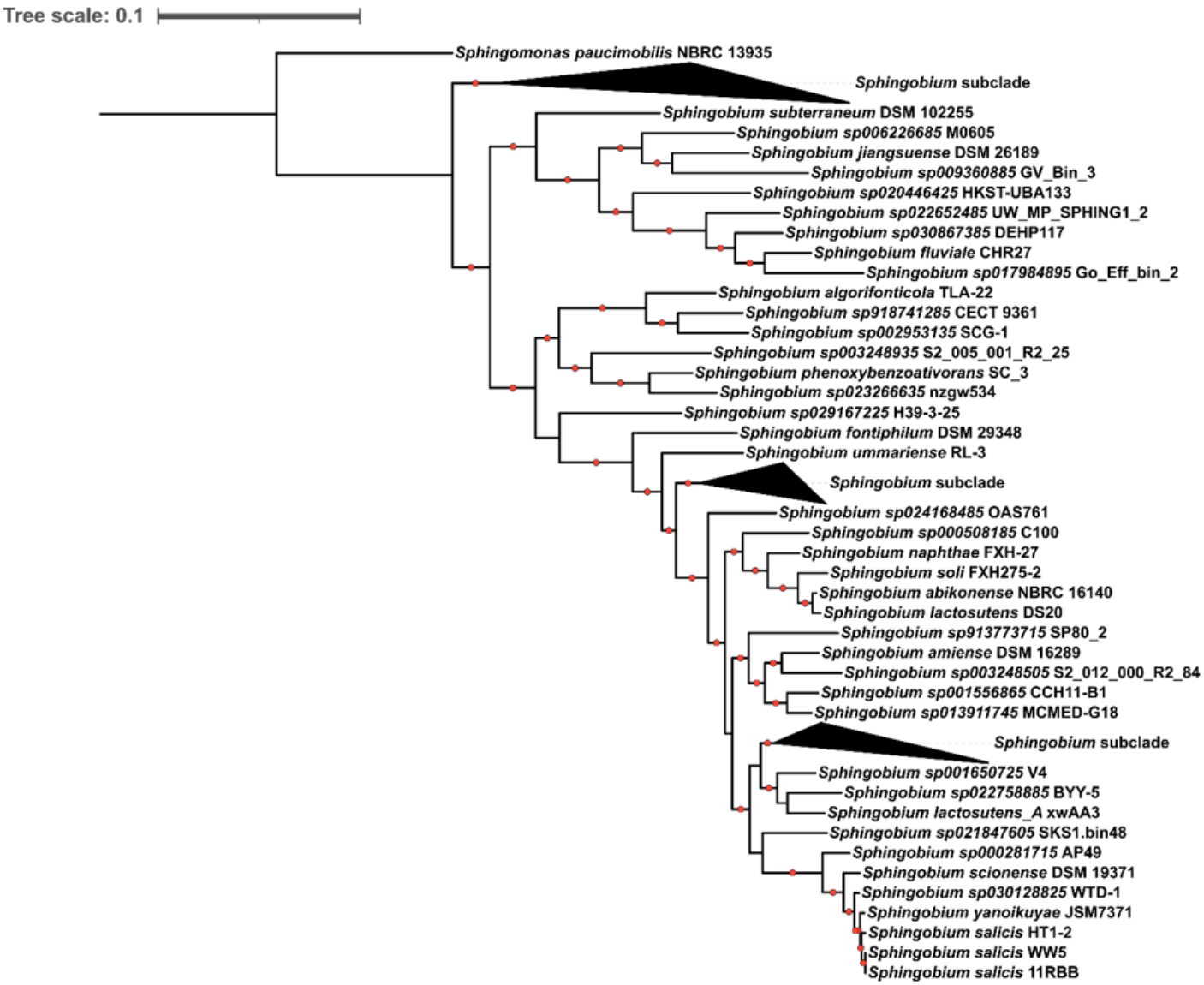


**Figure 2**

A maximum likelihood phylogenetic consensus tree was constructed based on 120 ubiquitous single-copy marker genes (bacterial 120 marker set; Parks et al. 2018). Phylogenetic distances were calculated using the Whelan and Goldman (WAG) protein substitution model. Bootstrap values were calculated from 1000 replications, and only nodes (indicated by red dots) with support values above 80% were displayed. Branches containing more than 10 representative genomes were collapsed into subclades for visual clarity. *Sphingomonas paucimobilis* NBRC 13935 was used as the outgroup.

# Supplementary Files

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

- TournaySupplementalDocs.zip