![American Society for Microbiology | Microbiology Resource Announcements](logos)

Complete Genome Sequence of *Amycolatopsis* sp. CA-230715, Encoding a 35-Module Type I Polyketide Synthase

Tue S. Jørgensen,[a] Tetiana Gren,[a] Daniel Oves-Costales,[b] Francisco Javier Ortiz-López,[b] Daniel Carretero-Molina,[b] Ignacio González,[b] Kai Blin,[a] Olga Genilloud,[b] Tilmann Weber[a]

aThe Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kongens Lyngby, Denmark
bFundación MEDINA, Centro de Excelencia en Investigación de Medicamentos Innovadores en Andalucía, Granada, Spain

Tue S. Jørgensen and Tetiana Gren contributed equally to this work. Author order was determined alternatingly with other equally contributed MRA manuscripts.

ABSTRACT We report the sequencing, assembly, and annotation of the genome of *Amycolatopsis* sp. CA-230715, a potentially interesting producer of natural products. The genome of CA-230715 was sequenced using PacBio, Illumina, and Nanopore technologies. It consists of a circular 10,363,158-nucleotide (nt) chromosome and a circular 12,080-nt plasmid.

The genus *Amycolatopsis* is a recognized source of secondary metabolites (1–3). Only a few complete gapless genome sequences from the *Amycolatopsis* genus exist in public databases. Here, we report the sequencing of *Amycolatopsis* sp. CA-230715, which was identified in antimicrobial screening against *Acinetobacter baumannii* MB5973. The strain was isolated from a soil sample collected in Berbérati (Central African Republic). The original colony was isolated from a serial dilution of a soil suspension plated onto HANOB medium (6.25 g/liter NaNO$_3$, 2.5 g/liter K$_2$HP$_4$, 0.65 g/liter MgSO$_4$, 1.25 g/liter humic acid, 0.020 g/liter benomyl, pH 7) after incubation for 5 weeks at 28°C/70% relative humidity. For DNA isolation, the strain was grown in liquid yeast extract-malt extract (YEME) medium (4). DNA was isolated using the Genomic-tip G100 kit (Qiagen, Venlo, Netherlands). PacBio RS II (Pacific Biosciences, Menlo Park, CA, USA) data were generated by Macrogen Inc. (Seoul, South Korea; DNA/polymerase binding kit P6; SMRT cell 8Pac v3 using g-TUBE-sheared and Blue Pippin-size-selected DNA), yielding 125,283 subreads with an N_{50} value of 16,071 nucleotides (nt). Subread generation and adapter removal were performed using SMRT Analysis v2.3 software. A KAPA HyperPlus library was sequenced on an Illumina MiSeq instrument (San Diego, CA, USA), yielding 4,477,879 read clusters (2 × 150 nt), totaling 1,273,625,221 nt. Nanopore data were generated on a MinION device using the SQK-RBK004 kit and a FLO-MIN106D R9.4 Rev-D flow cell (183,945 reads; N_{50}, 10,904 nt; total, 1,065,369,647 nt) (Oxford Nanopore Technologies, Oxford, UK). Default software parameters were used except where otherwise noted. The Illumina reads were adapter and quality trimmed using AdapterRemoval2 v2.1.7 (5) with the parameters --trimns --trimqualities. The Nanopore reads were demultiplexed and base called using Guppy v3.0.3, adapter trimmed using Porechop v0.2.4 (6), and assembled using Flye v2.4.1-geb89c9e (7). This assembly was then polished with the Illumina reads using the polishing module in Unicycler v0.4.7 (8) and used in a second round of assembly with Unicycler v0.4.7 (running SPAdes v3.13.0) (9), which combined the initial polished assembly with the Illumina and PacBio data sets. The contiguity and circular topology of the 10,363,158-nt chromosome (GC content, 69.7%) and circular 12,080-nt plasmid (GC content, 67.6%) were evaluated using the assembly graph from Unicycler and Bandage v0.8.1 (10). The circular chromosome was not rotated. The assembly had a BUSCO v3.1.0 (actinobacteria_odb9) (11) score of 100% complete genes (352 genes; 5 in duplicate). The genome sequence was annotated using Prokka v1.14.0 (12) with additional databases as described in reference 13. A total of 9 rRNAs and 104 tRNAs were found in

the annotation, as well as 9,465 coding DNA sequences (CDS), of which 6,405 (68%) were functionally annotated.

According to antiSMASH v6.0.0 (14), the strain harbors potentially novel biosynthetic gene clusters (BGCs), including a 35-module, 216-kb type I polyketide synthase BGC, one of the largest uninterrupted bacterial BGCs reported.

AutoMLST analysis (15) of the genome sequence has shown that *Amycolatopsis* sp. CA-230715 has 82.1% average nucleotide identity (ANI) similarity to *Amycolatopsis nigrescens* CSC17Ta-90, and GTDB-tk v1.5.1, R202 (16), places the strain within the *Amycolatopsis* genus.

Data availability. All data are available under BioProject accession number PRJNA639419. The raw reads have been deposited at the SRA under accession numbers SRR12367306 (Illumina), SRR12367305 (PacBio), and SRR12367307 (Nanopore). The GenBank accession numbers are CP059997.1 (chromosome) and CP059998.1 (plasmid).

REFERENCES

1. Song Z, Xu T, Wang J, Hou Y, Liu C, Liu S, Wu S. 2021. Secondary metabolites of the genus *Amycolatopsis*: structures, bioactivities and biosynthesis. Molecules 26:1884. https://doi.org/10.3390/molecules26071884.
2. Edenhart S, Denneler M, Spohn M, Doskocil E, Kavšček M, Amon T, Kosec G, Smole J, Bardl B, Biermann M, Roth M, Wohlleben W, Stegmann E. 2020. Metabolic engineering of *Amycolatopsis japonicum* for optimized production of [S,S]-EDDS, a biodegradable chelator. Metab Eng 60:148–156. https://doi.org/10.1016/j.ymben.2020.04.003.
3. Primahana G, Risdian C, Mozef T, Wink J, Surup F, Stadler M. 2021. Amycolatomycins A and B, cyclic hexapeptides isolated from an *Amycolatopsis* sp. 195334CR. Antibiotics 10:261. https://doi.org/10.3390/antibiotics10030261.
4. Kieser T, Bibb MJ, Buttner MJ, Chater KF, Hopwood DA. 2000. Practical Streptomyces genetics. John Innes Foundation, Norwich, United Kingdom.
5. Schubert M, Lindgreen S, Orlando L. 2016. AdapterRemoval v2: rapid adapter trimming, identification, and read merging. BMC Res Notes 9:88. https://doi.org/10.1186/s13104-016-1900-2.
6. Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Completing bacterial genome assemblies with multiplex MinION sequencing. Microb Genom 3:e000132. https://doi.org/10.1099/mgen.0.000132.
7. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone reads using repeat graphs. Nat Biotechnol 37:540–546. https://doi.org/10.1038/s41587-019-0072-8.
8. Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput Biol 13:e1005595. https://doi.org/10.1371/journal.pcbi.1005595.
9. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477. https://doi.org/10.1089/cmb.2012.0021.
10. Wick RR, Schultz MB, Zobel J, Holt KE. 2015. Bandage: interactive visualization of de novo genome assemblies. Bioinformatics 31:3350–3352. https://doi.org/10.1093/bioinformatics/btv383.
11. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31:3210–3212. https://doi.org/10.1093/bioinformatics/btv351.
12. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068–2069. https://doi.org/10.1093/bioinformatics/btu153.
13. Gren T, Jørgensen TS, Whitford CM, Weber T. 2020. High-quality sequencing, assembly, and annotation of the *Streptomyces griseofuscus* DSM 40191 genome. Microbiol Resour Announc 9:e01100-20. https://doi.org/10.1128/MRA.01100-20.
14. 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.
15. Alanjary M, Steinke K, Ziemert N. 2019. AutoMLST: an automated Web server for generating multi-locus species trees highlighting natural product potential. Nucleic Acids Res 47:W276–W282. https://doi.org/10.1093/nar/gkz282.
16. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. 2020. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. Bioinformatics 36:1925–1927. https://doi.org/10.1093/bioinformatics/btz848.