



Draft Genome Sequence of *Streptomyces albidoflavus* 09MW18-IS, Cultivated from the Atlantic Ocean off the Coast of Virginia

 Indu Sharma,^a Masuud Washington,^b  Jeremy Chen See,^c Rola Suleiman,^d  Regina Lamendella^c

^aDepartment of Biological Sciences, Hampton University, Hampton, Virginia, USA

^bThermo Fisher Scientific, Albany, New York, USA

^cJuniata College, Huntingdon, Pennsylvania, USA

^dUniversity of Virginia, Charlottesville, Virginia, USA

ABSTRACT We report a draft genome sequence for *Streptomyces albidoflavus* strain 09MW18-IS, isolated from the Atlantic slope off the coast of Virginia. The whole-genome sequence will provide novel insights into biosynthetic gene clusters and ecological adaptation in an oligotrophic environment.

Streptomyces species from marine environments are adapted to grow in the presence of seawater and can be found in sediments or in association with aquatic plants and invertebrates (1). The genus *Streptomyces* alone accounts for 80% of the natural products within the order *Actinomycetes* (2). Here, we sequenced the genome of *Streptomyces albidoflavus* 09MW18-IS, cultured from near-benthic seawater collected at a depth of 258 m (global positioning system [GPS] coordinates, 37°05'66.94"N, 73°51'13.06"W). The sampling site is located off the coast of Virginia along the Atlantic slope near Washington Canyon. The seawater sample was plated onto seawater complete medium as described previously (3). On day 10, a small pale-yellow colony with a powdery appearance was selected, and an axenic culture was obtained by streaking it onto solid artificial seawater medium.

Amplification of the 16S rRNA gene using the actinomycetes primers For243 (GGA TGA GCC CGC GGC CTA) and R1378 (CGG TGT GTA CAA GGC CCG GGA ACG) and sequencing confirmed that the isolate is a member of the genus *Streptomyces* (4–6).

Whole-genome sequencing was performed as described previously (3). Briefly, genomic DNA was extracted using the UltraClean microbial DNA isolation kit (Qiagen, Inc.) and sequenced using Illumina and Oxford Nanopore technology (ONT). For Illumina sequencing, the library was prepared using the Nextera XT library prep kit (Illumina, San Diego, CA), followed by sequencing using 150-bp paired-end chemistry on a NextSeq 550 platform. A total of 13,649,123 read pairs were generated. For ONT sequencing, the DNA was concentrated using a Vacufuge instrument (Thermo Fisher Scientific, Waltham, MA). The genomic DNA was neither sheared nor size selected. The library was prepared using the Nanopore rapid sequencing (SQK-RAD004) kit (Oxford Nanopore Technologies, Inc., Oxford, UK), with the protocol supplied by the manufacturer. Briefly, the DNA quality was checked using an Invitrogen Qubit 4 fluorometer and the 1× Qubit double-stranded DNA (dsDNA) high-sensitivity assay kit (Thermo Fisher Scientific). Adaptors were ligated and loaded onto the GridION R9.4.1 flow cell via the SpotON sample port. A total of 559,363 reads were generated with an N_{50} value of 7,797 bp. Poretools was used for base calling and demultiplexing (7); the adaptors were trimmed and the quality of the raw reads was checked using FastQC (8). However, error correction was not performed.

Trimmomatic v0.39 (9) was used for quality filtering and trimming the raw Illumina reads using the following parameters: LEADING:3 TRAILING:3 SLIDINGWINDOW:4:20 MINLEN:80. A total of 6,786,495 read pairs were retained, as well as a combined 4,881,460

Editor J. Cameron Thrash, University of Southern California

Copyright © 2022 Sharma 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 Indu Sharma, indu.sharma@hamptonu.edu.

The authors declare no conflict of interest.

Received 2 September 2021

Accepted 2 December 2021

Published 6 January 2022

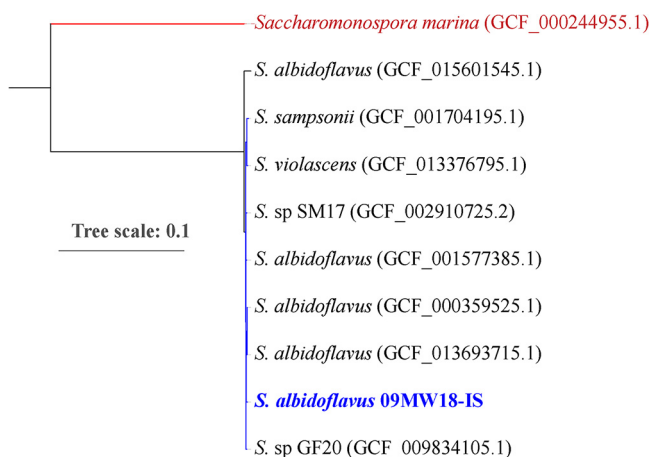


FIG 1 Estimated maximum-likelihood phylogenomic tree based on the concatenated amino acid alignment of 130 single-copy core genes specific for actinomycetes, with *Saccharomonospora* sp. as an outlier.

individual forward and reverse reads. For all the following programs, default settings were used. Hybrid assembly was performed using SPAdes v3.13.0 (10) with the filtered short-read Illumina sequences and raw long-read Oxford Nanopore sequences, and the assembly summary statistics were generated using QUAST v5.0.2 (11). The total length of the draft genome is 7,118,966 bp; it is composed of 81 contigs (N_{50} 153,728 bp; L_{50} 13), with an average GC content of 73.18%. Recruiting the quality-filtered and trimmed reads back to the assembly using Bowtie2 v2.3.4 (11) yielded a mean genome coverage of \sim 447-fold. The publicly available genome was annotated by Prokaryotic Genome Annotation Pipeline (PGAP) (12). The assembly contains 6,209 gene sequence predictions, 5,844 protein-coding sequences, 6 complete rRNA genes, 69 tRNAs, and 275 pseudogenes.

For taxonomic placement, GToTree v1.4.5 was used to create a phylogenomic tree of the nine closely related *Streptomyces* assemblies (Fig. 1) (13–18). Phylogenomics places *Streptomyces* sp. strain 09MW18 close to *S. albidoflavus* (GenBank accession number [GCF_0013693715.1](https://doi.org/10.1128/MRA.01089-19)) and within the *S. albidoflavus* group (Fig. 1).

Data availability. This whole-genome shotgun genome assembly has been deposited at NCBI GenBank under accession number [JAFRUC000000000.1](https://doi.org/10.1128/MRA.01089-19). The raw sequencing reads have been deposited in the NCBI Sequence Read Archive under accession numbers [SRX11958551](https://doi.org/10.1128/MRA.01089-19) and [SRX11958552](https://doi.org/10.1128/MRA.01089-19) and BioProject accession number [PRJNA562082](https://doi.org/10.1128/MRA.01089-19).

ACKNOWLEDGMENTS

This work was supported by National Science Foundation award 1601057.

We thank Deborah Bronk, president and CEO, Bigelow Laboratory for Ocean Sciences, Boothbay Harbor, ME, for participation in a research cruise to collect benthic water samples; John McNabb, research assistant professor, Department of Atmospheric and Planetary Sciences, Hampton University, Hampton, VA, for computational help in establishing a genome assembly pipeline at the Hampton University pirate cluster; and Wright Labs, Huntingdon, PA, for sequencing.

REFERENCES

- Jensen PR, Fenical W. 1994. Strategies for the discovery of secondary metabolites from marine bacteria: ecological perspectives. *Annu Rev Microbiol* 48: 559–584. <https://doi.org/10.1146/annurev.mi.48.100194.003015>.
- Jensen PR, Mincer TJ, Williams PG, Fenical W. 2005. Marine actinomycete diversity and natural product discovery. *Antonie Van Leeuwenhoek* 87: 43–48. <https://doi.org/10.1007/s10482-004-6540-1>.
- Sharma I, Lee MD. 2019. Draft genome sequence of *Cyclobacterium marinum* strain Atlantic-IS, isolated from the Atlantic slope off the coast of Virginia, USA. *Microbiol Resour Announc* 8:e01089-19. <https://doi.org/10.1128/MRA.01089-19>.
- Sanger F, Nicklen S, Coulson AR. 1977. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* 74:5463–5467. <https://doi.org/10.1073/pnas.74.12.5463>.
- Heuer H, Krsek M, Baker P, Smalla K, Wellington E. 1997. Analysis of actinomycete communities by specific amplification of genes encoding 16S rRNA and gel-electrophoretic separation in denaturing gradients. *Appl*

- Environ Microbiol 63:3233–3241. <https://doi.org/10.1128/aem.63.8.3233-3241.1997>.
6. Winker S, Woese CR. 1991. A definition of the domains Archaea, Bacteria and Eucarya in terms of small subunit ribosomal RNA characteristics. *Syst Appl Microbiol* 14:305–310. [https://doi.org/10.1016/S0723-2020\(11\)80303-6](https://doi.org/10.1016/S0723-2020(11)80303-6).
 7. Loman NJ, Quinlan AR. 2014. Poretools: a toolkit for analyzing Nanopore sequence data. *Bioinformatics* 30:3399–3401. <https://doi.org/10.1093/bioinformatics/btu555>.
 8. Andrews S. 2017. FastQC: a quality control tool for high throughput sequence data.
 9. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>.
 10. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Pribelski 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>.
 11. Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 29:1072–1075. <https://doi.org/10.1093/bioinformatics/btt086>.
 12. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Ciufu S, Li W. 2013. Prokaryotic Genome Annotation Pipeline. In *The NCBI handbook* [internet], 2nd ed. US National Center for Biotechnology Information, Bethesda, MD.
 13. Lee MD. 2019. GToTree: a user-friendly workflow for phylogenomics. *Bioinformatics* 35:4162–4164. <https://doi.org/10.1093/bioinformatics/btz188>.
 14. Edgar RC. 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5:113. <https://doi.org/10.1186/1471-2105-5-113>.
 15. Eddy SR. 2011. Accelerated profile HMM searches. *PLoS Comput Biol* 7:e1002195. <https://doi.org/10.1371/journal.pcbi.1002195>.
 16. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25:1972–1973. <https://doi.org/10.1093/bioinformatics/btp348>.
 17. Price MN, Dehal PS, Arkin AP. 2010. FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS One* 5:e9490. <https://doi.org/10.1371/journal.pone.0009490>.
 18. Cock PJA, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, Friedberg I, Hamelryck T, Kauff F, Wilczynski B, de Hoon MJL. 2009. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* 25:1422–1423. <https://doi.org/10.1093/bioinformatics/btp163>.