



Complete Genome Sequences of Two Novel Species from the *Pseudonocardiae* Family Isolated from the Persian Gulf

 Nasim Safaei,^a Boyke Bunk,^b Cathrin Spröer,^b Joachim Wink,^a  Yvonne Mast^b

^aDepartment of Microbial Strain Collection, Helmholtz Centre for Infection Research, Braunschweig, Germany

^bLeibniz Institute, German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany

ABSTRACT *Amycolatopsis* sp. strain DSM 110486 and *Pseudonocardia* sp. strain DSM 110487 are two novel actinomycete species that were isolated from Hengam Island beach sand from the Persian Gulf. Here, we present the complete genome sequences of DSM 110486 and DSM 110487, with sizes of 10.98 Mbp and 10.33 Mbp, respectively.

The family *Pseudonocardiae* includes several well-known natural-compound-producing genera, including *Amycolatopsis* and *Pseudonocardia*. In the context of a screening approach for novel natural compound producers, two new actinomycete strains (DSM 110486 and DSM 110487) have been isolated from a sand sample from Hengam Island, following the isolation method described by Safaei et al. (1). Hengam Island is a small island (37 km²) located south of Qeshm Island in the Persian Gulf (26.674191, 55.904151).

For DNA isolation, strains DSM 110486 and DSM 110487 were grown in glucose-yeast-malt (GYM) medium for 14 days at 30°C on a rotary shaker (160 rpm). DNA was isolated using a Genomic-tip 100/G (Qiagen, Hilden, Germany) according to the instructions of the manufacturer and was used for the preparation of long and short-read libraries. A SMRTbell template library for sequencing on the Sequel II platform was prepared according to the instructions from Pacific Biosciences (PacBio) (Menlo Park, CA, USA), following the procedure and checklist for preparing multiplexed microbial libraries using SMRTbell Express template preparation kit v2.0. Libraries for sequencing on the NextSeq 500 platform were prepared using the Nextera XT DNA library preparation kit (Illumina, San Diego, CA, USA) with modifications according to Baym et al. (2). Preassembly read quality control (QC) was performed using single-molecule real-time (SMRT) Link (Run QC module) and FastQC (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc>). Long-read genome assembly was performed using the Microbial Assembly protocol included in SMRT Link v10.1.0, applying target genome sizes of 10.9 and 10.3 Mbp for DSM 110486 and DSM 110487, respectively. The assembly process resulted in one circular chromosomal contig and one linear plasmid each. Genome completeness was retrieved from the Microbial Assembly results in terms of circular replicons or blunt-end linear replicons, as well as 0% missing bases. Error correction was performed by mapping of Illumina short reads onto the completed genome using BWA (3), with subsequent determination of a new consensus sequence (<https://github.com/JHartlich/AlternateReferenceMaker>). Genome annotation was performed with the NCBI Prokaryotic Genome Annotation Pipeline (PGAP). Default parameters were used for all software. All genome-sequence-related data are summarized in Table 1.

Phylogenomic analysis of the full-length genome sequences with the Type (Strain) Genome Server (TYGS) v1.0 (<https://tygs.dsmz.de>) (4) revealed that DSM 110486 and DSM 110487 are most closely related to the type strains *Amycolatopsis saalfeldensis* DSM

Editor J. Cameron Thrash, University of Southern California

Copyright © 2021 Safaei et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Yvonne Mast, yvonne.mast@dsmz.de.

The authors declare no conflict of interest.

Received 8 October 2021

Accepted 23 November 2021

Published 16 December 2021

TABLE 1 Sequencing and annotation data

Parameter	Data for:	
	DSM 110486	DSM 110487
Genome length (bp)	10,984,156	10,329,491
No. of chromosomes	1	1
No. of extrachromosomal elements	1	1
Genome coverage (×)	533	343
Avg G+C content (%)	70.2	71.8
No. of coding sequences	10,508	9,666
No. of tRNAs	70	47
No. of rRNAs	12	6
No. of PacBio reads	156,372	88,199
PacBio read N_{50} (bp)	40,948	65,481
No. of Illumina reads	3,002,725	4,701,445
Illumina paired-end read length (bp)	2 × 151	2 × 151
GenBank accession no.		
Chromosome	CP080519.1	CP080521.1
Plasmid	CP080520.1	CP080522.1
BioProject no.	PRJNA751440	PRJNA751442
SRA accession no.		
PacBio reads	SRX12487219	SRX12487266
Illumina reads	SRX12487220	SRX12487267

44993 and *Pseudonocardia hierapolitana* DSM 45671 (Fig. 1), with digital DNA-DNA hybridization (dDDH) values (formula d_4) of 27.2% and 36.4%, respectively, suggesting that both strains represent novel species. AntiSMASH v6.0 analysis (5) with the genome sequences of DSM 110486 and DSM 110487 led to the identification of 23 and 16 biosynthetic gene clusters, respectively, revealing the genetic potential of these strains for the production of novel natural products.

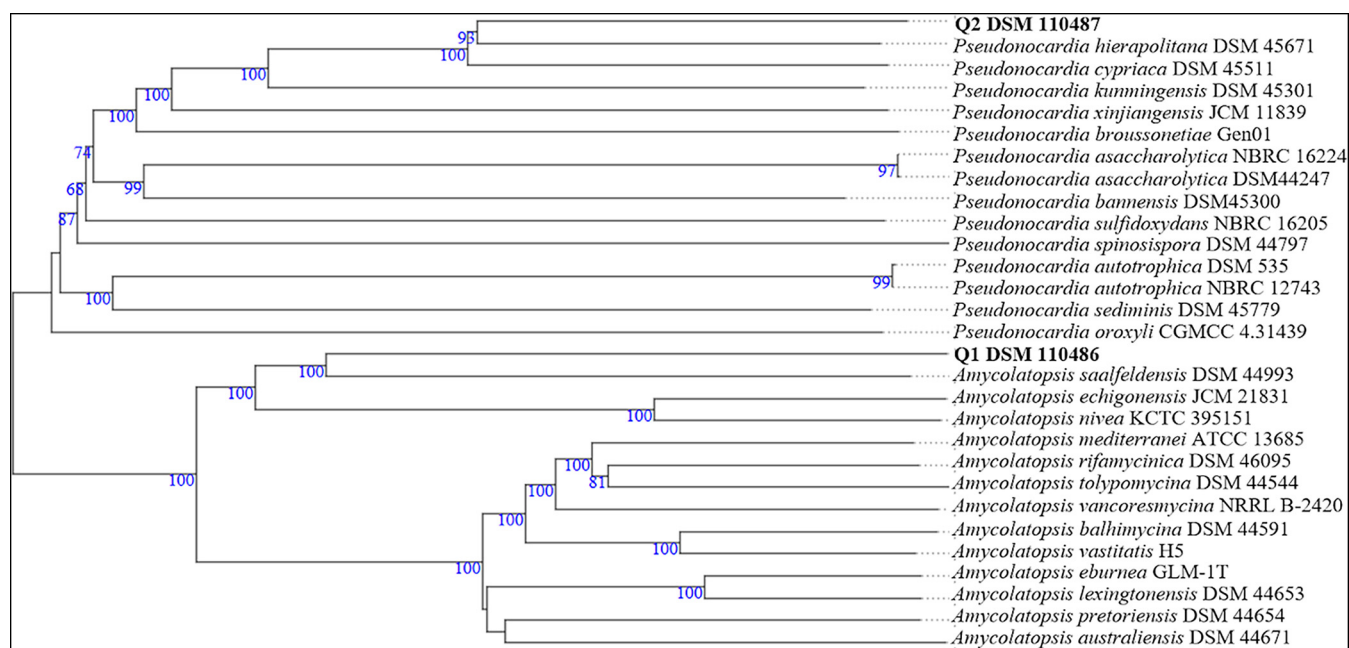


FIG 1 Whole-genome sequence tree generated with the TYGS web server for strains DSM 110486 and DSM 110487 and closely related species. The tree was inferred with FastME from Genome BLAST Distance Phylogeny (GBDP) distances calculated from the genome sequences. The branch lengths are scaled in terms of GBDP distance formula d_5 . The numbers above the branches are GBDP pseudo-bootstrap support values of >60% from 100 replications, with an average branch support value of 90.3%. The tree was rooted at the midpoint.

Data availability. Genome sequence-related data availability is listed in Table 1.

ACKNOWLEDGMENTS

We thank Nicole Heyer (Leibniz Institute, DSMZ) for excellent technical assistance and Katharina Huber (Leibniz Institute, DSMZ) for recommendations on biomass production and editing.

REFERENCES

1. Safaei N, Nouioui I, Mast Y, Zaburanyi N, Rohde M, Schumann P, Müller R, Wink J. 2021. *Kibdelosporangium persicum* sp. nov., a new member of the Actinomycetes from a hot desert in Iran. *Int J Syst Evol Microbiol* 71: 004625. <https://doi.org/10.1099/ijsem.0.004625>.
2. Baym M, Kryazhimskiy S, Lieberman TD, Chung H, Desai MM, Kishony RK. 2015. Inexpensive multiplexed library preparation for megabase-sized genomes. *PLoS One* 10:e0128036. <https://doi.org/10.1371/journal.pone.0128036>.
3. Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>.
4. Meier-Kolthoff JP, Göker M. 2019. TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy. *Nat Commun* 10: 2182. <https://doi.org/10.1038/s41467-019-10210-3>.
5. Blin K, Shaw S, Kloosterman AM, Charlop-Powers Z, Van Wezel GP, Medema MH, Weber T. 2021. AntiSMASH 6.0: improving cluster detection and comparison capabilities. *Nucleic Acids Res* 49:W29–W35. <https://doi.org/10.1093/nar/gkab335>.