



Complete Genome Sequence of the Collinolactone Producer *Streptomyces* sp. Strain Gö40/10

Eva B. Sterndorff,^a  Tue S. Jørgensen,^a Jaime Felipe Guerrero Garzón,^a Julian Schmid,^b Stephanie Grond,^b  Tilmann Weber^a

^aThe Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kongens Lyngby, Denmark

^bInstitut für Organische Chemie, Eberhard Karls Universität Tübingen, Tübingen, Germany

ABSTRACT The actinomycete *Streptomyces* sp. strain Gö40/10 has the potential to produce a range of secondary metabolites, one of which is collinolactone, a compound with neuroprotective properties and potential for pharmaceutical applications. The genome was sequenced with Oxford Nanopore Technologies MinION and Illumina MiSeq systems and consists of a single 9,635,564-nucleotide linear chromosome.

Streptomyces sp. strain Gö40/10 was isolated from soil collected in Bolivia in 1991 and grown in medium NL 8711, which is used routinely for agar slopes (1). This isolate has the ability to produce various complex secondary/specialized metabolites, including collinolactone (2–4), which exhibits neuroprotective properties through inhibition of intracellular oxidative stress on glutamate-sensitive cells and hence has potential as a drug against Alzheimer’s disease (4). Furthermore, *Streptomyces* sp. strain Gö40/10 was described to produce cineromycins and ansamycins (3) and potentially other complex polyketides or peptides. To get access to the biosynthetic gene clusters (BGCs) encoding these biosynthesis pathways, the whole-genome sequence of *Streptomyces* sp. strain Gö40/10 was determined.

The strain was grown in yeast extract-malt extract (YEME) medium with 3.4% sucrose at 30°C with shaking at 250 rpm (5). Purification of genomic DNA (gDNA) was performed by using the Qiagen Genomic-tip 100 kit (Venlo, Netherlands) with prolonged lysis, i.e., 2 h of incubation with B1 and enzyme at 37°C followed by 2 h with B2 at 50°C.

A library was constructed using the rapid barcoding kit (SQK-RBK004; Oxford Nanopore Technologies, Inc.) and run with a FLO-MIN106D (R9.4.1) flow cell on a MinION device. Base calling was done using Guppy base caller (v5.0.11+2b6dbff) in high-accuracy mode followed by demultiplexing by the Guppy barcoder. Adaptors were trimmed from Nanopore reads using Porechop (v0.2.4) (6), and reads smaller than 1,000 nucleotides (nt) were removed using FilTlong (v0.2.0) (7). After trimming, reads with a total of 930,546,501 nt and an N_{50} value of 16,263 nt were used for *de novo* assembly with Flye (v2.9-b1768) with the setting --nano-hq and a total of five polishing iterations (8).

Illumina MiSeq data were generated from a KAPA HyperPlus library (Roche, Switzerland), with size selection of approximately 450 to 600 nt, resulting in 2.3 million clusters of read length 2×151 bp. Illumina reads were trimmed using Trim Galore with Cutadapt (v2.10), applying the settings --length 100 and --quality 20 (9).

Illumina data were aligned to the Nanopore assembly using bowtie2-align (v2.3.4.1), with an overall alignment rate of 97.08% (10). Finally, the Nanopore assembly was polished with the Illumina data using the Unicycler (v0.4.8) polishing module, resulting in 4,429 changes in base positions (11).

The topology of the genome was determined from the Flye assembly graph and consisted of a single, linear chromosome with a total assembly length of 9,635,564 nt with terminal inverted repeats of 59,464 nt, a GC content of 72.33%, and mean Nanopore coverage of $95\times$. Gö40/10 was classified by GTDB-tk (v1.7.0) (12) as a member of

Editor Irene L. G. Newton, Indiana University, Bloomington

Copyright © 2022 Sterndorff 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 Tilmann Weber, tiwe@biosustain.dtu.dk, or Stephanie Grond, stephanie.grond@uni-tuebingen.de.

The authors declare no conflict of interest.

Received 7 March 2022

Accepted 8 July 2022

Published 27 July 2022

Streptomyces cinerochromogenes, and the closest autoMLST (v231601f) (13) match was *Streptomyces reticuli* (GenBank accession number [LN997842.1](https://doi.org/10.1093/nar/gkab335)) at 92.8% average nucleotide identity (ANI). The benchmarking universal single-copy orthologs (BUSCO) score (14) was 100% complete BUSCO genes (v4.0.5, actinobacteria_class_odb10), with 3 genes existing in duplicate. PGAP (v2021-11-29.build5742) was used for annotation and yielded 8,226 coding sequences (CDSs), of which 7,950 were functionally annotated, with 7 rRNA operons and 72 tRNAs.

Analysis with antiSMASH (v6.0.alpha1-0bc1c66) (15) with the parameters --cb-general --cb-subclusters, and --cb-knownclusters revealed 39 regions containing secondary metabolite BGCs, including 3 non-ribosomal peptide synthetase (NRPS), 7 polyketide, 4 terpene, and 6 hybrid BGCs.

Data availability. The data are available under NCBI BioProject accession number [PRJNA721311](https://doi.org/10.1093/nar/gkab335). Raw reads were deposited in the SRA under BioSample accession numbers [SAMN18970196](https://doi.org/10.1093/nar/gkab335) (Illumina) and [SAMN18970196](https://doi.org/10.1093/nar/gkab335) (Nanopore). The genome sequence can be found under GenBank accession number [CP084203.1](https://doi.org/10.1093/nar/gkab335).

ACKNOWLEDGMENTS

We thank Oliwia Vuksanovic and Alexandra Hoffmeyer for their invaluable technical assistance.

This work was funded by grants NNF16OC0021746 and NNF20CC0035580 from the Novo Nordisk Foundation.

REFERENCES

1. Schuhmann T. 2005. Untersuchungen zur Biosynthese und Aktivität ausgewählter Plecomakrolide sowie chemisches Screening von Actinomyceten. Dr. rer. nat. dissertation. Georg August University of Göttingen, Göttingen, Germany.
2. Bode HB, Bethe B, Höfs R, Zeeck A. 2002. Big effects from small changes: possible ways to explore nature's chemical diversity. *ChemBioChem* 3:619–627. [https://doi.org/10.1002/1439-7633\(20020703\)3:7<619::AID-CBIC619>0.CO;2-9](https://doi.org/10.1002/1439-7633(20020703)3:7<619::AID-CBIC619>0.CO;2-9).
3. Schiewe HJ, Zeeck A. 1999. Cineromycins, γ -butyrolactones and ansamycins by analysis of the secondary metabolite pattern created by a single strain of *Streptomyces*. *J Antibiot (Tokyo)* 52:635–642. <https://doi.org/10.7164/antibiotics.52.635>.
4. Schmid JC, Frey K, Scheiner M, Garzon JFG, Stafforst L, Fricke JN, Schuppe M, Schiewe H, Zeeck A, Weber T, Uson I, Kemkemer R, Decker M, Grond S. 2021. The structure of cyclodecatriene collinolactone, its biosynthesis and semisynthetic analogues: effects of monoastrial phenotype and protection from intracellular oxidative stress. *Angew Chem Int Ed Engl* 60:23212–23216. <https://doi.org/10.1002/anie.202106802>.
5. Kieser T, Bibb MJ, Buttner MJ, Chater K. 2000. Practical *Streptomyces* genetics. John Innes Foundation, Norwich, United Kingdom.
6. Wick RR. 2018. Porechop. <https://github.com/rrwick/Porechop>.
7. Wick RR. 2018. Filtlong. <https://github.com/rrwick/Filtlong>.
8. 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>.
9. Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J* 17:10. <https://doi.org/10.14806/ej.17.1.200>.
10. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9:357–359. <https://doi.org/10.1038/nmeth.1923>.
11. 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>.
12. Chaumeil PA, Mussig AJ, Hugenholtz P, Parks DH. 2019. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* 36:1925–1927. <https://doi.org/10.1093/bioinformatics/btz848>.
13. 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>.
14. Manni M, Berkeley MR, Seppely M, Simão FA, Zdobnov EM. 2021. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol Biol Evol* 38:4647–4654. <https://doi.org/10.1093/molbev/msab199>.
15. 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>.