



Data Article

Genomic data resource of type strains of genus *Pseudoxanthomonas*

Kanika Bansal¹, Sanjeet Kumar¹, Prashant P. Patil, Shikha Sharma, Prabhu B. Patil*

Bacterial Genomics and Evolution Laboratory, CSIR-Institute of Microbial Technology, Sector-39A, Chandigarh 160036, India

ARTICLE INFO

Article history:

Received 22 October 2021

Revised 28 March 2022

Accepted 29 March 2022

Available online 15 April 2022

Dataset link: [Genomic data resource of type strains of genus *Pseudoxanthomonas* \(Original data\)](#)

Keywords:

Pseudoxanthomonas

Type strains

Phylogenomics

OrthoANI

Whole genome sequencing

Illumina MiSeq

ABSTRACT

Genus *Pseudoxanthomonas* represents a relatively newly characterized group of gamma-proteobacterium of environmental origin. Species of the genus have very similar morphology to strains belonging to *Xanthomonas*, *Xylella* and *Stenotrophomonas*. However, the genome resource of this genus was largely unexplored. The species belonging to the genus are from a wide range of environmental sites including hydrocarbon polluted fields. Here, we have provided the whole genome sequence of all available type strains of the genus of *Pseudoxanthomonas*. In order to deduce the differences with closely related genera, we have employed the whole genome-based investigation of the type species of genus *Pseudoxanthomonas*.

© 2022 The Authors. Published by Elsevier Inc.
This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>)

DOI of original article: [10.1016/j.ygeno.2021.09.021](https://doi.org/10.1016/j.ygeno.2021.09.021)

* Corresponding author.

E-mail address: pbpatil@imtech.res.in (P.B. Patil).

Social media: [@Kanikabansal91](#) (K. Bansal), [@LifeInfoATGC](#) (S. Kumar)

¹ These authors contributed equally to this work.

<https://doi.org/10.1016/j.dib.2022.108145>

2352-3409/© 2022 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>)

Specifications Table

Subject	Biological sciences
Specific subject area	Microbiology: Bacteriology
Type of data	Whole genome sequence assembled genome with gene annotation and phylogeny of genus <i>Pseudoxanthomonas</i> and its related genera.
How the data were acquired	Whole genome sequencing (WGS) library was prepared for Illumina MiSeq sequencing platform. Assembly of the raw reads were performed using SPAdes v3.10.
Data format	Raw
Parameters for data collection	Analyzed
Description of data collection	Sequencing library for all the available type strains were prepared for Illumina MiSeq following manufacturer's instructions. Sequencing was performed with 2°250 bp paired end sequencing kit.
Data source location	WGS data obtained from the sequencer was quality trimmed by control software of Illumina MiSeq. Raw reads were <i>de novo</i> assembled into high quality draft genome was performed using SPAdes v3.10 and quality checked using CheckM v1.10
Data accessibility	Institution: CSIR-Institute of Microbial Technology, Chandigarh City/Town/Region: Chandigarh Country: INDIA NCBI: MWIP000000000: https://www.ncbi.nlm.nih.gov/nucleotide/MWIP000000000 NCBI: PDWO000000000: https://www.ncbi.nlm.nih.gov/nucleotide/PDWO000000000 NCBI: PDWN000000000: https://www.ncbi.nlm.nih.gov/nucleotide/PDWN000000000 NCBI: PDWT000000000: https://www.ncbi.nlm.nih.gov/nucleotide/PDWT000000000 NCBI: PDWS000000000: https://www.ncbi.nlm.nih.gov/nucleotide/PDWS000000000 NCBI: PDWW000000000: https://www.ncbi.nlm.nih.gov/nucleotide/PDWW000000000 NCBI: PDWU000000000: https://www.ncbi.nlm.nih.gov/nucleotide/PDWU000000000 NCBI: PDWR000000000: https://www.ncbi.nlm.nih.gov/nucleotide/PDWR000000000 NCBI: PDWL000000000: https://www.ncbi.nlm.nih.gov/nucleotide/PDWL000000000 NCBI: PDWQ000000000: https://www.ncbi.nlm.nih.gov/nucleotide/PDWQ000000000 NCBI: PDWM000000000: https://www.ncbi.nlm.nih.gov/nucleotide/PDWM000000000 NCBI: PDWP000000000: https://www.ncbi.nlm.nih.gov/nucleotide/PDWP000000000 NCBI: PDWV000000000: https://www.ncbi.nlm.nih.gov/nucleotide/PDWV000000000 NCBI: PDWK000000000: https://www.ncbi.nlm.nih.gov/nucleotide/PDWK000000000 NCBI: QOVG000000000: https://www.ncbi.nlm.nih.gov/nucleotide/QOVG000000000
Related research article	Bansal, K., Kumar, S., Kaur, A., Singh, A. & Patil, P. B. (2021) Deep phylo-taxono genomics reveals <i>Xylella</i> as a variant lineage of plant associated <i>Xanthomonas</i> and supports their taxonomic reunification along with <i>Stenotrophomonas</i> and <i>Pseudoxanthomonas</i> . https://doi.org/10.1016/j.jygeno.2021.09.021 [1].

Value of the Data

- Species of genus *Pseudoxanthomonas* are from contaminated sites such as: heavy metal, oil, hydrocarbons etc. Genome resource of strains from such extreme environmental conditions will aid in identification of genomic signatures underlying their bioremediation potential.
- These assembled genomes can be reused as a reference by the taxonomist and microbiologist in order to distinguish any putative species of the genera *Pseudoxanthomonas*.
- Present genome resource of type strains will be valuable in addressing the taxonomic ambiguities of the family *Lysobacteraceae* and order *Lysobacterales*.

1. Data Description

Here, we have performed whole-genome sequencing of the 15 type strains of genus *Pseudoxanthomonas* comprising of 14 valid species and one non-valid type strain of *P. jiangsuensis* DSM 22398^T based on LPSN latest classification v2.0. Whole genome data of *P. dokdonensis* DSM 21858^T, *P. indica* P15^T and *P. spadix* BD-a59 were obtained from the public repository of NCBI (Table 1). *P. helianthi* NRBC 110414^T [2] and *P. putridarboris* LMG 25968^T [3] were not retrieved and thus whole genome sequence information is not included in the study. 16S rRNA

Table 1
Genome assembly statistics of the species of genus *Pseudoxanthomas*.

Strain name	Genome size (bps)	Fold	# Contigs	N50 (bps)	% GC	Completeness/ Contamination	# CDS	tRNA + rRNA	#Putative Plasmids	Accession number	Refs.
<i>P. broegbernensis</i> DSM 12573 ^T	3,547,767	267x	157	175,278	70.6	99.66/1.74	3024	54 + 3	6	MWIP00000000	Current study
<i>P. kaohsiungensis</i> DSM 17583 ^T	3,774,556	269x	100	199,653	69.68	99.66/1.42	3420	52 + 3	6	PDW000000000	Current study
<i>P. daejeonensis</i> DSM 17801 ^T	3,563,566	131x	38	227,294	68.89	99.66/0.11	3143	58 + 3	3	PDWN00000000	Current study
<i>P. yeongjuensis</i> DSM 18204 ^T	3,937,688	137x	28	610,842	65.11	99.95/0.76	3389	53 + 3	1	PDWT000000000	Current study
<i>P. sacheonensis</i> DSM 19373 ^T	4,036,514	101x	44	165,077	64.3	100/0.91	3582	50 + 3	1	PDWS000000000	Current study
<i>P. japonensis</i> DSM 17109 ^T	4,075,711	111x	74	108,688	67.3	99.95/0.34	3673	50 + 3	4	PDWW000000000	Current study
<i>P. wuyuanensis</i> DSM 100640 ^T	4,686,433	112x	86	149,132	65.75	100/1.08	4003	51 + 6	6	PDWU000000000	Current study
<i>P. sangjuensis</i> DSM 28345 ^T	3,289,016	91x	76	105,206	68.68	99.95/0.46	2908	51 + 3	3	PDWR000000000	Current study
<i>P. jiangsuensis</i> DSM 22398 ^T	3,790,571	171x	154	121,587	70.35	99.31/1.03	3389	50 + 3	10	PDWL000000000	Current study
<i>P. kalamensis</i> DSM 18571 ^T	3,034,522	239x	162	478,688	65.88	99.84/1.50	2697	48 + 3	-	PDWQ000000000	Current study
<i>P. koreensis</i> KCTC12208 ^T	3,049,736	74x	93	60,801	70.17	99.48/0.04	2681	52 + 3	7	PDWM000000000	Current study
<i>P. suwonensis</i> DSM 17175 ^T	3,428,455	65x	66	104,754	70.36	99.66/0.34	3071	55 + 3	2	PDWP000000000	Current study
<i>P. mexicana</i> DSM17121 ^T	3,965,467	32x	115	63,754	67.4	99.49/1.71	3660	54 + 6	10	PDWV000000000	Current study
<i>P. taiwanensis</i> DSM22914 ^T	3,043,352	69x	160	39,985	72.08	99.16/0.34	2729	56 + 3	14	PDWK000000000	Current study
<i>P. gei</i> KCTC32298 ^T	3,431,103	173x	50	332,530	65.46	99.95/0.41	3,122	45 + 4	-	QOVQ000000000	Current study
<i>P. dokdonensis</i> DSM21858 ^T	3,553,658	170x	34	-	64.48	99.59/0.41	3153	50 + 2	NA	LDJL010000000	[4]
<i>P. indica</i> P15 ^T	3,960,920	-	3	-	65.4	99.89/0.41	3593	49 + 6	NA	FUZV010000000	DOE-Joint Genome Institute [22]
<i>P. spadix</i> BD-a59	3,452,554	-	1	-	67.65	97.1/1.3	3153	50 + 3	NA	CP003093	

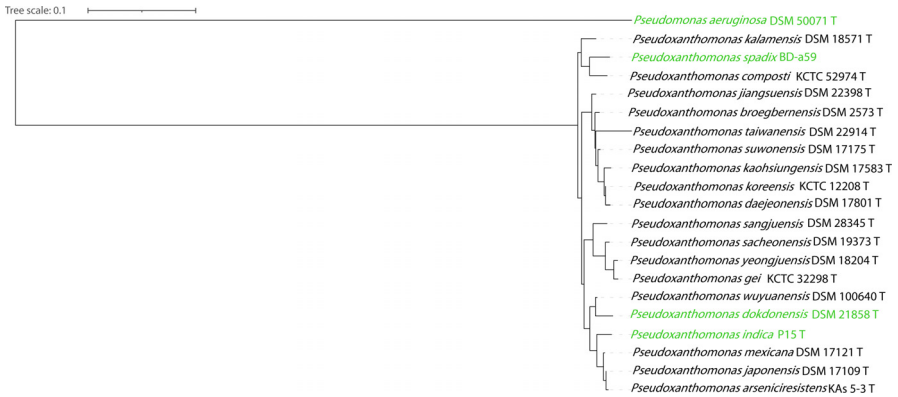


Fig. 1. Phylogenetic tree construction with maximum-likelihood method based on the 16S rRNA gene sequence of different species of genus *Pseudoxanthomonas*. *Pseudomonas aeruginosa* DSM 50071^T was used as outgroup. Genomes sequenced in the present study are in black and genomes from public repository are in green color. Bootstrap values are shown at the node of each cluster in blue color as percentage of 1000 replicates.

based phylogeny of all the twenty species of the genus *Pseudoxanthomonas* is depicted in Fig. 1. Whole genome sequence of the type strains of the genus *Pseudoxanthomonas* can be a valuable resource in taxonogenomics study of family *Lysobacteraceae* and its close relatives such as *Xanthomonas* and *Stenotrophomonas* [4,5]. Extreme environmental isolates such as *P. taiwanensis* [6] could be one of the key biotechnologically importance species to explore the heat stress mechanism. Genome resource of species of *P. broegbernensis*, *P. indica*, *P. kalamensis*, *P. kaohsiungensis*, *P. sacheonensis*, *P. spadix* and *P. jiangsuensis* [7–13] could be used for studying the stress tolerant genomics determinants.

2. Experimental Design, Materials and Methods

2.1. Bacterial strains and culture conditions

Type strains of the genus *Pseudoxanthomonas* were procured from two culture collection of Korean Collection for Type Cultures (KCTC) and The Leibniz Institute DSMZ-German Collection of Microorganisms and Cell Cultures (Table 1). Ampoules containing respective bacterial cultures were processed in the recommended media and condition in accordance with the bacterial strains collection.

2.2. Genome sequencing, assembly and annotation

Bacterial genomic DNA was extracted using ZR Fungal/Bacterial DNA MiniPrep Kit (Zymo Research, Irvine, CA, USA) and quantified using Qubit 2.0 Fluorometer (Thermo Fisher Scientific, Waltham, MA, USA). 1 ng of DNA sample was used in the preparation of Illumina sequencing libraries using Nextera XT sample preparation kit with dual indexing following provider’s instructions. Sequencing libraries were pooled and sequenced in-house on Illumina MiSeq platform with 2*250 bp paired-end sequencing kit.

The raw sequencing reads were assembled into the high-quality draft genome using SPAdes v3.10 [14] which is a de Bruijn graph-based assembler for the bacterial genome. Quality of the assembled genome was accessed using QUAST v4.4 [15] and overall coverage of the assembled genome was calculated using BBMap [16]. Presence of putative plasmid in the assembled

genome was accessed using plasmidSPAdes [17] with a minimum cut-off of 1Kb length. The assembled genomes were annotated using the NCBI prokaryotic genome annotation pipeline [18]. Assembly information with the putative number of plasmids is summarized in Table 1.

2.3. Phylogenetic assessment

Phylogenetic analysis based on the traditional 16S rRNA gene sequence was performed, for which 16S rRNA gene sequence was fetched from the respective assembled genome using from a standalone academic version of RNAmmer v1.2 [19] except for the type strains for species *P. spadix*, *P. helianthi* and *P. putridarboris*. 16S rRNA for these 3 species were taken from LPSN of the respective species definition. Multiple sequence alignment of 16S rRNA gene sequences was performed using ClustalW [20]. Phylogenetic tree based on Maximum Likelihood method with 1000 bootstrap replication was generated using MEGA v7.0.18 [21].

Ethics Statement

There is no ethical concern involved in the study.

Author Contributions

SK, SS and PPP have carried out strain procurement from culture collection and strain revival. KB and SK have performed whole genome sequencing and submission of assembled genomes to NCBI. PBP has conceived the study and participated in the design. All the authors have read and approved the manuscript.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships which have or could be perceived to have influenced the work reported in this article.

Data Availability

Genomic data resource of type strains of genus *Pseudoxanthomonas* (Original data) (NCBI genome).

Acknowledgments

All the work in the present study was supported by “Mega-genomic insights into co-evolution of rice and its Microbiome” (MICRA/MLP0020) and “GEAR- Genomic Evolutionary and Big-Data Analytic Strategies to Address Antimicrobial Resistance” (MLP0016).

References

- [1] K. Bansal, et al., Deep phylo-taxono genomics reveals *Xylella* as a variant lineage of plant associated *Xanthomonas* and supports their taxonomic reunification along with *Stenotrophomonas* and *Pseudoxanthomonas*, *Genomics* 113 (6) (2021) 3989–4003.
- [2] C. Kittiwongwattana, C. Thawai, *Pseudoxanthomonas helianthi* sp. nov., isolated from roots of Jerusalem artichoke (*Helianthus tuberosus*), *Int. J. Syst. Evol. Microbiol.* 66 (12) (2016) 5034–5038.

- [3] J.K. Lee, et al., *Pseudoxanthomonas putridarboris* sp. nov. isolated from rotten tree, *Int. J. Syst. Evol. Microbiol.* 67 (6) (2017) 1807–1812.
- [4] P.P. Patil, et al., Genome sequence of type strains of genus *Stenotrophomonas*, *Front. Microbiol.* 7 (2016) 309.
- [5] S. Kumar, et al., Phylogenomics insights into order and families of *Lysobacterales*, *Access Microbiol.* 1 (2) (2019) 1–9.
- [6] M.Y. Chen, et al., *Pseudoxanthomonas taiwanensis* sp. nov., a novel thermophilic, N₂O-producing species isolated from hot springs, *Int. J. Syst. Evol. Microbiol.* 52 (6) (2002) 2155–2161.
- [7] W. Finkmann, et al., Characterization of N₂O-producing *Xanthomonas*-like isolates from biofilters as *Stenotrophomonas nitritireducens* sp. nov., *Luteimonas mephitis* gen. nov., sp. nov. and *Pseudoxanthomonas broegbernsensis* gen. nov., sp. nov. *Int. J. Syst. Evol. Microbiol.* 50 (1) (2000) 273–282.
- [8] K. Kumari, et al., *Pseudoxanthomonas indica* sp. nov., isolated from a hexachlorocyclohexane dumpsite, *Int. J. Syst. Evol. Microbiol.* 61 (9) (2011) 2107–2111.
- [9] R.M. Harada, S. Campbell, Q.X. Li, *Pseudoxanthomonas kalamensis* sp. nov., a novel gammaproteobacterium isolated from Johnston Atoll, North Pacific Ocean, *Int. J. Syst. Evol. Microbiol.* 56 (5) (2006) 1103–1107.
- [10] J.S. Chang, et al., *Pseudoxanthomonas kaohsiungensis*, sp. nov., a novel bacterium isolated from oil-polluted site produces extracellular surface activity, *Syst. Appl. Microbiol.* 28 (2) (2005) 137–144.
- [11] D.S. Lee, et al., *Pseudoxanthomonas sacheonensis* sp. nov., isolated from BTEX-contaminated soil in Korea, transfer of *Stenotrophomonas dokdonensis* Yoon et al. (2006) to the genus *Pseudoxanthomonas* as *Pseudoxanthomonas dokdonensis* comb. nov. and emended description of the genus *Pseudoxanthomonas*, *Int. J. Syst. Evol. Microbiol.* 58 (9) (2008) 2235–2240.
- [12] C.C. Young, et al., *Pseudoxanthomonas spadix* sp. nov., isolated from oil-contaminated soil, *Int. J. Syst. Evol. Microbiol.* 57 (8) (2007) 1823–1827.
- [13] G.L. Wang, et al., *Pseudoxanthomonas jiangsuensis* sp. nov., a DDT-degrading bacterium isolated from a long-term DDT-polluted soil, *Curr. Microbiol.* 62 (6) (2011) 1760–1766.
- [14] A. Bankevich, et al., SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing, *J. Comput. Biol.* 19 (5) (2012) 455–477.
- [15] A. Gurevich, et al., QUAST: quality assessment tool for genome assemblies, *Bioinformatics* 29 (8) (2013) 1072–1075.
- [16] Bushnell, B., BBMap: a fast, accurate, splice-aware aligner. (2014). (No. LBNL-7065E). Lawrence Berkeley National Lab.(LBNL), Berkeley, CA (United States) doi: <https://www.osti.gov/servlets/purl/1241166>.
- [17] D. Antipov, et al., PlasmidSPAdes: assembling plasmids from whole genome sequencing data, *Bioinformatics* 32 (22) (2016) 3380–3387.
- [18] T. Tatusova, et al., NCBI prokaryotic genome annotation pipeline, *Nucleic Acids Res.* 44 (14) (2016) 6614–6624.
- [19] K. Lagesen, et al., RNAmmer: consistent and rapid annotation of ribosomal RNA genes, *Nucleic Acids Res.* 35 (9) (2007) 3100–3108.
- [20] M.A. Larkin, et al., Clustal W and Clustal X version 2.0, *Bioinformatics* 23 (21) (2007) 2947–2948.
- [21] S. Kumar, G. Stecher, K. Tamura, MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets, *Mol. Biol. Evol.* 33 (7) (2016) 1870–1874.
- [22] S.H. Lee, et al., Complete genome sequence of the BTEX-degrading bacterium *Pseudoxanthomonas spadix* BD-a59, *J. Bacteriol.* 194 (2) (2012) 544.