Complete Genome Sequence of *Streptomyces* sp. Strain CA-256286

Charlotte Beck,[a] Tetiana Gren,[a] Tue S. Jørgensen,[a] Ignacio González,[b] Fernando Román-Hurtado,[b] Daniel Oves-Costales,[b] Olga Genilloud,[b] Tilmann Weber[a]

[a]The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kongens Lyngby, Denmark
[b]Fundación MEDINA, Granada, Spain

Charlotte Beck, Tetiana Gren, and Tue S. Jørgensen contributed equally to this work. The author order was determined alphabetically.

ABSTRACT Here, we report the sequencing, assembly, and annotation of the genome of *Streptomyces* sp. strain CA-256286. The genome consists of a linear 7,726,360-nucleotide chromosome and a linear 466,817-nucleotide putative plasmid. This strain is predicted to produce a range of novel secondary metabolites.

*S*treptomyces sp. strain CA-256286 was isolated from a soil sample collected at Pilar de la Mola (Formentera, Spain) from a muddy torrent and was treated with dry heat (100°C for 1 h). The original colony was isolated from a soil serial dilution suspension plated onto M3 medium supplemented with nalidixic acid (20 mg/liter) (1), after incubation for 6 weeks at 28°C in 70% relative humidity. The strain was grown in yeast extract-malt extract (YEME) medium for DNA isolation according to the method described by Kieser et al. (2). The genomic DNA for sequencing was purified using the Genomic-tip 100 kit (Qiagen, Venlo, Netherlands). The genomic DNA was sheared using a g-TUBE (Covaris, Inc., Woburn, MA, USA), followed by BluePippin size selection (Sage Science, Beverly, MA, USA). Macrogen, Inc. (Seoul, South Korea), generated the RS II (Pacific Biosciences [PacBio], Menlo Park, CA, USA) data (87,704 subreads, with an N_{50} value of 18,359 nucleotides [nt]) using the single-molecule real-time (SMRT) Cell 8PAC v3, the DNA polymerase binding kit P6, and two SMRT cells. PacBio polymerase reads were partitioned to subreads by Macrogen using SMRT Analysis v2.3 (PacBio) (http://www.pacb.com/support/software-downloads), which removes adapter-related sequences. Default software parameters were used except where otherwise noted. Flye v2.8 (3) was used to assemble the PacBio subreads with the parameter --iterations 5 for a total of five rounds of polishing using the PacBio data, also including quality filtering, which excluded subreads shorter than 5,000 nt. The resulting assembly consists of two contigs. A 7,726,360-nt chromosome sequence (GenBank accession number CP071044) with inverted repeat chromosome ends of 26,587 nt, which is characteristic of *Streptomyces* strains, was identified. In addition, a 466,817-nt linear putative plasmid (GenBank accession number CP071045) was found. Both contigs were found to be linear, based on Bandage v0.8.1 (4) visualization of the Flye assembly graph. The coverages for the chromosome and the putative plasmid are 108× and 95×, respectively, indicating that the copy number for the putative plasmid is 1. The genome of *Streptomyces* sp. strain CA-256286 was annotated using Prokka v1.14.6 (5), with the switches --cdsrnaolap --rnammer (6) --increment 10. In addition to default databases, the Pfam-A v32.0 database was used along with the genomes of six actinobacterial species with manual or otherwise high-quality annotations (see reference 7 for details). The five genes closest to the chromosome ends all lack functional annotation. The chromosome harbors 6,744 protein-coding genes, 6 rRNA operons, and 79 tRNAs, and the putative plasmid contains 466 protein-coding genes. The GC content of the chromosome is 71.6%, and that of the plasmid is 68.4%. Our initial assessment of the production capabilities of *Streptomyces* sp. strain CA-256286 was supported by

genome mining analyses with antiSMASH v6.0.0alpha1-60bffdb (8). In total, 41 regions were predicted, covering 6 polyketide synthase (PKS), 5 nonribosomal peptide synthetase (NRPS), 8 PKS/NRPS, 9 terpene, 1 melanin, 4 lassopeptide, 1 lanthipeptide, 2 butyrolactone, 2 siderophore, and 3 ectoine biosynthetic gene clusters. According to autoMLST (9), the closest relative is *Streptomyces* sp. NRRL S-623 (NCBI assembly accesion number GCF_000725705.1) with 99.4% estimated average nucleotide identity.

Data availability. Raw PacBio data and the assembly and annotation of the CA-256286 genome are available via BioProject PRJNA689435. The NCBI GenBank accession numbers are CP071044 (chromosome) and CP071045 (plasmid).

ACKNOWLEDGMENT

REFERENCES

1. Mincer TJ, Jensen PR, Kauffman CA, Fenical W. 2002. Widespread and persistent populations of a major new marine actinomycete taxon in ocean sediments. Appl Environ Microbiol 68:5005–5011. https://doi.org/10.1128/AEM.68.10.5005-5011.2002.
2. Kieser T, Bibb MJ, Buttner MJ, Chater KF, Hopwood DA. 2000. Practical Streptomyces genetics. John Innes Foundation, Norwich, United Kingdom.
3. 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.
4. 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.
5. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068–2069. https://doi.org/10.1093/bioinformatics/btu153.
6. Lagesen K, Hallin P, Rødland EA, Staerfeldt H-H, Rognes T, Ussery DW. 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res 35:3100–3108. https://doi.org/10.1093/nar/gkm160.
7. Gren T, Jørgensen TS, Whitford CM, Weber T. 2020. High-quality sequencing, assembly, and annotation of the *Streptomyces griseofuscus* DSM 40191. Microbiol Resour Announc 9:e01100-20. https://doi.org/10.1128/MRA.01100-20.
8. Blin K, Shaw S, Steinke K, Villebro R, Ziemert N, Lee SY, Medema MH, Weber T. 2019. antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. Nucleic Acids Res 47:W81–W87. https://doi.org/10.1093/nar/gkz310.
9. 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.