



Complete Genome Sequence of *Nocardia* sp. Strain CS682, a Producer of Antibacterial Compound Nargenicin A1

 Divesh Dhakal,^a Vijay Rayamajhi,^a Hue Thi Nguyen,^a Purna Bahadur Poudel,^a Jae Kyung Sohng^{a,b}

^aDepartment of Life Science and Biochemical Engineering, Sun Moon University, Asan-si, Chungnam, Republic of Korea

^bDepartment of Pharmaceutical Engineering and Biotechnology, Sun Moon University, Asan-si, Chungnam, Republic of Korea

ABSTRACT *Nocardia* sp. strain CS682 is a rare actinobacterium with a promising ability to produce secondary metabolites such as nargenicin A1 (an effective antibacterial compound) and IBR-3 (a UV-protectant molecule). Here, we report the complete genome sequence of *Nocardia* sp. CS682, obtained by PacBio sequencing as a single contig with 8,919,230 bp (GC content, 63.3%).

Nocardia spp. are rare actinobacteria that are generally characterized as a source of diverse biomolecules with antibacterial, antifungal, cytotoxic, and immunosuppressive activities (1, 2). The major secondary metabolite derived from *Nocardia* sp. strain CS682 is nargenicin A1, an antibacterial polyketide with promising activity against various Gram-positive pathogenic bacteria (3). Similarly, IBR-3, a novel tetrahydroxynaphthalene (THN) derivative with UV protection effects, was also isolated from strain CS682 (4).

Strain CS682 was isolated from soil samples from Jeonnam, Republic of Korea, while screening for bacterial strains producing potential antibiotics against methicillin-resistant *Staphylococcus aureus* (MRSA) (3). For genomic extraction, strain CS682 was cultured in brain heart infusion (BHI) medium at 37°C and 200 rpm for 5 days (5). The genomic DNA of *Nocardia* sp. CS682 was isolated and purified using the TIANamp bacterial DNA kit. A PacBio SMRTbell library was constructed and sequenced on the PacBio RS platform (Macrogen, Inc., Republic of Korea) (6, 7). The Hierarchical Genome Assembly Process (HGAP) algorithm was executed in the Web-based graphical user interface (GUI) in the SMRT Analysis pipeline version 2.3.0 for preparing the PacBio library and for high-quality *de novo* assembly of the genome from PacBio raw reads. There were 152,943 sequence reads with a mean subread length of 9,897 bp, resulting in 189-fold sequencing coverage of the genome. Furthermore, a consensus sequence of higher quality was generated by validating the mapping reads against assembled contigs and error correcting using Quiver version 2.1.0 (8). Finally, the CS682 genome was assembled into a single contig. The genome was annotated with the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) version 4.5 (9). The autoMLST pipeline (<https://automlst.ziemertlab.com/download>) (10) was used for determining closely related genomes based on alignment of core genes, and the closest species were determined based on percent average nucleotide identity (ANI). For determining the biosynthetic gene clusters (BGCs) in the genome, antiSMASH bacterial version 4.0 (11) was used. Default parameters were used for all software unless otherwise specified.

The assembled genome of strain CS682 was obtained as a single contig with 8,919,230 bp (GC content, 63.3%). A total of 7,856 open reading frames (ORFs), 56 tRNAs, 3 rRNA genes (5S, 16S, and 23S rRNA), and 3 noncoding RNA (ncRNA) genes were identified (Table 1). The estimated ANI values of *Nocardia* sp. CS682 with different *Nocardia* spp. were 86.7% (*Nocardia altamirensis* NBRC 108246; RefSeq assembly accession no. [GCF_001612685](https://.ncbi.nlm.nih.gov/assembly/GCF_001612685)), 86.7% (*Nocardia brasiliensis* ATCC 700358; [GCF_000250675](https://ncbi.nlm.nih.gov/assembly/GCF_000250675)), and 86.7% (*Nocardia brasiliensis* NBRC 14402; [GCF_000308475](https://ncbi.nlm.nih.gov/assembly/GCF_000308475)). These observations

Citation Dhakal D, Rayamajhi V, Nguyen HT, Poudel PB, Sohng JK. 2019. Complete genome sequence of *Nocardia* sp. strain CS682, a producer of antibacterial compound nargenicin A1. *Microbiol Resour Announc* 8:e01098-19. <https://doi.org/10.1128/MRA.01098-19>.

Editor David Rasko, University of Maryland School of Medicine

Copyright © 2019 Dhakal 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 Divesh Dhakal, medivesh@gmail.com, or Jae Kyung Sohng, sohng@sunmoon.ac.kr.

Received 5 September 2019

Accepted 24 October 2019

Published 5 December 2019

TABLE 1 Major genomic features of *Nocardia* sp. CS682

Genomic feature	Data
Genome size (bp)	8,919,230
GC content (%)	67.3
Contig N_{50} (bp)	14,297
No. of reads	152,943
No. of predicted ORFs	7,856
No. of predicted tRNA genes	56
Total no. of predicted rRNA genes (5S, 16S, and 23S)	3
No. of predicted pseudogenes	182
No. of predicted ncRNA genes	3
No. of predicted CRISPR arrays	4
BioProject identifier	PRJNA473642
BioSample identifier	SAMN09280253
Sequence Read Archive accession no.	SRX6973800, SRX6973801, and SRX6973802
GenBank genome accession no.	CP029710
No. of gene clusters predicted by antiSMASH	44

establish that *Nocardia* sp. CS682 has higher similarities to *Nocardia altamirensis* and *Nocardia brasiliensis* than to other species. The antiSMASH 4.0 results revealed that strain CS682 contains 44 putative BGCs, including clusters for polyketides, nonribosomal peptides, ectoine, terpenes, etc. The genomic architecture of strain CS682 also provided insight on precise locations of BGCs of nargenicin A1 (2) and IBR-3 (3) within the genome.

Data availability. The genome sequence of *Nocardia* sp. strain CS682 can be accessed in the NCBI database under accession no. CP029710. The PacBio raw reads have been submitted to the Sequence Read Archive (SRA) under accession no. SRX6973800, SRX6973801, and SRX6973802.

ACKNOWLEDGMENT

This work was supported by National Research Foundation of Korea (NRF) grants funded by the South Korean government (MEST) (grants NRF-2017R1A2A2A05000939 to J.K.S. and NRF-2017R1D1A1B03036273 to D.D.).

REFERENCES

- Dhakal D, Pokhrel AR, Shrestha B, Sohng JK. 2017. Marine rare actinobacteria: isolation, characterization, and strategies for harnessing bioactive compounds. *Front Microbiol* 8:1106. <https://doi.org/10.3389/fmicb.2017.01106>.
- Dhakal D, Rayamajhi V, Mishra R, Sohng JK. 2019. Bioactive molecules from *Nocardia*: diversity, bioactivities and biosynthesis. *J Ind Microbiol Biotechnol* 46:385–407. <https://doi.org/10.1007/s10295-018-02120-y>.
- Sohng JK, Yamaguchi T, Seong CN, Baik KS, Park SC, Lee HJ, Jang SY, Simkhada JR, Yoo JC. 2008. Production, isolation and biological activity of nargenicin from *Nocardia* sp. CS682. *Arch Pharm Res* 31:1339–1345. <https://doi.org/10.1007/s12272-001-2115-0>.
- Mishra R, Dhakal D, Han JM, Lim HN, Jung HJ, Yamaguchi T, Sohng JK. 2019. Production of a novel tetrahydroxynaphthalene (THN) derivative from *Nocardia* sp. CS682 by metabolic engineering and its bioactivities. *Molecules* 24:244. <https://doi.org/10.3390/molecules24020244>.
- Dhakal D, Sohng JK. 2015. Laboratory maintenance of *Nocardia* species. *Curr Protoc Microbiol* 39:10F.1.1–10F.1.8. <https://doi.org/10.1002/9780471729259.mc10f01s39>.
- Rhoads A, Au KF. 2015. PacBio sequencing and its applications. *Genomics Proteomics Bioinformatics* 13:278–289. <https://doi.org/10.1016/j.gpb.2015.08.002>.
- Montecillo AD, Raymundo AK, Papa IA, Aquino GMB, Jacildo AJ, Stothard P, Rosana A. 2018. Near-complete genome sequence of *Ralstonia solanacearum* T523, a phylotype I tomato phytopathogen isolated from the Philippines. *Microbiol Resour Announc* 7:e01048-18. <https://doi.org/10.1128/MRA.01048-18>.
- Chin CS, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE, Turner SW, Korlach J. 2013. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat Methods* 10:563–569. <https://doi.org/10.1038/nmeth.2474>.
- Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J. 2016. NCBI Prokaryotic Genome Annotation Pipeline. *Nucleic Acids Res* 44:6614–6624. <https://doi.org/10.1093/nar/gkw569>.
- 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>.
- Blin K, Wolf T, Chevrette MG, Lu X, Schwalen CJ, Kautsar SA, Suarez Duran HG, de Los Santos ELC, Kim HU, Nave M, Dickschat JS, Mitchell DA, Shelest E, Breitling R, Takano E, Lee SY, Weber T, Medema MH. 2017. antiSMASH 4.0—improvements in chemistry prediction and gene cluster boundary identification. *Nucleic Acids Res* 45:W36–W41. <https://doi.org/10.1093/nar/gkx319>.