AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

GENOME SEQUENCES

Draft Genome Sequence of *Streptomyces* sp. Isolate H28 from the Meycauayan River, Philippines

Aroin B. Verdera,[a] Andrew D. Montecillo,[b] Marie Christine M. Obusan[a,c]

[a]Microbial Ecology of Terrestrial and Aquatic Systems Laboratory, Institute of Biology, University of the Philippines Diliman, Quezon City, Philippines
[b]Institute of Biological Sciences, University of the Philippines Los Baños, Laguna, Philippines
[c]Natural Sciences Research Institute, University of the Philippines Diliman, Quezon City, Philippines

ABSTRACT In this paper, we report the draft genome sequence of *Streptomyces* sp. isolate H28, isolated from sediments of the Meycauayan River in the Philippines. This species exhibits production of melanin as well as the ability to utilize and degrade both high-density polyethylene (HDPE) and low-density polyethylene (LDPE).

A *Streptomyces* species called isolate H28 was isolated from sediments of the Meycauayan River, a segment of the Marilao-Meycauayan-Obando River System (MMORS) in the Philippines (14°44′14.29″N, 120°57′26.11″E) that receives effluents from gold smelters, a public market, and households.

Pure culture of isolate H28 was grown and maintained in soyabean-casein digest agar (15 g tryptone, 5 g soya peptone, 5 g NaCl, and 15 g agar per liter). The genomic DNA of isolate H28 was extracted using a KingFisher cell and tissue DNA kit (Thermo Fisher Scientific, USA) according to the manufacturer's protocol. A Nextera XT library prep kit was used for the preparation of the sequencing library, and Illumina MiSeq 2 × 300-bp reads were used for sequencing. A total of 1,390,602 paired-end reads comprising 379,629,799 bases were obtained. The Q-score heat map shows that the first ~300 bp of read 1 sequences were above Q30, and the first ~220 bp of read 2 sequences were above Q30.

Quality control of raw reads was done using the program fastp v0.19.8 (1) using the following parameters: -3, -q 20, and -T 114. After filtering, there were a total of 999,998 reads and 180.917 million bases. Of these, 167.337 million are Q20 bases (92.49%) and 143.117 million are Q30 bases (79.11%). Genomic assembly was done using SPAdes v3.13 (2) using the default parameters and the --careful option. The assessment of the quality of the assembly was done using QUAST (3). The assembly, with an estimated genome coverage of 26×, comprises 1,181 contigs (1,180 scaffolds) with a total of 6,697,207 bases. The largest contig and scaffold is 58,123 bp long. It has a GC content of 72.73%, and the N_{50} and L_{50} lengths are 9,931 bp and 207 bp, respectively.

Using the Microbial Genomes Atlas (MiGA) v0.7.15.2 (4), the closest taxonomic relative was found to be *Streptomyces viridosporus* T7A NZ (GenBank accession number CP023700, GenBank assembly accession number GCA_008704515.1) (*P* value, 0.419) with 87.87% average nucleotide identity (ANI) and a 69.37% fraction of proteins shared. Isolate H28 likely belongs to a species not represented in the database (*P* value, 0.0043). Using MiGA, the assembly was found to have 105 out of the 106 (99.1%) essential genes found in *Bacteria* and *Archaea* in addition to 6,558 predicted proteins with an average length of 296.44 amino acids.

A total of 7,103 new features were predicted via the Rapid Annotations using Subsystems Technology (RAST) tool kit (5). Of these, 319 are noncoding and 6,784 are coding genes. Overall, the coding genes were identified to have 3,352 distinct functions. Genes coding for proteins that are associated with plastic degradation, such as

Citation Verdera AB, Montecillo AD, Obusan MCM. 2021. Draft genome sequence of *Streptomyces* sp. isolate H28 from the Meycauayan River, Philippines. Microbiol Resour Announc 10:e01269-20. https://doi.org/10.1128/MRA.01269-20.

Editor Frank J. Stewart, Georgia Institute of Technology

Address correspondence to Aroin B. Verdera, abverdera@up.edu.ph.

Received 2 November 2020
Accepted 3 December 2020
Published 7 January 2021

esterase, lipase, P450 hydroxylase, multicopper oxidase, and hydroxylase were found. Genes coding for putative tyrosinase and tyrosinase cofactor that are associated with melanin production were also found.

To identify secondary metabolite gene clusters, antiSMASH v5.1.2 (6) was used. The genome was found to have 30 biosynthetic gene clusters coding for a total of 13 secondary metabolite types, including melanin. The following 30 secondary metabolite regions were identified: lanthipeptide (1 region), terpene (8 regions), siderophore (3 regions), melanin (3 regions), nonribosomal peptides (NRPS) (6 regions), ectoine (1 region), butyrolactone (1 region), bacteriocin (2 regions), type II polyketide synthase (T2PKS) (1 region), furan (1 region), ladderane (1 region), indole (1 region), and type I polyketide synthase (T1PKS) (1 region).

All software programs were run using the default parameters unless mentioned otherwise.

Data availability. This whole-genome sequencing project has been deposited at GenBank under the accession numbers JACYXY010000001 to JACYXY010001180, BioProject number PRJNA665970, BioSample number SAMN16275722, and Sequencing Read Archive (SRA) number SRR13061778.

ACKNOWLEDGMENTS

This project was funded by the Department of Science and Technology Philippine Council for Industry, Energy, and Emerging Technology Research and Development (DOST-PCIEERD) and the University of the Philippines Diliman (UPD) Natural Sciences Research Institute (project code PCIEERD-BIO-17-03).

The sequencing services were provided by the Philippine Genome Center (PGC), UPD, under the Genomics Junior Camp (project number 18-N08).

We declare no conflicts of interest.

REFERENCES

1. Chen S, Zhou Y, Chen Y, Gu J. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 34:i884–i890. https://doi.org/10.1093/bioinformatics/bty560.
2. 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.
3. 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.
4. Rodriguez-R LM, Gunturu S, Harvey WT, Rosselló-Mora R, Tiedje JM, Cole JR, Konstantinidis KT. 2018. The Microbial Genomes Atlas (MiGA) webserver: taxonomic and gene diversity analysis of *Archaea* and *Bacteria* at the whole genome level. Nucleic Acids Res 46:W282–W288. https://doi.org/10.1093/nar/gky467.
5. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O. 2008. The RAST server: Rapid Annotations using Subsystems Technology. BMC Genomics 9:75. https://doi.org/10.1186/1471-2164-9-75.
6. 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.