Pseudogemmobacter sonorensis sp. nov., a new alphaproteobacterium isolated from the slime flux of a tree (*Populus fremontii*) in the Sonoran Desert (Arizona, USA)

Nóra Tünde Lange-Enyedi[1,2], Erika Tóth[1,3], Gorkhmaz Abbaszade[1,4], Péter Németh[5,6], Laurence A. J. Garvie[7], Jacqueline Wolf[8], Meina Neumann-Schaal[8], Bernadett Khayer[9], György Sipos[2], Judit Makk[1]

[1]Department of Microbiology, Faculty of Science, Eötvös Loránd University, Pázmány Péter stny, 1/C H-1117 Budapest, Hungary

[2]Functional Genomics and Bioinformatics Group, Faculty of Forestry, University of Sopron, Bajcsy-Zsilinszky út 4, H-9400, Sopron, Hungary

[3]Health Promotion and Education Research Team, Hungarian Academy of Sciences, Budapest, Hungary

[4]Department of Applied Microbial Ecology, Helmholtz Centre for Environmental Research – UFZ, Permoserstr. 15, 04318 Leipzig, Germany

[5]Institute for Geological and Geochemical Research, HUN-REN Research Centre for Astronomy and Earth Sciences, Budaörsi út 45 H-1112 Budapest, Hungary

[6]Research Institute of Biomolecular and Chemical Engineering, Nanolab, University of Pannonia, Egyetem út 10 H-8200 Veszprém, Hungary

[7]School of Earth and Space Exploration, Arizona State University, 781 East Terrace Road, Tempe, AZ 85287-6004, USA

[8]Leibniz Institute DSMZ – German Collection of Microorganisms and Cell Cultures, Inhoffenstrasse 7B, D-38124, Braunschweig, Germany

[9]Department of Public Health Laboratory and Methodology, National Center Public Health and Pharmacy, Albert Flórián út 2-6 H-1097 Budapest, Hungary

Correspondence: Judit Makk, Tel.: +36-1-381-2177, Fax.: +36-1-381-2178, e-mail: makk.judit@ttk.elte.hu

Running title: *Pseudogemmobacter sonorensis* sp. nov.

The subject category for the Contents list: New Taxa - Pseudomonadota

Keywords: *Pseudogemmobacter*; *Populus fremontii*; slime flux, Sonoran Desert

Depositories: The GenBank/EMBL/DDBJ accession number for the 16S rRNA gene sequence of strain PA1-206B^T is OP709268. The whole-genome shotgun project of the strain PA1-206B^T has been deposited at DDBJ/ENA/GenBank under the accession JBEFZE000000000. The version described in this paper is JBEFZE010000000.

Bacteria; Pseudomonadota; Alphaproteobacteria; Rhodobacterales; Paracoccaceae

Abstract

The Gram-stain negative, aerobic, non-motile, non-spore-forming, irregular rod-shaped bacterial strain, PA1-206B^T was isolated from tree wound exudate of the *Populus fremontii* trunk in the Sonoran Desert (USA), and its taxonomic position was investigated by a polyphasic approach. Strain PA1-206B^T grows optimally at 28-30 ºC and pH 6 to 10 without NaCl. Based on 16S rRNA gene sequence analysis, this isolate showed only 95.9 % sequence similarity to the type strain of *Pseudogemmobacter hezensis* and similarity of 93.9-95.4 % to

the other species of the genera *Pseudogemmobacter*, *Falsigemmobacter*, *Rhodobacter*, *Neotabrizicola Paracoccus*, *Xinfangfangia*, *Fuscovulum*, *Pseudotabrizicola*, and *Tabrizicola*. The major isoprenoid quinone of the strain was ubiquinone Q-10. The predominant fatty acids (> 5%) were $C_{18:1}$ ω7c, $C_{16:0}$ and 11-methyl-$C_{18:1}$ ω7c. Diphosphatidylglycerol, phosphatidylglycerol, phosphatidylethanolamine, an unidentified aminolipid, and two unidentified phospholipids were present. The assembled draft genome of strain PA1-206B^T had 115 contigs with a total length of 4.5 Mb and a G+C content of 67.4 mol%. The overall genome-related indices (ANI <80.4 %, AAI <70.6 %, dDDH <21.6 %) with respect to close relatives were below the corresponding threshold to demarcate bacterial species and nearly reached the genus level but the whole genome-based taxonomic approaches rejected the description of a new genus. Strain PA1-206B^T (=DSM 115559^T =NCAIM B.02680^T) is suggested as the type strain of a novel species of the genus *Pseudogemmobacter*, for which the name *Pseudogemmobacter sonorensis* sp. nov. is proposed.

Introduction

The family *Paracoccaceae* (illegitimate synonym "Rhodobacteraceae") belonging to the order *Rhodobacterales*, class *Alphaproteobacteria* within the *Pseudomonadota*, comprises 78 taxa with validly published names at the time of writing (see List of Prokaryotic Names with Standing in Nomenclature; https://lpsn.dsmz.de/family/paracoccaceae). The taxonomical lineage of the family was revised numerous times recently [1–3]. The family demonstrates adaptability to a wide range of environmental conditions, reflecting their diverse ecological roles and metabolic capabilities, and the members of the family have been found in various marine and non-marine environments, including seawater, marine animal tissues, seaweeds, soil, rhizosphere, activated sludge, and a soda lake [4–10]. Among the described species,

Sinorhodobacter populi and *Pseudogemmobacter hezensis* have been described from the wound exudates of *Populus* trees [1, 11]. In this study, a novel species of this family is described based on a polyphasic characterization for a bacterial strain, PA1-206B^T, isolated from a tree (*Populus fremontii*) wound in the Sonoran Desert.

Strain PA1-206B^T was isolated from a slime flux collected from mineralized tree wounds of *P. fremontii* in south central Arizona along the Verde River around 33°44′33.60″N, 111°39′33.97″W [12]. *Populus fremontii* is a native tree found in the southwestern USA and northern to central Mexico. Wetwood occurs in the heartwood of poplar trees (including *P. fremontii*) and is characterized by dark brown staining and infusion. When infected with bacteria, yeast, and other fungi, the resulting exudate forms a smelly, foaming mass known as slime flux [13, 14]. The slime flux jelly on the tree wound is likely a bacterial exopolysaccharide (EPS), consistent with the diverse prokaryotic community identified from wetwood on *P. fremontii* and Balsamifera trees in California [15]. Wetwood is a natural occurrence and is not considered detrimental to the health of the tree, although its presence may contribute to the general decline of trees, particularly old trees, and those with low vitality [16], therefore describing and analysing bacteria from slime flux is influential.

Methods

The jelly-like slime flux sample was scraped from the surface of a *Populus* tree using a sterile spatula [12]. The collected sample was mashed in 4.5 ml physiological saline solution in a sterile marble mortar. A dilution series of the resulting solution was spread on R2A agar medium (DSMZ medium 830, https://mediadive.dsmz.de/) using the dilution-plating method.

Strain PA1-206B^T was isolated from the diluted samples on R2A agar plates with pH 7.0 after incubation at 28 °C for 5 days. The strain was maintained on R2A agar medium.

Genomic DNA was isolated from strain PA1-206B^T according to the methods of Enyedi et al. [17] and the 16S rRNA gene was amplified with primers 27F and 1492R as previously described by Makk et al. [18]. Purification and sequencing of PCR products were carried out by Eurofins BIOMI (Gödöllő, Hungary). A nearly full-length sequence of the 16S rRNA gene (1333 bp) was compiled using the MEGA7 software [19]. The GenBank/EMBL/DDBJ accession number for the 16S rRNA gene sequence of strain PA1-206B^T is OP709268. The 16S rRNA gene sequence of strain PA1-206B^T was aligned with sequences available from the EzBioCloud database [20]. The alignment was used to calculate the distance matrix corrected by the Tamura 3-parameter model and a phylogenetic tree was reconstructed using the maximum likelihood method. A discrete Gamma distribution was used to model evolutionary rate differences among sites. The rate-variation model allowed for some sites to be evolutionarily invariable. The phylogenetic tree was constructed with MEGA7, and the result of the model test based on the maximum likelihood fits of 24 different nucleotide substitution models was applied. Bootstrap analysis was based on 500 resamplings [21].

Full genome sequencing of strain PA1-206B^T was performed as a service provided by Eurofins BIOMI (Gödöllő, Hungary). Genomic DNA was extracted using the NucleoSpin Microbial DNA Kit (Macherey-Nagel). The library was prepared using the Nextera DNA Flex Library Preparation Kit (Illumina). The quality of the library was checked using Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific), and Agilent 4150 TapeStation D1000 ScreenTape System (Agilent). 2x 250 bp paired-end sequencing was carried out on the Illumina Miseq System using an Illumina v2 500 cycle sequencing kit. The sequence read quality was checked by FastQC [22], and low-quality read sequences were trimmed by trimmomatic v0.39 tools [23]. *De novo* assembly of raw reads was performed using the SPAdes v.3.15.5

tool in careful mode [24], and the quality of the resulting assembly (assembly size, number of contigs, N50, L50) was assessed by QUAST 5.2 [25]. GC content was calculated by BBmap 39.11. package [26]. Contigs shorter than 500 nt were excluded from the assembly. Coverage was calculated with coverage calculator master v0.0.1 (https://github.com/GenomicaMicrob/coverage_calculator). Possible contamination of the genome was checked based on the 16S rRNA gene data on the ContEst16S platform [27] and no contamination was detected. The partial 16S rRNA gene sequence of strain PA1-206B^T obtained by the Sanger method was compared with the extracted 16S rRNA gene sequence from the genome assembly by the Pairwise nucleotide sequence alignment for taxonomy tool (https://www.ezbiocloud.net/tools/pairAlign). Raw sequencing reads were deposited in the NCBI Sequence Read Archive (SRA) under the BioProject ID PRJNA1120638 and accession number SRR29909998. The whole-genome shotgun project has been deposited at DDBJ/ENA/GenBank under accession JBEFZE000000000. The version described in this paper is JBEFZE010000000.

Overall genomic relatedness indices (OGRIs) values were calculated between the genome sequence of PA1-206B^T and the available reference genomes of its relatives (*Pseudogemmobacter hezensis* D13-10-4-6^T, *Pseudogemmobacter bohemicus* Cd-10^T, *Falsigemmobacter intermedius* 119/4^T, *Pseudogemmobacter humi* CIP 111625^T, *Xinfangfangia soli* CCTCC AB 2017177^T, *Tabrizicola aquatica* RCRI19^T, *Falsirhodobacter deserti* W402^T, *Paracoccus limosus* JCM 17370^T, *Paracoccus sanguinis* DSM 29303^T) downloaded from the NCBI database (results shown in Table 1). Average nucleotide identity (ANI) and average amino acid identity (AAI) were calculated using the Enveomics online tools (http://enve-omics.ce.gatech.edu/) [28]. For AAI, the whole genome of PA1-206B^T was annotated with GeneMarkS (http://exon.gatech.edu/genemark/genemarks.cgi) [29]. The Type Strain Genome Server (https://tygs.dsmz.de/) [30] was used for digital DNA–DNA

hybridization (dDDH), G+C difference calculation and to construct phylogenomic tree by Genome Blast Distance Phylogeny (GBDP) approach [31] based on whole genome sequences of closely related strains identified by the EzBioCloud database. The resulting intergenomic distances were used to infer a balanced minimum evolution tree with branch support via FASTME 2.1.6.1 including SPR postprocessing [32]. The proteome tree was constructed from whole-proteome-based GBDP distances. The branch lengths are scaled via GBDP distance formula d_5. Branch support was inferred from 100 pseudobootstrap replicates each. The trees were rooted at the midpoint [33].

To accurately establish the taxonomic classification of the genome, the phylogenomic tree was constructed using the Genome Taxonomic Database Toolkit 2.4.0 (GTDB-Tk) [34] based on 120 ubiquitously conserved bacterial marker genes by Kbase server [35]. The trees were visualized by the Interactive Tree of Life (iTol) tool (https://itol.embl.de/) [36] and PhyD3 [37].

For genus delineation, the AAI cut off value within *Paracoccaceae* family members were identified. In total, 4033 existing genomes of the *Paracoccaceae* family were downloaded from NCBI genome database (accessed on 01.11.2024). Genome quality was assessed by CheckM v1.2.3 (https://github.com/Ecogenomics/CheckM) [38] and after removing the genomes with less than 90 % completeness and more than 5 % contamination, remaining 2049 genomes were used for the determination of AAI cut off value by using CompareM v0.1.2 tool (https://github.com/Ecogenomics/CheckMhttps://github.com/donovan-h-parks/CompareM), which uses the mean amino acid identity of orthologous genes between a given pair of genomes.

The draft genome was analysed for prediction of genome features through the Rast server version 2.0 (http://rast.nmpdr.org/) [39] and the DDBJ Fast Annotation and Submission Tool (DFAST) [40]. After submission, the NCBI Prokaryotic Genome Annotation Pipeline (PGAP)

(https://www.ncbi.nlm.nih.gov/genome/annotation_prok/) also annotated the genome. Functional genes that were investigated as having possible roles in metabolic pathways were checked by the KEGG database [41]. A circular graphical display of the genome and applicable genes was prepared by the Proksee circular genome visualization tool with the Prokka annotation tool [42].

In this study, the type strains *Pseudogemmobacter hezensis* KCTC 82215^{T}, *Falsigemmobacter intermedius* DSM 28642^{T,} and *Tabrizicola aquatica* JCM 17277^{T} were used as reference strains for comparison of phenotypic properties of strain PA1-206B^{T} with several genera under the same laboratory conditions. The colony morphology of strain PA1-206B^{T} was observed on R2A agar medium (pH 7) after incubation at 28 °C for 5 days by direct observation of single colonies (Supplementary Fig. S1). Growth of the strain was tested on TSA (DSMZ Medium 535) and Nutrient agar (DSMZ Medium 1). The morphology (size, shape, arrangement) of the cells grown on R2A medium (pH 7) at 28 °C for 3 days on R2A agar was studied in native preparations and after Gram-staining according to Claus [43] by Nikon light microscopy. Electron microscopic investigations were carried out as indicated by Tóth et al. [44] using a JEOL JSM-IT700HR scanning electron microscope (Supplementary Fig. S1). Growth under anaerobic conditions, catalase and oxidase activity, oxidative and fermentative degradation of glucose, aerobic nitrate reduction, indole and H_2S production, casease, urease, phosphatase activity, and hydrolysis of gelatine, starch, Tween 80 and aesculin were studied as described by Makk et al. [18]. Temperature and pH optima as well as salt tolerance were determined based on the observed growth intensity at 4, 10, 15, 20, 28, 30, 35, 40, and 45 °C, at pH from 3 to 13 (1 pH unit intervals), and from 0 to 5 % (w/v) NaCl concentration (0.5 % intervals), as described previously [18] (in R2A medium). Motility was assessed by phase-contrast microscopy in wet-mount preparations and in deep R2A medium with 0.3 % (w/v) agar, respectively. Acid production from different carbon sources as the sole

source of carbon and enzyme activities were tested by using the API 50CH and API ZYM systems (bioMérieux) according to the manufacturer's instructions, except that the strips were incubated for 5 h (API ZYM) and for 24–72 h (API 50CH).

Chemotaxonomic analyses of isoprenoid quinones, cellular fatty acids, and polar lipids were performed as previously described in detail [45].

Results

The closest relatives of strain PA1-206B^T based on 16S rRNA gene pairwise sequence similarity values are as follows: *Pseudogemmobacter hezensis* D13-10-4-6^T, 95.9 %; *Tabrizicola fusiformis* SY72^T, 95.5 %; *Pseudogemmobacter humi* IMT-291^T, 95.4 %; *Neotabrizicola shimadae* N10^T, 95.3 %; *Pseudogemmobacter bohemicus* Cd-10^T, 95.1 % and *Rhodobacter ruber* CCP-1^T, 95.0 %. PA1-206B^T showed lower similarity to other species than the generally accepted threshold of 98.7 %. Other taxa of the family *Paracoccaceae* were even more distantly related (with less than 94 % sequence similarity), and they were placed on separated branches on the phylogenetic tree of 16S rRNA-gene sequences (Fig. 1), except for *Falsigemmobacter intermedius* 119/4^T with low bootstrap value that showed only 94.1 % sequence similarity. Nevertheless, previous studies also demonstrated that 16S rRNA gene sequences are not eligible to determine the taxonomical outline of the family *Paracoccaceae* [1–3].

The assembled genome of strain PA1-206B^T had a total length of 4.5 Mb, 115 contigs with 129.4× average coverage depth, N50 and L50 values of 125071 nt and 11 nt, respectively. The DNA G+C content of the strain PA1-206B^T was 67.4 mol% (± 6 mol% standard deviation) as determined by the draft genome sequence. The partial 16S rRNA gene sequence of strain PA1-206B^T obtained by the Sanger method was compared with the extracted 16S rRNA gene sequence from the genome assembly and showed 100 % similarity.

OGRI values between the genome of strain PA1-206B^T and the available genomes of close relative species were calculated (Table 1). ANI values were well below the species threshold of 95 %. Based on current AAI genus delineation cutoff of 70% proposed for the *Paracoccaceae* family (70%) by Huang et al. [3], the genome of strain PA1-206B^T belongs to a new genus. However, objective taxonomic classification of the genome, conducted using the Genome Database Taxonomy toolkit (GTDB-Tk, database release 202), considering 120 ubiquitous single-copy proteins from the family *Paracoccaceae*, placed the strain PA1-206B^T within the genus *Pseudogemmobacter* (Supplementary Fig. S2*)*. In this case, the 70% genus delineation threshold does not adequately reflect the observed AAI value between the genome of PA1-206B^T and other species within the genus *Pseudogemmobacter*, suggesting that this threshold may not be appropriate for defining genus-level relationships in this context (see Table 1). Consequently, this study proposes a new genus delineation AAI cutoff of 68% for the *Paracoccaceae* family, based on AAI analyses of 2049 high-quality genomes within the family. This threshold lies within typical genus-level cutoffs for many groups in *Paracoccaceae* family (Table 1, Supplementary Fig. S3). Digital DNA–DNA hybridization (dDDH) values of strain PA1-206B^T were far below 70 % based on the calculations of the Type Strain Genome Server. G+C difference was above 1 mol% except for *Xinfangfangia soli* CCTCC AB 2017177^T, *Pseudogemmobacter humi* CIP 111625^T, and *Tabrizicola aquatica* RCRI19^T (Table 1) but these strains showed low AAI, ANI and dDDH similarities. The ANI, AAI, dDDH, and G+C difference values suggest that strain PA1-206B^T represents a new member of the genus *Pseudogemmobacter*.

The phylogenetic position of strain PA1-206B^T within the genus *Pseudogemmobacter* was confirmed by the results of the phylogenomic analysis based on the whole-proteome-based tree (Fig. 2). Strain PA1-206B^T and other *Pseudogemmobacter* spp. showed a separated branch from the genera *Falsigemmobacter, Rhodobacter, Neotabrizicola, Paracoccus,*

Xinfangfangia, *Fuscovulum*, *Pseudotabrizicola* and *Tabrizicola* on the whole-genome-based tree as well (Fig. 2, Supplementary Fig. S4).

For the annotation of the draft genome, 4594 coding sequences (CODs) and 318 subsystems were identified based on RAST analysis. Prokka annotation showed 4250 clusters of orthologous genes (COGs) and 1972 hypothetical proteins. DFAST analysis revealed 3 rRNAs (5S, 16S, and 23S rRNAs) and 55 tRNAs within the genome (Fig. 3). Based on RAST analysis, 300 genes were involved in the metabolism of amino acids and derivatives, and 178 genes were associated with the metabolism of proteins including gelatinase. The draft genome of strain PA1-206B^T revealed 196 gene clusters in the metabolism of carbohydrates, for example, D-glucose, chitin, and N-acetylglucosamine. 122 genes were identified that were involved in the metabolism of cofactors, vitamins, and prosthetic groups (biotin, thiamine, pyridoxine, folate) but genes were not identified in carotenoid, bacteriochlorophyll synthesis, and photosynthesis. Among them, 121 genes were affiliated with respiration processes, and 109 genes were linked with the metabolism of nucleosides and nucleotides. Genes encoding proteins involved in the metabolism of CO_2 fixation were not found, and autotrophic metabolism was not suggested. The number of genes involved in flagellar motility was only 5, and genes related to nitrogen fixation were not prevalent in the draft genome.

Regarding the isolation source, auxin biosynthesis and the degradation of aromatic compounds (salicylate, benzoate, and catechol) were found in the gene clusters, that can enhance bacterial metabolism on the surface of a plant with lignin-containing cell wall. The draft genome of strain PA1-206B^T contains more genes involved in these pathways than its closest relative strain, *Pseudogemmobacter hezensis* D13-10-4-6^T (Supplementary Table S1). The genes (23 by number) of diverse membrane transport systems were identified in the draft genome, that enable to maintain the transport of ions, the pH and osmolarity homeostasis in the slime flux [12], such as the magnesium and cobalt efflux protein, the Na^+/H^+ antiporter

and the copper-translocating P-type ATPase, alkaline phosphatase. Similarly, genes linked to potassium metabolism including the potassium efflux system KefA protein, large-conductance mechanosensitive channel, and the Kup system potassium uptake protein were identified in the genome. Branched-chain amino acid transporters were also present that convert into negatively charged glutamic acid [46]. The genome contains several gene clusters against osmotic and oxidative stress.

RAST functional analysis revealed that the genome of the most closely related type strain *Pseudogemmobacter hezensis* D13-10-4-6^T contains the same gene subsystems but there were quantitative differences in the gene sets. Among all, 1101 gene functions were common, 148 were identified only in strain PA1-206B^T, and 121 were identified only in *Pseudogemmobacter hezensis* D13-10-4-6^T (Supplementary Table S1).

Differential phenotypic and genotypic characteristics of strain PA1-206B^T and its closest relatives are given in Table 2, Supplementary Fig. S1, S5 and Tables S2-S4. The cells of strain PA1-206B^T are immobile irregular rods, and they produce transparent colonies in accordance to the genomic data. Strain PA1-206B^T did not grow at elevated concentrations of NaCl than other members of the genera *Pseudogemmobacter*, *Falsigemmobacter*, and *Tabrizicola*. The hydrolysis of gelatine, acid production from D-sucrose and D-trehalose, and the presence of alkaline phosphatase distinguished strain PA1-206B^T from its closest relatives.

The predominant quinone of strain PA1-206B^T was ubiquinone Q-10 accompanied by Q-9. The major cellular fatty acids present in strain PA1-206B^T were $C_{18:1}$ ω7c, $C_{16:0}$, and 11-methyl-$C_{18:1}$ ω7c (Supplementary Table S2), which were partially similar to the fatty acid profile of the closest related type strains, and with significant differences in their quantities. Diphosphatidylglycerol (DPG), phosphatidylglycerol (PG), phosphatidylethanolamine (PE), an unidentified aminolipid (AL), and two unidentified phospholipids (PLs) were the major

polar lipids of PA1-206B^T (Supplementary Fig. S5). PA1-206B^T differed from its closest relatives due to the absence of phosphatidylmonomethylethanolamine (PMME) and phosphatidylcholine (Table 2).

On the basis of differences in the phenotypic (hydrolysis of gelatine, acid production from D-sucrose and D-trehalose, and the presence of alkaline phosphatase), chemotaxonomic, phylogenetic (low sequence similarity in the 16S rRNA gene, distinct position within the phylogenetic trees) and phylogenomic data (low values of overall genome-related indices) presented here, strain PA1-206B^T represents a novel species in the genus *Pseudogemmobacter*, for which the name *Pseudogemmobacter sonorensis* sp. nov. is proposed.

Description of *Pseudogemmobacter sonorensis* sp. nov.

Pseudogemmobacter sonorensis (so.no.ren´sis. N.L. adj. *sonorensis* of the Sonoran, named after the Sonoran Desert, where the organism was collected).

Cells are irregular rod-shaped (0.6–0.7 × 1.0–1.5 μm) and non-motile. Cells of PA1-206B^T occur form pairs or aggregates. It grows well on R2A, Nutrient, and TGY media. Colonies on R2A agar plates are shiny, mucoid, creamy-white translucent, circular, and 2–3 mm in diameter after 3 days of cultivation at 28 °C. Growth occurs at 15–35 °C (optimum at 28-30 °C), pH 6.0–10.0 (optimum at pH 7.0–9.0). The strains could not grow with 1-5 % (w/v) NaCl. Oxidase- and catalase-positive. Negative for Voges-Proskauer test, glucose fermentation, indole production, aerobic nitrite reduction to N_2, phosphatase activity, hydrolysis of urea, aesculine, casein, starch, Tween 80 as well as H_2S production. Positive for hydrolysis of gelatine and weakly positive for aerobic nitrate reduction to nitrite. The API ZYM test showed positive for alkaline phosphatase, esterase (C4), leucine arylamidase, valine

arylamidase, naphthol-AS-BI-phosphohydrolase, α-glucosidase, N-acetyl-β-glucosaminidase, and cells are weakly positive for esterase lipase (C8). Enzyme production was negative for lipase (C14), cystine arylamidase, trypsin, α-chymotrypsin, acid phosphatase, α-galactosidase, β-galactosidase, β-glucuronidase, β-glucosidase, α-mannosidase, and α-fucosidase. Based on API 50CH strips, acid production was observed only from D-glucose, D-fructose, D-sucrose, D-trehalose, and a weak reaction from D-mannitol. The major respiratory quinone is Q-10. The polar lipid profile is composed of DPG, PG, PE as major components, and unidentified AL and two unidentified PL as minor components. The DNA G+C content is 67.4 mol% (calculated from the genome sequence).

The type strain, PA1-206B^T (=DSM 115559^T =NCAIM B.02680^T) was isolated from the wound exudate of a tree (*Populus fremontii*) in the Sonoran Desert (USA). The GenBank accession number for the 16S rRNA gene sequence of strain PA1-206B^T is OP709268. The Whole Genome Shotgun project of strain PA1-206B^T has been deposited at DDBJ/ENA/GenBank under accession JBEFZE000000000. The version described in this paper is version JBEFZE010000000.

Acknowledgements

This work was supported by the Ministry of Innovation and Technology of Hungary from the National Research, Development and Innovation Fund [grant number ANN141894]. The study was funded by the Scientific Foundations of Education Research Program of the Hungarian Academy of Sciences. We thank Gesa Martens, Birgit Grün and Anika Wasner for excellent technical assistance.

Conflicts of interest

The authors declare that there are no conflicts of interest.

References

1. **Ma T, Xue H, Piao C, Liu C, Yang M, *et al.*** Reclassification of 11 members of the family Rhodobacteraceae at genus and species levels and proposal of *Pseudogemmobacter hezensis* sp. nov. *Front Microbiol* 2022;13:849695.
2. **Zhang DF, He W, Shao Z, Ahmed I, Zhang Y, *et al.*** Phylotaxonomic assessment based on four core gene sets and proposal of a genus definition among the families Paracoccaceae and Roseobacteraceae. *Int J Syst Evol Microbiol* 2023;73:006156.
3. **Huang Z, Li M, Oren A, Lai Q**. Genome-based analysis of the family Paracoccaceae and description of *Ostreiculturibacter nitratireducens* gen. nov., sp. nov., isolated from an oyster farm on a tidal flat. *Front Microbiol* 2024;15:1376777.
4. **Nokhal TH, Schlegel HG**. Taxonomic study of *Paracoccus denitrificans*. *Int J Syst Bacteriol* 1983;33:26–37.
5. **Szuróczki S, Abbaszade G, Bóka K, Schumann P, Neumann-Schaal M, *et al.*** *Szabonella alba* gen. nov., sp. nov., a motile alkaliphilic bacterium of the family Rhodobacteraceae isolated from a soda lake. *Int J Syst Evol Microbiol* 2022;72:005219.
6. **Maszenan AM, Seviour RJ, Patel BKC, Rees GN, McDougall BM**. *Amaricoccus* gen. nov., a Gram-negative coccus occurring in regular packages or tetrads,isolated from activated sludge biomass, and descriptions of *Amaricoccus veronensis* sp. nov., *Amaricoccus tamworthensis* sp. nov., *Amaricoccus macauensis* sp. nov., and *Amaricoccus kaplicensis* sp. nov. *Int J Syst Bacteriol* 1997;47:727–734.
7. **Nedashkovskaya OI, Otstavnykh NY, Gun S, Andrey K, Zhukova N V, *et al.*** Algicella marina gen. nov., sp. nov., a novel marine bacterium isolated from a Pacific red alga. *Arch Microbiol* 2022;204:487.
8. **Chhetri G, Kang M, Kim J, Kim I, So Y, *et al.*** *Fuscibacter oryzae* gen. nov., sp. nov., a phosphate-solubilizing bacterium isolated from the rhizosphere of rice plant. *Antonie van Leeuwenhoek, Int J Gen Mol Microbiol* 2021;114:1453–1463.

9. **Li F, Huang Y, Hu W, Li Z, Wang Q, *et al.*** *Mesobaculum littorinae* gen. Nov., sp. nov., a novel bacterium isolated from a sea snail littorina scabra. *Int J Syst Evol Microbiol* 2021;71:004821.

10. **Simon M, Scheuner C, Meier-Kolthoff JP, Brinkhoff T, Wagner-Döbler I, *et al.*** Phylogenomics of Rhodobacteraceae reveals evolutionary adaptation to marine and non-marine habitats. *ISME J* 2017;11:1483–1499.

11. **Xu GT, Piao C, Chang JP, Guo LM, Yang XQ, *et al.*** *Sinorhodobacter populi* sp. Nov., isolated from the symptomatic bark tissue of *Populus* x *euramericana* canker. *Int J Syst Evol Microbiol* 2019;69:1220–1224.

12. **Garvie LAJ**. Seasonal formation of ikaite in slime flux jelly on an infected tree (*Populus fremontii*) wound from the Sonoran Desert. *Sci Nat* 2022;109:48.

13. **Hofstra TS, Stromberg JC, Stutz JC**. Factors associated with wetwood intensity of *Populus fremontii* (Fremont cottonwood) in Arizona. *Gt Basin Nat* 1999;59:85–91.

14. **Roberts H, Hsiang T**. Incidence of slime flux in deciduous trees of Southern Ontario. *For Chron* 2018;94:151–154.

15. **Hofstra T**. *Distribution, genetic diversity and compositional variation of wetwood prokaryote communities within and among two native California Populus species*. University of California, Santa Cruz; 2003.

16. **Hamilton WD**. Wetwood and slime flux in landscape trees. *J Arboric* 1980;6:247–249.

17. **Enyedi NT, Anda D, Borsodi AK, Szabó A, Pál SE, *et al.*** Radioactive environment adapted bacterial communities constituting the biofilms of hydrothermal spring caves (Budapest, Hungary). *J Environ Radioact* 2019;203:8–17.

18. **Makk J, Enyedi NT, Tóth E, Anda D, Szabó A, *et al.*** *Deinococcus fonticola* sp. nov., isolated from a radioactive thermal spring in Hungary. *Int J Syst Evol Microbiol* 2019;69:1724–1730.

19. **Kumar S, Stecher G, Tamura K**. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Mol Biol Evol* 2016;33:1870–1874.

20. **Chalita M, Kim YO, Park S, Oh HS, Cho JH, *et al.*** EzBioCloud: a genome-driven database

and platform for microbiome identification and discovery. *Int J Syst Evol Microbiol* 2024;74:006421.

21. **Tamura K**. Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G+C-content biases. *Mol Biol Evol* 1992;9:678–687.

22. **Andrews S**. *FastQC: a quality control tool for high throughput sequence data*. Cambridge; 2010.

23. **Bolger AM, Lohse M, Usadel B**. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 2014;30:2114–2120.

24. **Prjibelski A, Antipov D, Meleshko D, Lapidus A, Korobeynikov A**. Using SPAdes de novo assembler. *Curr Protoc Bioinforma* 2020;70:e102.

25. **Mikheenko A, Prjibelski A, Saveliev V, Antipov D, Gurevich A**. Versatile genome assembly evaluation with QUAST-LG. *Bioinformatics* 2018;34:i142–i150.

26. **Bushnell B, Rood J, Singer E**. BBMerge – Accurate paired shotgun read merging via overlap. *PLoS One* 2017;12:1–15.

27. **Lee I, Chalita M, Ha SM, Na SI, Yoon SH, *et al.*** ContEst16S: An algorithm that identifies contaminated prokaryotic genomes using 16S RNA gene sequences. *Int J Syst Evol Microbiol* 2017;67:2053–2057.

28. **Rodriguez-R LM, Konstantinidis KT**. The enveomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes. *PeerJ Prepr* 2016;4:e1900v1.

29. **Besemer J, Lomsadze A, Borodovsky M**. GeneMarkS: A self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. *Nucleic Acids Res* 2001;29:2607–2618.

30. **Meier-Kolthoff JP, Göker M**. TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy. *Nat Commun* 2019;10:2182.

31. **Meier-Kolthoff JP, Auch AF, Klenk HP, Göker M**. Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics*;14. Epub ahead of print 2013. DOI: 10.1186/1471-2105-14-60.

32. **Lefort V, Desper R, Gascuel O**. FastME 2.0: A comprehensive, accurate, and fast distance-

based phylogeny inference program. *Mol Biol Evol* 2015;32:2798–2800.

33. **Farris JS**. Estimating phylogenetic trees from distance matrices. *Am Nat* 1972;106:645–667.

34. **Chaumeil PA, Mussig AJ, Hugenholtz P, Parks DH**. GTDB-Tk: A toolkit to classify genomes with the genome taxonomy database. *Bioinformatics* 2020;36:1925–1927.

35. **Arkin AP, Cottingham RW, Henry CS, Harris NL, Stevens RL, *et al.*** KBase: The United States department of energy systems biology knowledgebase. *Nat Biotechnol* 2018;36:566–569.

36. **Letunic I, Bork P**. Interactive tree of life (iTOL) v5: An online tool for phylogenetic tree display and annotation. *Nucleic Acids Res* 2021;49:W293–W296.

37. **Kreft L, Botzki A, Coppens F, Vandepoele K, Van Bel M**. PhyD3: A phylogenetic tree viewer with extended phyloXML support for functional genomics data visualization. *Bioinformatics* 2017;33:2946–2947.

38. **Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW**. CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* 2015;25:1043–1055.

39. **Aziz RK, Bartels D, Best A, DeJongh M, Disz T, *et al.*** The RAST Server: Rapid annotations using subsystems technology. *BMC Genomics* 2008;9:75.

40. **Tanizawa Y, Fujisawa T, Nakamura Y**. DFAST: A flexible prokaryotic genome annotation pipeline for faster genome publication. *Bioinformatics* 2018;34:1037–1039.

41. **Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K**. KEGG: New perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res* 2017;45:D353–D361.

42. **Grant JR, Enns E, Marinier E, Mandal A, Herman EK, *et al.*** Proksee: In-depth characterization and visualization of bacterial genomes. *Nucleic Acids Res* 2023;51:W484–W492.

43. **Claus D**. A standardized Gram staining procedure. *World J Microbiol Biotechnol* 1992;8:451–452.

44. **Tóth EM, Kéki Z, Bohus V, Borsodi AK, Márialigeti K, *et al.*** *Aquipuribacter hungaricus* gen. nov., sp. nov., an actinobacterium isolated from the ultrapure water system of a power

plant. *Int J Syst Evol Microbiol* 2012;62:556–562.

45. **Toumi M, Whitman WB, Kyrpides NC, Woyke T, Wolf J, *et al.*** *Antiquaquibacter oligotrophicus* gen. nov., sp. nov., a novel oligotrophic bacterium from groundwater. *Int J Syst Evol Microbiol* 2023;73:006205.

46. **Lebre PH, Cowan DA**. Genomics of Alkaliphiles. *Alkaliphiles Biotechnol* 2020;135–155.

47. **Kämpfer P, Jerzak L, Wilharm G, Golke J, Busse HJ, *et al.*** *Gemmobacter intermedius* sp. nov., isolated from a white stork (*Ciconia ciconia*). *Int J Syst Evol Microbiol* 2015;65:778–783.

48. **Tarhriz V, Thiel V, Nematzadeh G, Hejazi MA, Imhoff JF, *et al.*** *Tabrizicola aquatica* gen. nov. sp. nov., a novel alphaproteobacterium isolated from Qurugöl Lake nearby Tabriz city, Iran. *Antonie van Leeuwenhoek, Int J Gen Mol Microbiol* 2013;104:1205–1215.

Legends

Fig. 1. 16S rRNA gene sequence-based maximum likelihood tree showing the phylogenetic positions of strain PA1-206B^T among closely related members of the family *Paracoccaceae*. GenBank accession numbers are given in parentheses. Bootstrap values (above 50 %) based on 500 resamplings are shown at branch nodes (Tamura 1992 [21]). Bar, 0.05 nucleotide substitutions per position.

Fig. 2. Phylogenomic tree indicating the position of strain PA1-206B^T among closely related taxa based on whole-proteome data. GenBank accession numbers are given in parentheses. Tree inferred with FastME 2.1.6.1 (Lefort et al. 2015 [32]) from whole-proteome-based GBDP distances. The branch lengths are scaled via GBDP distance formula d_5. Branch values are GBDP psedo-bootstrap support values > 50 % from 100 replications, with an average branch support of 85.8 %. The tree was midpoint-rooted (Farris, 1972 [33]).

Fig. 3. A circular graphical display of the genome and applicable genes of strain PA1-206B^T. This includes coding sequences (CDS) on the forward strand, CDS on the reverse strand, RNA genes, repeat regions and GC content. The figure was prepared by Proksee circular genome visualization tool (Grant et al. 2023 [42]).

Table 1. Average nucleotide identity (ANI), average amino acid identity (AAI), digital DNA–DNA hybridization (dDDH) and G+C difference values between the genome sequence of PAI-206B^T and the available reference genomes of its relatives: *Pseudogemmobacter hezensis* D13-10-4-6^T (GCF_013155295.1), *Pseudogemmobacter bohemicus* Cd-10^T (GCF_003290025.1), *Pseudogemmobacter humi* CIP 111625^T (GCF_900609055.1), *Falsigemmobacter intermedius* 119/4^T (GCA_004054105.1), *Xinfangfangia soli* CCTCC AB 2017177^T (GCA_015999335.1), *Tabrizicola aquatica* RCRI19^T (GCF_002900975.1), *Falsirhodobacter deserti* W402^T (GCF_004015795.1), *Paracoccus limosus* JCM 17370^T (GCF_009711185.1), *Paracoccus sanguinis* OM2164^T (GCF_012689545.1). Numbers in parentheses indicate the number of proteins on which AAI was based.

Table 2. Differential characteristics of the strain of *Pseudogemmobacter sonorensis* PA1-206B^T and the closely related species.
Characters are scored as: +, positive; -, negative; w, weak positive reaction; DPG, diphosphatidylglycerol; PG, phosphatidylglycerol; PE, phosphatidylethanolamine; AL, unidentified aminolipid; PLx, unidentified phospholipids; PC, phosphatidylcholine; PME, phosphatidylmonomethylethanolamine; Lx, unidentified polar lipid. The data were obtained from this study unless otherwise is indicated. Data taken from: a, Ma et al. 2022 [1]; b, Kämpfer et al. 2015 [47]; c, Tarhriz et al. 2013 [48].

Supplementary Fig. S1. Scanning electron micrograph showing the general morphology of cell of strain PA1-206B^T after growth at 28 °C in R2A broth.

Supplementary Fig. S2. Trimmed phylogenomic tree of the family *Paracoccaceae* based on bacterial 120 conserved proteins of 2049 genomes concatenated by Genome Taxonomic Database Toolkit on Kbase server. Colour legends: purple – *Pseudogemmobacter* spp., yellow - strain PA1-206B^T.

Supplementary Fig. S3. The average amino acid identity (AAI) distribution of the *Paracoccaceae* family from 2049 high-quality genomes calculated with CompareM v0.1.2.

Supplementary Fig. S4. Phylogenomic tree indicating the position of strain PA1-206B^T among closely related taxa. Tree inferred with FastME 2.1.6.1 (Lefort et al. 2015 [32]) from GBDP distances calculated from genome sequences. The branch lengths are scaled in terms of GBDP distance formula d5. The numbers above branches are GBDP pseudo-bootstrap support values > 50 % from 100 replications, with an average branch support of 54.4 %. The tree was rooted at the midpoint (Farris, 1972 [33]) and visualized with PhyD3 (Kreft et al. 2017 [37]).

Supplementary Fig. S5. Two-dimensional TLC of polar lipids of strain PA1-206B^T after spraying with ninhydrin and heating at 100 °C for 10 minutes (A, aminolipids), after spraying with molybdenum blue (B, phospholipids), after spraying with 20 % (w/v) ethanolic phosphomolybdic acid (Sigma) and subsequent heating at 200 °C for 10 min (C, total lipids) and after spraying with α-naphthol reagent and heating at 100 °C for 5 minutes (D, no purple spots of glycolipids). Images A and B originate from the same chromatogram. Chloroform/methanol/water (65:25:4, by vol.) was used in the first direction (1), followed by chloroform/acetic acid/methanol/water (80:15:12:4, by vol.) in the second direction (2). Abbreviations: PG, phosphatidylglycerol; DPG, diphosphatidylglycerol; PE, phosphatidylethanolamine; PL, unidentified phospholipid; AL, unidentified aminolipid.

Supplementary Table S1. Distribution of genes based on RAST functional categories in the genomes of novel strain *Pseudogemmobacter sonorensis* PA1-206B^T and the closest type strain *Pseudogemmobacter hezensis* D13-10-4-6^T.

Supplementary Table S2. Fatty acid profiles of strains: 1, *Pseudogemmobacter sonorensis* PA1-206B^T (data from this study); 2, *Pseudogemmobacter hezensis* KCTC 82215^T (data from Ma et al. 2022 [1]); 3, *Falsigemmobacter intermedius* DSM 28642^T (data from Kämpfer et al. 2015 [47]); 4, *Tabrizicola aquatica* JCM 17277^T (data from Szuróczki et al. 2022 [5]).

Major fatty acids (> 5 %) are shown in bold type. Fatty acids amounting to <1% of the total fatty acids are listed as tr, trace (<1 %); -, Not detected.

* trace amounts identified as $C_{16:1}$ ω7c in *Pseudogemmobacter sonorensis*, ** identified as unknown 11.799 by the MIDI system, ***identified as part of Summed Feature 8 (18:1 ω7c/18:1 ω6c) by the MIDI system, but identified as C18:1ω7c by GC/MS.

Supplementary Table S3. API ZYM enzyme activities of strains: 1, *Pseudogemmobacter sonorensis* PA1-206B^T; 2, *Pseudogemmobacter hezensis* KCTC 82215^T; 3, *Falsigemmobacter intermedius* DSM 28642^T; 4, *Tabrizicola aquatica* JCM 17277^T. Characters are scored as: +, positive; -, negative; w, weak positive reaction.

Supplementary Table S4. API 50CH carbon source utilization of strains: 1, *Pseudogemmobacter sonorensis* PA1-206B^T; 2, *Pseudogemmobacter hezensis* KCTC 82215^T; 3, *Falsigemmobacter intermedius* DSM 28642^T; 4, *Tabrizicola aquatica* JCM 17277^T. Characters are scored as: +, positive; -, negative; w, weak positive reaction.

Tables

Table 1. Average nucleotide identity (ANI), average amino acid identity (AAI), digital DNA–DNA hybridization (dDDH) and G+C difference values between the genome sequence of PAI-206B^T and the available reference genomes of its relatives: *Pseudogemmobacter hezensis* D13-10-4-6^T (GCF_013155295.1), *Pseudogemmobacter bohemicus* Cd-10^T (GCF_003290025.1), *Pseudogemmobacter humi* CIP 111625^T (GCF_900609055.1), *Falsigemmobacter intermedius* 119/4^T (GCA_004054105.1), *Xinfangfangia soli* CCTCC AB 2017177^T (GCA_015999335.1), *Tabrizicola aquatica* RCRI19^T (GCF_002900975.1), *Falsirhodobacter deserti* W402^T (GCF_004015795.1), *Paracoccus limosus* JCM 17370^T (GCF_009711185.1), *Paracoccus sanguinis* OM2164^T (GCF_012689545.1). Numbers in parentheses indicate the number of proteins on which AAI was based.

Strain	**ANI (%)**	**AAI (%)**	**dDDH (%)**	**G+C difference (mol%)**
Pseudogemmobacter hezensis D13-10-4-6^T	80.39	69.59 (2568)	17.3	4.50
Pseudogemmobacter bohemicus Cd-10^T	78.97	68.92 (2679)	17.3	4.00
Pseudogemmobacter humi CIP 111625^T	80.30	70.62 (2571)	21.6	0.85
Falsigemmobacter intermedius 119/4^T	76.67	61.18 (2235)	13.6	4.74
Xinfangfangia soli CCTCC AB 2017177^T	78.55	66.19 (2381)	17.1	0.20
Tabrizicola aquatica RCRI19^T	77.98	66.88 (2314)	15.9	0.98
Falsirhodobacter deserti W402^T	77.41	61.98 (1908)	14.3	3.95
Paracoccus limosus JCM 17370^T	77.50	59.77 (2220)	15.2	1.31
Paracoccus sanguinis OM2164^T	80.65	58.35 (971)	14.4	3.51

Table 2. Differential characteristics of the strain of *Pseudogemmobacter sonorensis* PA1-206B^T and the closely related species.

Characteristic	***Pseudogemmobacter sonorensis*** **PA1-206B^T**	***Pseudogemmobacter hezensis*** **KCTC 82215^T**	***Falsigemmobacter intermedius*** **DSM 28642^T**	***Tabrizicola aquatica*** **JCM 17277^T**
Colony pigmentation	Creamy white	Creamy white	Cream	Cream
Cell size (μm)	1.0-1.5 x 0.8	1.6–2.0 x 0.8–1.0[a]	1.0–1.2 x 2.0–5.0[b]	0.9×1.3–3[c]
Cell shape	irregular rod-shaped	ovoid to rod-shaped[a]	rod- to irregular shaped[b]	rod-shaped[c]
Gram strain	-	-[a]	-[b]	-[c]
Source of isolation	wound exudate of tree	bark sample of tree[a]	white stork[b]	freshwater lake[c]
Growth temperature (°C)	15-35	15-35[a]	15-55[b]	15-45[c]
NaCl (% w/v) tolerance	0	0-4	0-5	0-5
pH for growth:				
range	6.0-10.0	5.0-10.0	5.0-10.0	6.0-13.0
optimum	8.0-9.0	7.0-8.0	8.0	7.0-12.0
Acid production from D-glucose (aerob)	-	+	-	-
Hydrolysis of gelatine	+	-	-	-
Nitrate to nitrite	w	w	-	+
Phosphatase activity	-	+	w	-
Enzyme activity (API ZYM):				
Alkaline phosphatase	+	-	-	-
Esterase(C4)	+	-	+	-
Esterase lipase (C8)	w	-	w	+
Leucin arylamidase	+	+	-	+
Valin arylamidase	+	w	-	-
Naphtol-AS-BI-phosphohydrolase	+	w	w	+
α-galactosidase	-	-	-	+
β-galactosidase	-	-	-	+
α-glucosidase	+	+	-	+
β-glucosidase	-	-	-	+
N-Acetyl-β-glucosaminidase	+	+	-	-
Acid production from (API 50CH):				
Erythritol	-	w	-	-
D-arabinose	-	w	-	-
L-arabinose	-	+	-	-
D-ribose	-	w	-	-
D-xylose	-	+	-	-
D-galactose	-	+	-	-
D-glucose	+	+	-	-
D-fructose	+	w	-	-
L-rhamnose	-	w	-	-
D-mannitol	w	-	-	-
N-acetyl-glucosamine	-	w	-	-
D-saccharose	+	-	-	-
D-trehalose	+	-	-	-
D-lyxose	-	+	-	-
D-fucose	-	+	-	-
L-fucose	-	+	-	-
Major polar lipids	DPG,PG,PE,AL,PL_{1-2}	PME,DPG,PE,PG,PC,PL,L_6[a]	PME, DPG, PG, PE, PC,AL_{1-3},L_5[b]	PC,PG,DPG,PE_1[c]
DNA G+C content (mol%)	67.4	62.9[a]	64[b]	65.9[c]

Characters are scored as: +, positive; -, negative; w, weak positive reaction; DPG, diphosphatidylglycerol; PG, phosphatidylglycerol; PE, phosphatidylethanolamine; AL, unidentified aminolipid; PLx, unidentified phospholipids; PC, phosphatidylcholine; PME, phosphatidylmonomethylethanolamine; Lx, unidentified polar lipid. The data were obtained from this study unless otherwise is indicated. Data taken from: a, Ma et al. 2022 [1]; b, Kämpfer et al. 2015 [47]; c, Tarhriz et al. 2013 [48].

Figure_1 Click here to access/download;Figure;Fig. 1.png

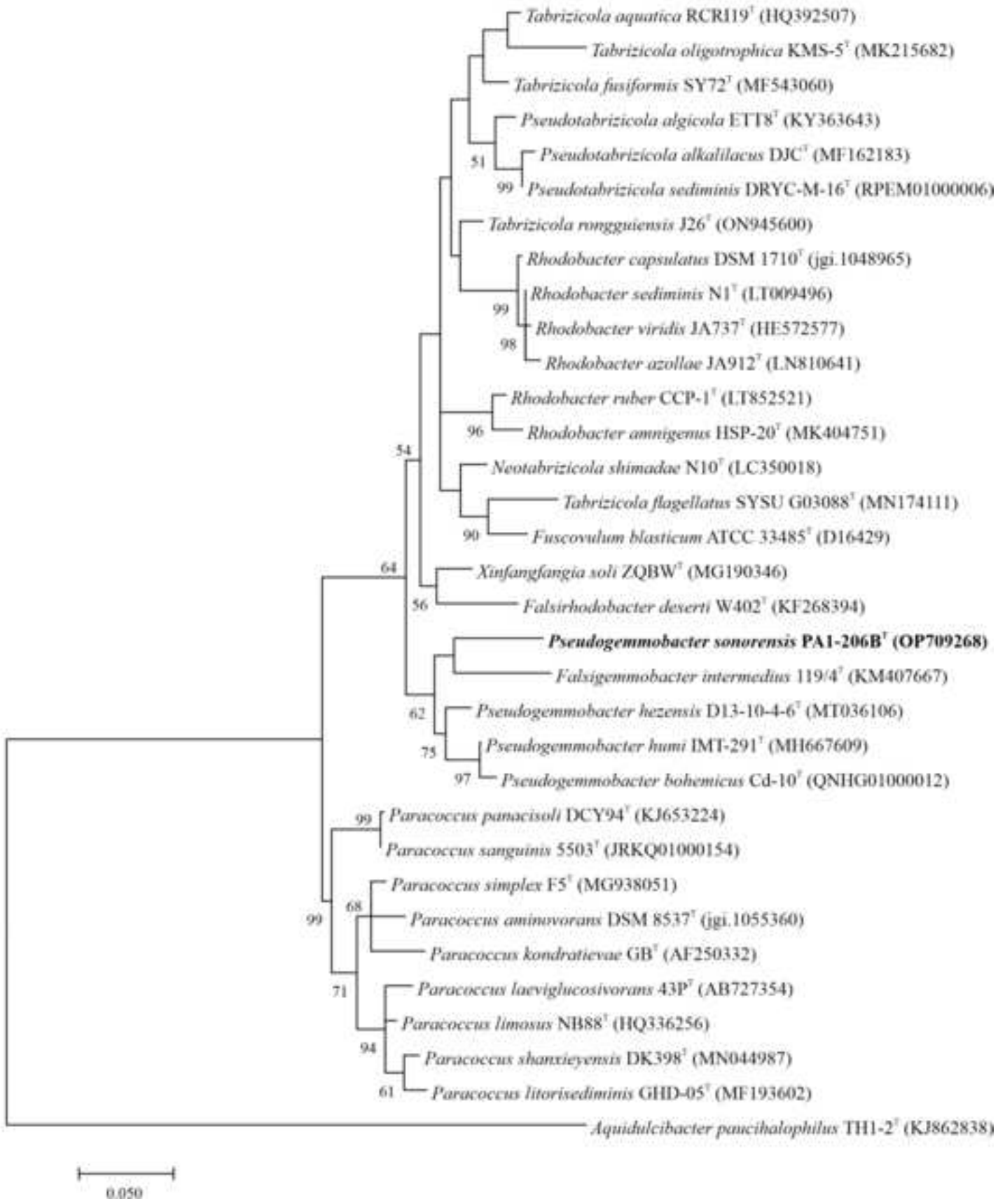


Figure_2

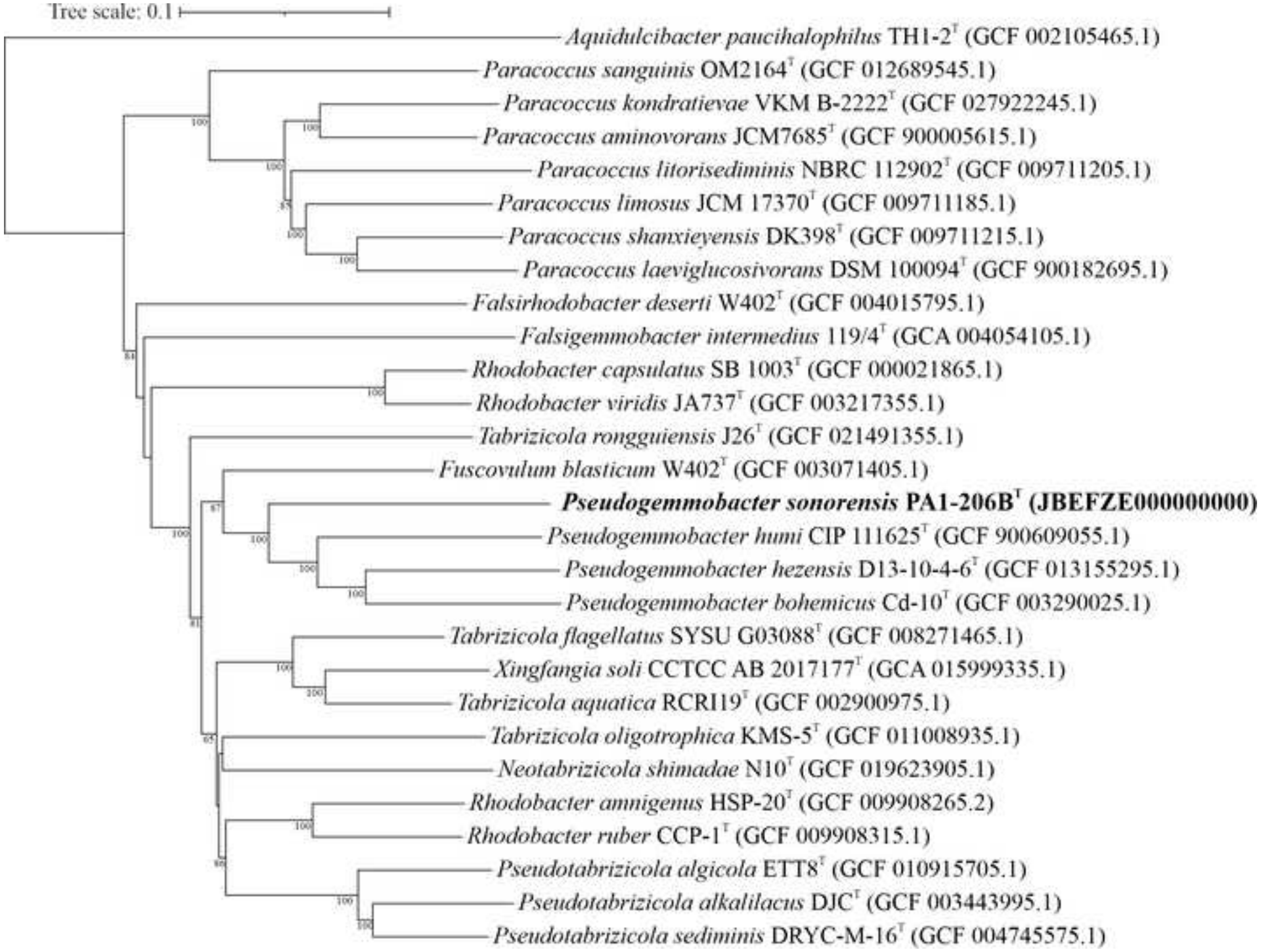
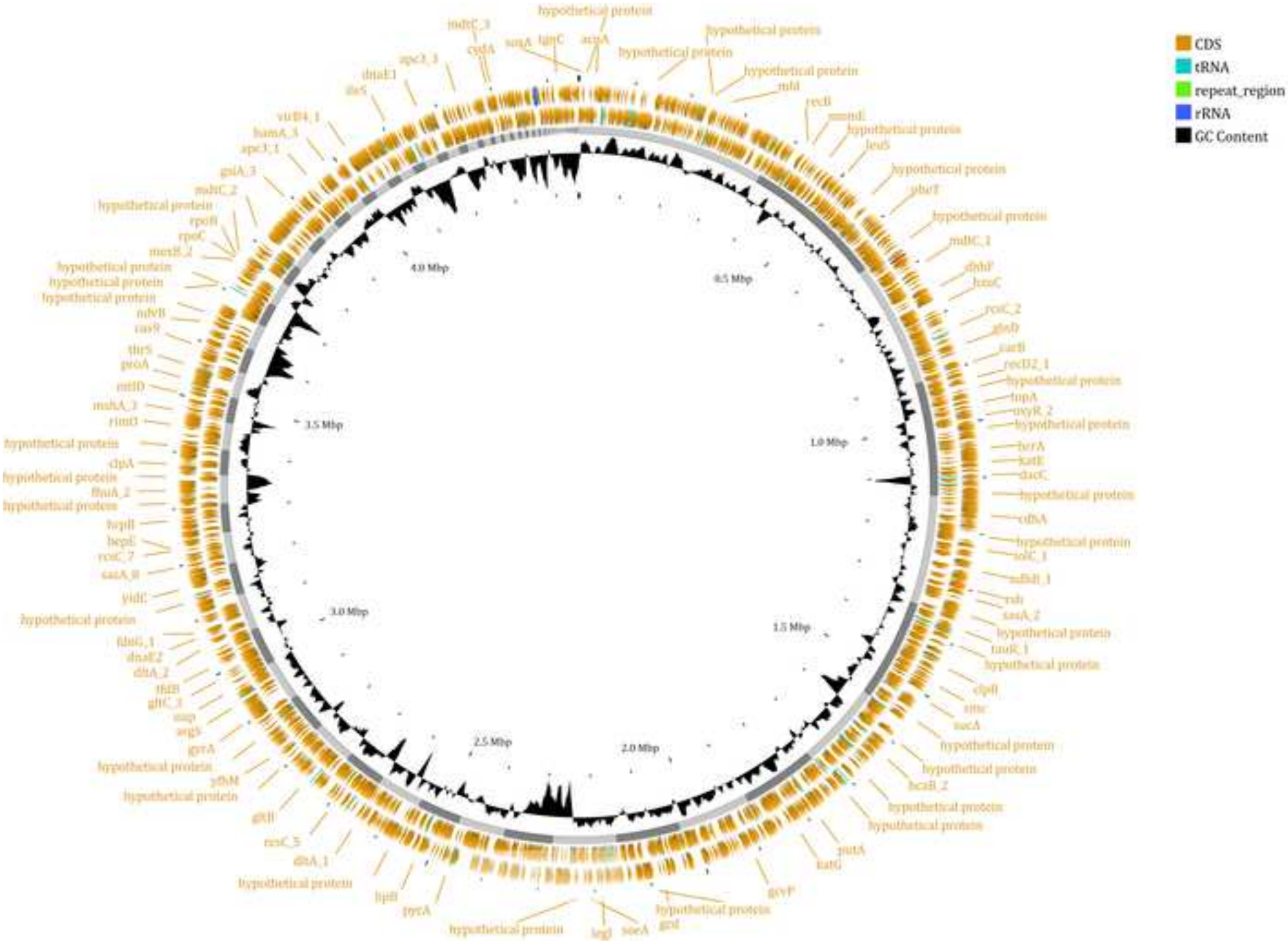


Figure_3

Click here to access/download;Figure;Fig.3..png



Pseudogemmobacter sonorensis sp. nov., a new alphaproteobacterium isolated from the slime flux of a tree (*Populus fremontii*) in the Sonora Desert (Arizona, USA).

Nóra Tünde Lange-Enyedi[1,2], Erika Tóth[1,3], Gorkhmaz Abbaszade[1,4], Péter Németh[5,6], Laurence A. J. Garvie[7], Jacqueline Wolf[8], Meina Neumann-Schaal[8], Bernadett Khayer[9], György Sipos[2], Judit Makk[1]

[1]Department of Microbiology, Faculty of Science, Eötvös Loránd University, Pázmány Péter stny, 1/C H-1117 Budapest, Hungary

[2]Functional Genomics and Bioinformatics Group, Faculty of Forestry, University of Sopron, Bajcsy-Zsilinszky út 4, H-9400, Sopron, Hungary

[3]Health Promotion and Education Research Team, Hungarian Academy of Sciences, Budapest, Hungary

[4]Department of Applied Microbial Ecology, Helmholtz Centre for Environmental Research – UFZ, Permoserstr. 15, 04318 Leipzig, Germany

[5]Institute for Geological and Geochemical Research, HUN-REN Research Centre for Astronomy and Earth Sciences, Budaörsi út 45 H-1112 Budapest, Hungary

[6]Research Institute of Biomolecular and Chemical Engineering, Nanolab, University of Pannonia, Egyetem út 10 H-8200 Veszprém, Hungary

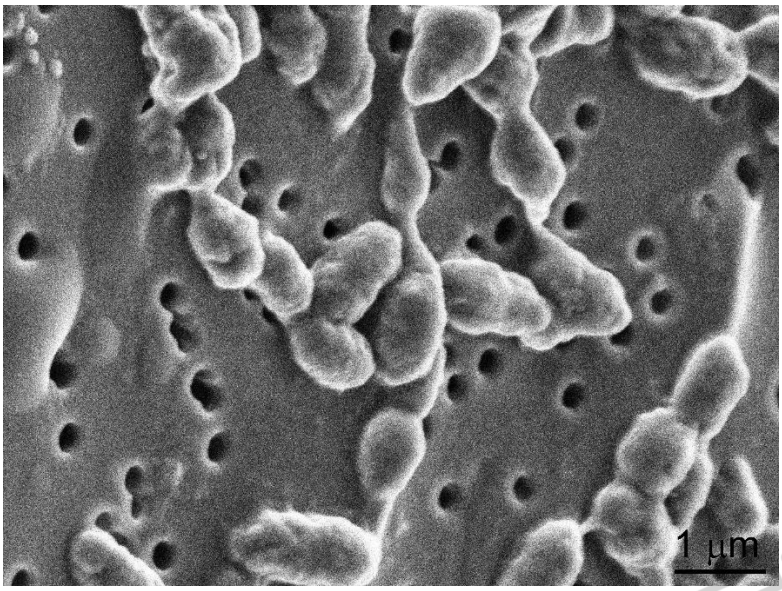
[7]School of Earth and Space Exploration, Arizona State University, 781 East Terrace Road, Tempe, AZ 85287-6004, USA

[8]Leibniz Institute DSMZ – German Collection of Microorganisms and Cell Cultures, Inhoffenstrasse 7B, D-38124, Braunschweig, Germany

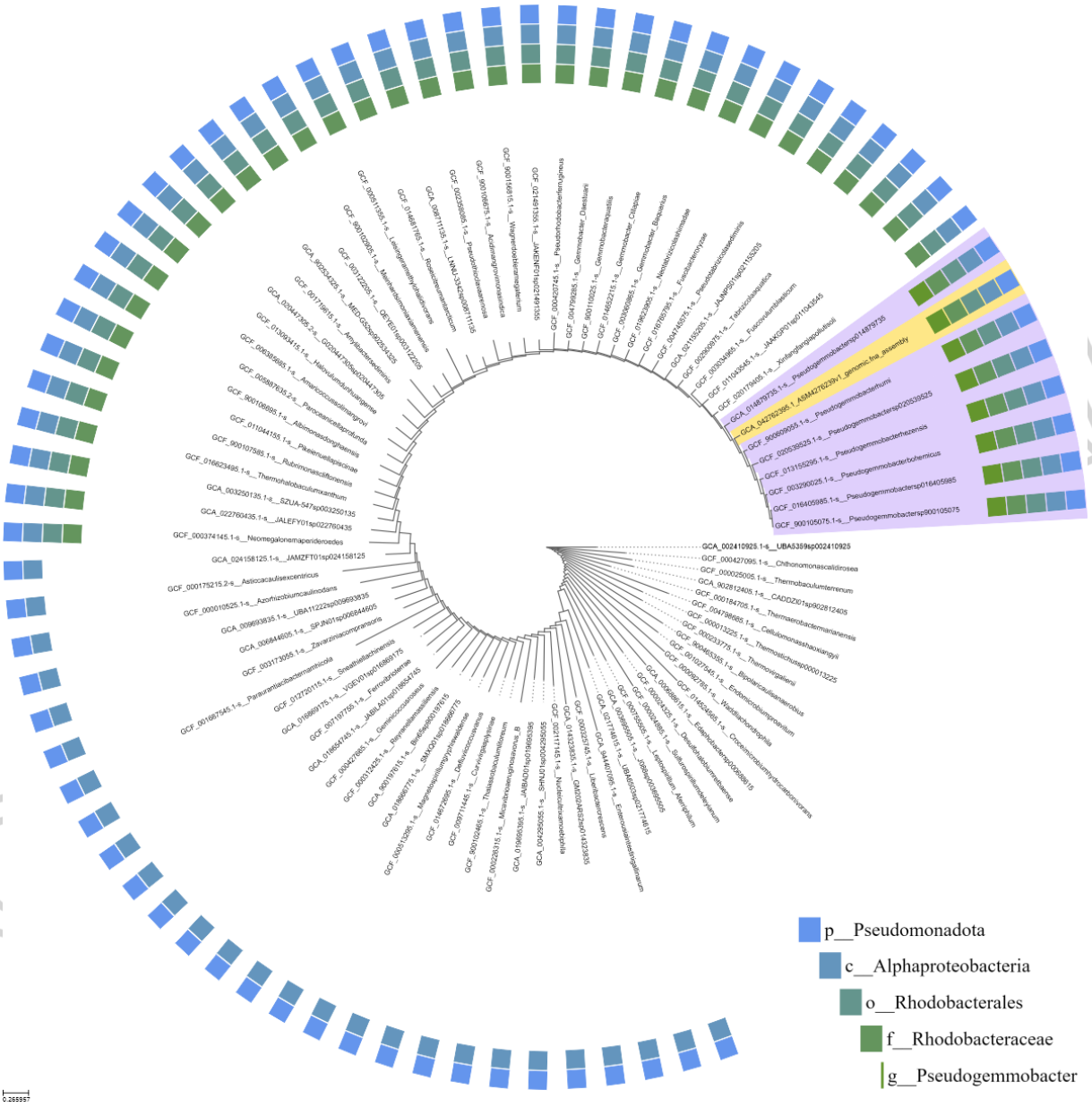
[9]Department of Public Health Laboratory and Methodology, National Center Public Health and Pharmacy, Albert Flórián út 2-6 H-1097 Budapest, Hungary

Correspondence author:
Judit Makk
e-mail: makk.judit@ttk.elte.hu

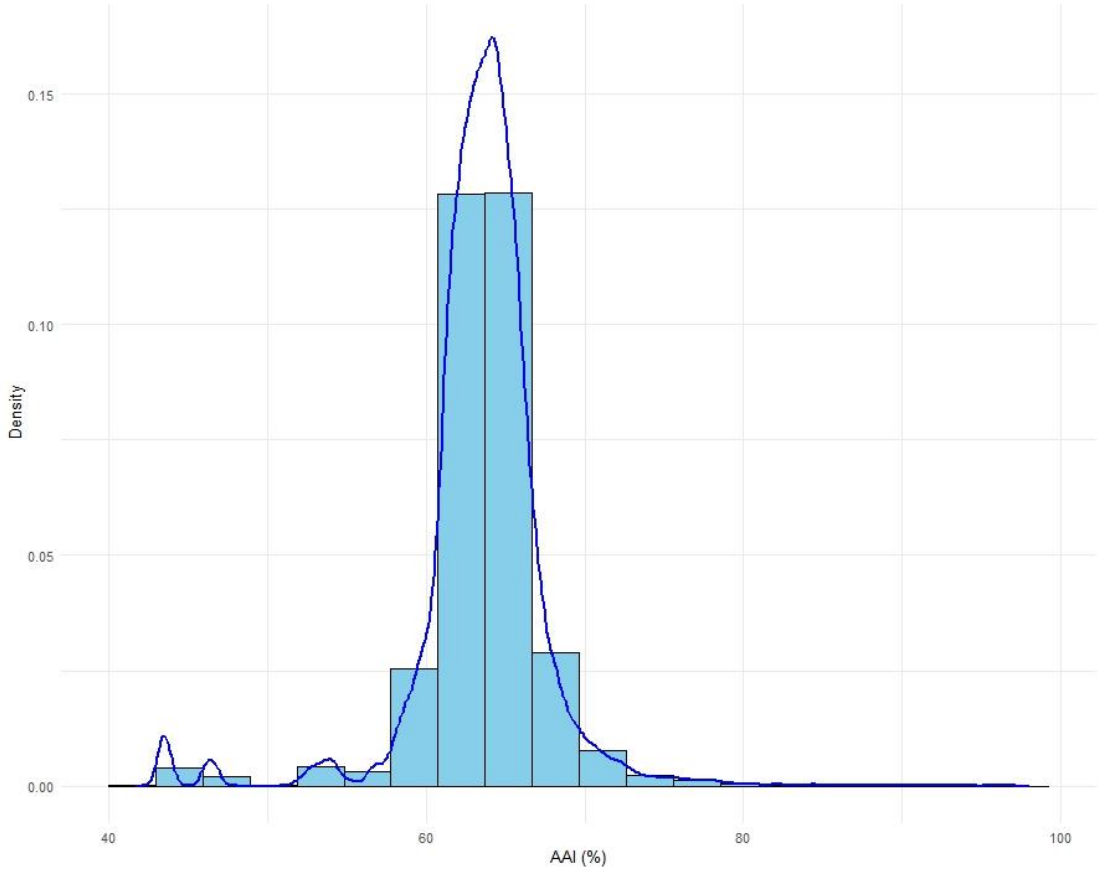
Supplementary Fig. S1. Scanning electron micrograph showing the general morphology of cell of strain PA1-206B^T after growth at 28 °C in R2A broth.



Supplementary Fig. S2. Trimmed phylogenomic tree of the family *Paracoccaceae* based on bacterial 120 conserved proteins of 2049 genomes concatenated by Genome Taxonomic Database Toolkit on Kbase server. Colour legends: purple – *Pseudogemmobacter* spp., yellow - strain PA1-206B^{T}.

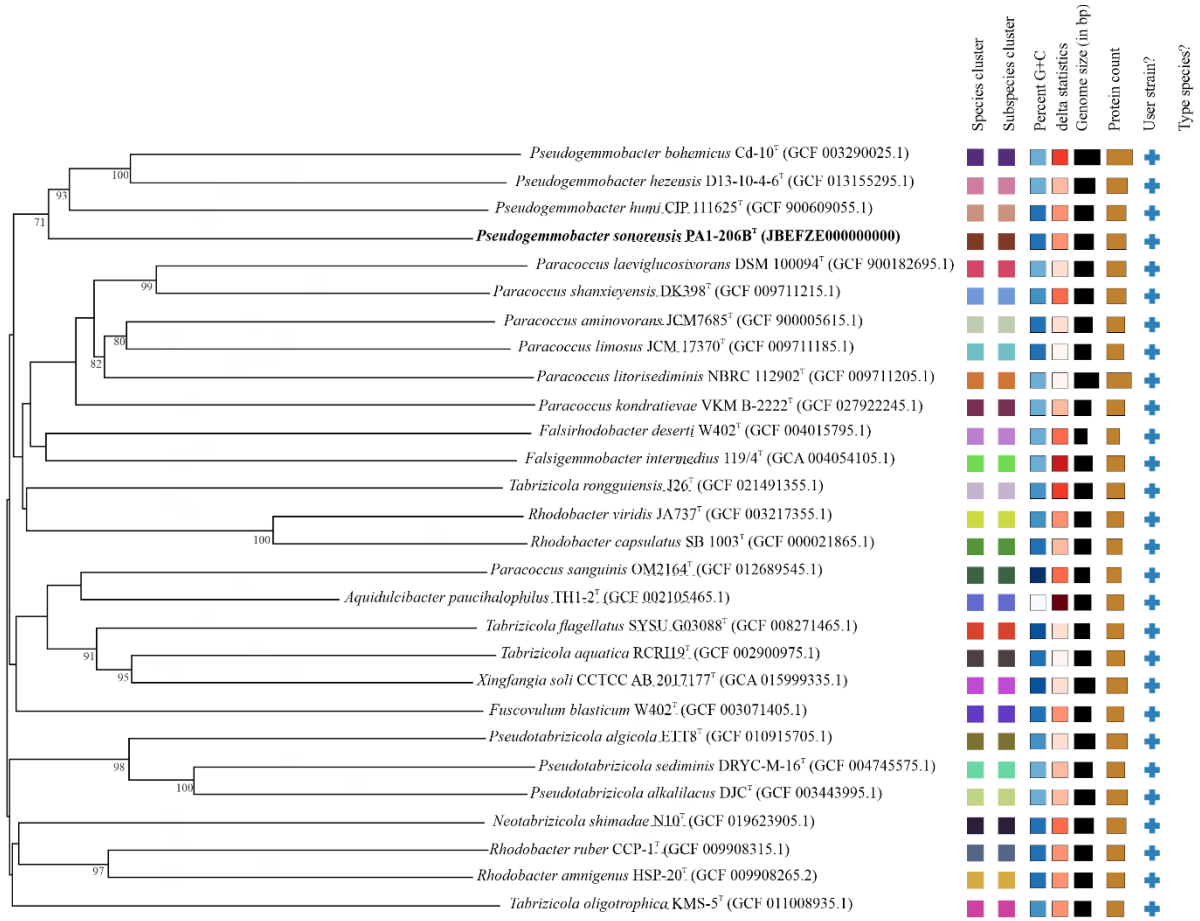


Supplementary Fig. S3. The average amino acid identity (AAI) distribution of the *Paracoccaceae* family from 2049 high-quality genomes calculated with CompareM v0.1.2.

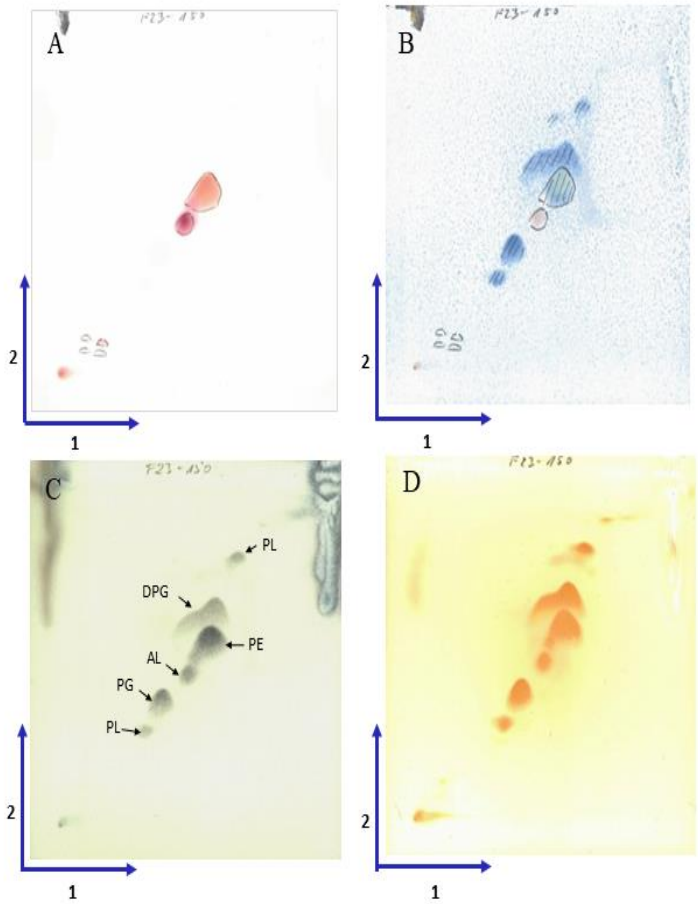


Min.	1st Quantile	Median	Mean	3rd Quantile	0.95 Quantile	Max.
41.75	62.1	63.7	63.6	65.4	68.8	98.1

Supplementary Fig. S4. Phylogenomic tree indicating the position of strain PA1-206B^T among closely related taxa. Tree inferred with FastME 2.1.6.1 (Lefort et al. 2015 [32]) from GBDP distances calculated from genome sequences. The branch lengths are scaled in terms of GBDP distance formula d5. The numbers above branches are GBDP pseudo-bootstrap support values > 50 % from 100 replications, with an average branch support of 54.4 %. The tree was rooted at the midpoint (Farris, 1972 [33]) and visualized with PhyD3 (Kreft et al. 2017 [37]).



Supplementary Fig. S5. Two-dimensional TLC of polar lipids of strain PA1-206B^T after spraying with ninhydrin and heating at 100 °C for 10 minutes (A, aminolipids), after spraying with molybdenum blue (B, phospholipids), after spraying with 20% (w/v) ethanolic phosphomolybdic acid (Sigma) and subsequent heating at 200 °C for 10 min (C, total lipids) and after spraying with α-naphthol reagent and heating at 100 °C for 5 minutes (D, no purple spots of glycolipids). Images A and B originate from the same chromatogram. Chloroform/methanol/water (65:25:4, by vol.) was used in the first direction (1), followed by chloroform/acetic acid/methanol/water (80:15:12:4, by vol.) in the second direction (2). Abbreviations: PG, phosphatidylglycerol; DPG, diphosphatidylglycerol; PE, phosphatidylethanolamine; PL, unidentified phospholipid; AL, unidentified aminolipid.



Supplementary Table S1. Distribution of genes based on RAST functional categories in the genomes of novel strain *Pseudogemmobacter sonorensis* PA1-206B^T and the closest type strain *Pseudogemmobacter hezensis* D13-10-4-6^T.

Subsystem features	***Pseudogemmobacter hezensis* D13-10-4-6^T**	***Pseudogemmobacter sonorensis* PA1-206B^T**	**Common features**
Amino acids and derivatives	11	22	148
Carbohydrates	24	8	103
Cell wall and capsule	1	2	17
Clustering-based subsystems	11	17	143
Cofactors, vitamins, prosthetic groups, pigments	15	8	80
DNA metabolism	5	7	43
Dormancy and sporulation	0	0	1
Fatty acids, lipids, and isoprenoids	0	1	34
Iron acquisition and metabolism	3	9	4
Membrane transport	21	7	36
Metabolism of aromatic compounds	2	11	9
Miscellaneous	0	4	24
Motility and chemotaxis	0	0	5
Nitrogen metabolism	3	1	8
Nucleosides and nucleotides	2	11	63
Phages, prophages, transposable elements, plasmids	1	4	17
Phosphorus metabolism	3	1	15
Potassium metabolism	0	0	3
Protein metabolism	4	11	146
RNA metabolism	0	2	30
Regulation and cell signaling	2	6	8
Regulons	0	0	1
Respiration	1	7	92
Secondary metabolism	0	0	4
Stress response	10	4	32
Sulfur metabolism	2	2	15
Virulence, disease and defense	0	3	20
Sum of subsystem feature counts	**121**	**148**	**1101**

Supplementary Table S2. Fatty acid profiles of strains: 1, *Pseudogemmobacter sonorensis* PA1-206B^T (data from this study); 2, *Pseudogemmobacter hezensis* KCTC 82215^T [data from Ma et al. 2022 [1]); 3, *Falsigemmobacter intermedius* DSM 28642^T (data from Kämpfer et al. 2015 [47]); 4, *Tabrizicola aquatica* JCM 17277^T (data from Szuróczki et al. 2022 [5]).

Major fatty acids (> 5 %) are shown in bold type. Fatty acids amounting to <1% of the total fatty acids are listed as tr, trace (<1 %); -, Not detected.

* trace amounts identified as $C_{16:1}$ ω7c in *Pseudogemmobacter sonorensis*, ** identified as unknown 11.799 by the MIDI system, ***identified as part of Summed Feature 8 (18:1 ω7c/18:1 ω6c) by the MIDI system, but identified as C18:1ω7c by GC/MS.

Fatty acid	1	2	3	4
$C_{12:0}$	tr	-	-	-
$C_{14:0}$	tr	-	-	-
$C_{15:0}$	tr	-	-	-
$C_{16:0}$	**19.2**	**5.4**	3.0	tr
$C_{17:0}$	2.3	-	3.7	-
$C_{18:0}$	3.6	4.1	-	1.4
$C_{10:0}$ 3-OH	2.2	2.4	3.1	3.4
$C_{14:0}$ 3-OH	tr	-	-	-
$C_{18:0}$ 3-OH	-	2.6	-	-
$C_{12:1}$ ω7c	3.8	-	2.8**	3.2
$C_{16:1}$ ω7c/$C_{16:1}$ ω6c*	tr	1.5	**6.6**	tr
$C_{17:1}$ ω6c	tr	-	-	-
$C_{17:1}$ ω7c	-	-	-	tr
$C_{18:1}$ ω7c***	**59.5**	**81.1**	**80.9**	**85.9**
$C_{18:1}$ ω5c	-	-	-	tr
11-methyl-$C_{18:1}$ ω7c	**7.3**	-	-	1.3
$C_{19:1}$ iso ω5c	-	-	-	3.0
$C_{20:1}$ iso ω7c	-	-	-	tr

Supplementary Table S3. API ZYM enzyme activities of strains: 1, *Pseudogemmobacter sonorensis* PA1-206B^T; 2, *Pseudogemmobacter hezensis* KCTC 82215^T; 3, *Falsigemmobacter intermedius* DSM 28642^T; 4, *Tabrizicola aquatica* JCM 17277^T. Characters are scored as: +, positive; -, negative; w, weak positive reaction.

Enzyme activities	**1**	**2**	**3**	**4**
Alkaline phosphatase	+	-	-	-
Esterase (C4)	+	-	+	-
Esterase lipase (C8)	w	-	w	+
Lipase (C14)	-	-	-	-
Leucin arylamidase	+++	++	-	+
Valin arylamidase	++	w	-	-
Cystine arylamidase	-	-	-	-
Trypsin	-	-	-	-
α-chymotrypsin	-	-	-	-
Acid phosphatase	-	-	-	-
Naphtol-AS-BI-phosphohydrolase	+	w	w	+
α-galactosidase	-	-	-	+
β-galactosidase	-	-	-	+
β-glucuronidase	-	-	-	-
α-glucosidase	+	+	-	+
β-glucosidase	-	-	-	+
N-Acetyl-β-glucosaminidase	+	+++	-	-
α-mannosidase	-	-	-	-
α-fucosidase	-	-	-	-

Supplementary Table S4. API 50CH carbon source utilization of strains: 1, *Pseudogemmobacter sonorensis* PA1-206B^T; 2, *Pseudogemmobacter hezensis* KCTC 82215^T; 3, *Falsigemmobacter intermedius* DSM 28642^T; 4, *Tabrizicola aquatica* JCM 17277^T. Characters are scored as: +, positive; -, negative; w, weak positive reaction.

Carbon source	1	2	3	4	Carbon source	1	2	3	4
Glycerol	-	-	-	-	Salicin	-	-	-	-
Erythritol	-	w	-	-	D-cellobiose	-	-	-	-
D-arabinose	-	w	-	-	D-maltose	-	-	-	-
L-arabinose	-	+	-	-	D-lactose	-	-	-	-
D-ribose	-	w	-	-	D-melibiose	-	-	-	-
D-xylose	-	+	-	-	D-sucrose	+	-	-	-
L-xylose	-	-	-	-	D-trehalose	+	-	-	-
D-adonitol	-	-	-	-	Inulin	-	-	-	-
Methyl-β-D-xylopyranoside	-	-	-	-	D-melezitose	-	-	-	-
D-galactose	-	+	-	-	D-raffinose	-	-	-	-
D-glucose	+	+	-	-	Amidon	-	-	-	-
D-fructose	+	w	-	-	Glycogen	-	-	-	-
D-mannose	-	-	-	-	Xylitol	-	-	-	-
L-sorbose	-	-	-	-	Gentiobiose	-	-	-	-
L-rhamnose	-	w	-	-	D-turanose	-	-	-	-
Dulcitol	-	-	-	-	D-lyxose	-	+	-	-
Inositol	-	-	-	-	D-tagatose	-	-	-	-
D-mannitol	w	-	-	-	D-fucose	-	+	-	-
D-sorbitol	-	-	-	-	L-fucose	-	+	-	-
Methyl-α-D-mannopyranoside	-	-	-	-	D-arabitol	-	-	-	-
Methyl-α-D-glucopyranoside	-	-	-	-	L-arabitol	-	-	-	-
N-acetyl-glucosamine	-	w	-	-	Potassium-gluconate	-	-	-	-
Amygdalin	-	-	-	-	Potassium-2-ketogluconate	-	-	-	-
Arbutin	-	-	-	-	Potassium-5-ketogluconate	-	-	-	-
Esculin	-	-	-	-					