



# Draft Genome Sequences of *Penicillium* spp. from Deeply Buried Oligotrophic Marine Sediments

Morgan S. Sobol,<sup>a</sup> Tatsuhiko Hoshino,<sup>b</sup> Taiki Futagami,<sup>c</sup> Fumio Inagaki,<sup>b</sup> Brandi Kiel Reese<sup>a</sup>

<sup>a</sup>Department of Life Sciences, Texas A&M University—Corpus Christi, Corpus Christi, Texas, USA

<sup>b</sup>Kochi Institute for Core Sample Research, Japan Agency for Marine-Earth Science and Technology (JAMSTEC), Nankoku, Japan

<sup>c</sup>Education and Research Centre for Fermentation Studies, Faculty of Agriculture, Kagoshima University, Kagoshima, Japan

**ABSTRACT** Here, we report genome sequences of two *Penicillium* isolates from below the seafloor of the oligotrophic South Pacific Gyre. These genomes are the first reported for fungi from deeply buried marine sediment. Both genomes will provide valuable information regarding the role of fungi and carbon cycling in the energy-limited subsurface biosphere.

Recent advances in molecular and culturing techniques have verified the presence and activity of eukaryotic fungi in deeply buried marine sediment. However, the physiological and metabolic capabilities of fungi in the deep marine subsurface are still poorly understood (1, 2). Whole-genome sequencing is invaluable to delineate the role that marine subsurface fungi play in global biogeochemical cycles and their survivability in an energy-limited system. Here, we present draft genome sequences of two *Penicillium* isolates from the oligotrophic sediments of the South Pacific Gyre. Isolates of *Penicillium* sp. SPG-F1 and SPG-F15 were collected during the Integrated Ocean Drilling Program Expedition 329 at 124 m below the seafloor (mbsf) and 12 mbsf, respectively. Sediment was enriched using 10% marine broth media (Difco 2216; BD Diagnostics), and the fungi were subsequently isolated and maintained in potato dextrose broth (Difco 254920; BD Diagnostics) at ambient temperature prior to nucleic acid extraction.

High-quality genomic DNA was extracted by following a cetyltrimethylammonium bromide (CTAB)-based method (3) with modifications. An Illumina paired-end library was prepared following the manufacturer's protocol (TruSeq DNA PCR-Free Sample Preparation Guide, part number 15036187 Rev. A; [http://research.lunefeld.ca/ngs/truseq\\_dna\\_pcrfree\\_sampleprep\\_guide\\_15036187\\_a.pdf](http://research.lunefeld.ca/ngs/truseq_dna_pcrfree_sampleprep_guide_15036187_a.pdf)) to generate 101-bp length paired-end reads. The isolates were sequenced using a HiSeq 2000 sequencer (Illumina, San Diego, CA) at Macrogen in Seoul, South Korea, which produced 162,517,156 reads for *Penicillium* sp. SPG-F15 and 149,610,352 reads for *Penicillium* sp. SPG-F1.

Low-quality reads (quality score [Q] of <20) were removed, and Illumina adapters were trimmed from the reads using Trim Galore! The average read length after trimming was 101 bp for both isolates. The genomes were assembled *de novo* using SPAdes v3.11.0 (4). In SPAdes, the coverage cutoff was set to 10× and the careful mode turned on to reduce the number of mismatches and errors. The genome of *Penicillium* sp. SPG-F15 was assembled into 348 contigs totaling 36.13 Mbp in length with a 46.52% G+C content. The assembly coverage was 436×. Isolate *Penicillium* sp. SPG-F1 was assembled into 457 contigs totaling 32.06 Mbp with a 48.26% G+C content. The assembly coverage was 446×. The lengths and G+C contents of both genomes were comparable to those of other *Penicillium* species of continental origin (5). The genomes

**Citation** Sobol MS, Hoshino T, Futagami T, Inagaki F, Kiel Reese B. 2019. Draft genome sequences of *Penicillium* spp. from deeply buried oligotrophic marine sediments. *Microbiol Resour Announc* 8:e01613-18. <https://doi.org/10.1128/MRA.01613-18>.

**Editor** Frank J. Stewart, Georgia Institute of Technology

**Copyright** © 2019 Sobol 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 Brandi Kiel Reese, [brandi.reese@tamucc.edu](mailto:brandi.reese@tamucc.edu).

This is a contribution to the Deep Carbon Observatory. This is Center for Dark Energy Biosphere Investigations (CDEBI) contribution number 458.

**Received** 27 November 2018

**Accepted** 9 January 2019

**Published** 7 February 2019

contained approximately 98% complete orthologs according to Benchmarking Universal Single-Copy Orthologs (BUSCO) software (6).

Repetitive elements were identified and masked with RepeatModeler v1.0.11 (7) and RepeatMasker v4.0.7 (8). A total of 3,550,914 bp (9.82%) of repetitive elements were masked in *Penicillium* sp. SPG-F15, whereas *Penicillium* sp. SPG-F1 had 928,433 bp (2.90%) masked. For annotation, a MAKER2 pipeline was used (9, 10) that incorporated *ab initio* gene predictions from both Augustus (11) and GeneMark-ES (12). A total of 11,243 protein-coding sequences were found in *Penicillium* sp. SPG-F15 and 11,581 were found in *Penicillium* sp. SPG-F1, which is consistent with those in other *Penicillium* species (5). A phylogenomics approach was used to identify both organisms as *Penicillium*. Single-copy homologues from both genomes in this study were clustered and compared with *Penicillium* and *Aspergillus* genomes (downloaded from NCBI GenBank and the Joint Genome Institute) using GET\_homologues (13). The ortholog search was carried out with OrthoMCL using a default cutoff of  $1e-05$ .

The predicted protein sequences were given to Blast2GO version 5.1 for putative gene ontology assignment, classifying enzyme codes, and Kyoto Encyclopedia of Genes and Genomes mapping (14). The proteins were also annotated with the Carbohydrate-Active enZymes (CAZy) database (15). Genes involved in hydrocarbon, lignin, lignocellulose, and carbonate degradation were present in both genomes, but a more detailed analysis of these metabolic pathways will be assessed in a future publication. Both genomes will serve as important resources for understanding the influence that fungi have on recalcitrant carbon degradation in the marine subsurface.

**Data availability.** The genomic sequences and assemblies have been deposited at GenBank under BioProject accession numbers [PRJNA435885](https://www.ncbi.nlm.nih.gov/submit/bioproject/PRJNA435885) (*Penicillium* sp. SPG-F15) and [PRJNA435890](https://www.ncbi.nlm.nih.gov/submit/bioproject/PRJNA435890) (*Penicillium* sp. SPG-F1). This announcement represents the first versions of both genomes.

## ACKNOWLEDGMENTS

We thank the shipboard scientific party, ship and drilling crews, and technical staffs for supporting seafloor core sampling during the Integrated Ocean Drilling Program (IODP) Expedition 329. We also thank Martha Ariza and Erick Carvalho for their assistance.

Sequencing of the genomes was supported by Japan Society for the Promotion of Science (JSPS) Grants-in-Aid for Science Research (number 26251041 to F.I. and number 17H03956 to T.H.). Funding for ecophysiological characterization was provided by an Integrated Ocean Drilling Program postexpedition award to B.K.R.

## REFERENCES

- Edgcomb VP, Beaudoin D, Gast R, Biddle JF, Teske A. 2011. Marine subsurface eukaryotes: the fungal majority. *Environ Microbiol* 13: 172–183. <https://doi.org/10.1111/j.1462-2920.2010.02318.x>.
- Orsi W, Biddle JF, Edgcomb V. 2013. Deep sequencing of seafloor eukaryotic rRNA reveals active fungi across marine subsurface provinces. *PLoS One* 8:e56335. <https://doi.org/10.1371/journal.pone.0056335>.
- Saghai-Marouf MA, Soliman KM, Jorgensen RA, Allard R. 1984. Ribosomal DNA spacer-length polymorphisms in barley: Mendelian inheritance, chromosomal location, and population dynamics. *Proc Natl Acad Sci USA* 81:8014–8018. <https://doi.org/10.1073/pnas.81.24.8014>.
- 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>.
- Nielsen JC, Grijseels S, Prigent S, Ji B, Dainat J, Nielsen KF, Frisvad JC, Workman M, Nielsen J. 2017. Global analysis of biosynthetic gene clusters reveals vast potential of secondary metabolite production in *Penicillium* species. *Nat Microbiol* 2:17044. <https://doi.org/10.1038/nmicrobiol.2017.44>.
- Simao FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31:3210–3212. <https://doi.org/10.1093/bioinformatics/btv351>.
- Smit A, Hubley R. 2015. RepeatModeler Open-1.0, 2008–2015. <http://www.repeatmasker.org>.
- Smit A, Hubley R, Green P. 2015. RepeatMasker Open-4.0, 2013–2015. <http://repeatmasker.org>.
- Holt C, Yandell M. 2011. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics* 12:491. <https://doi.org/10.1186/1471-2105-12-491>.
- Koutsovoulos G, Makepeace B, Tanya VN, Blaxter M. 2014. Palaeosymbiosis revealed by genomic fossils of *Wolbachia* in a strongyloidean nematode. *PLoS Genet* 10:e1004397. <https://doi.org/10.1371/journal.pgen.1004397>.
- Stanke M, Waack S. 2003. Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics* 19:ii215–ii225. <https://doi.org/10.1093/bioinformatics/btg1080>.
- Ter-Hovhannisyan V, Lomsadze A, Chernoff YO, Borodovsky M. 2008. Gene prediction in novel fungal genomes using an *ab initio* algorithm

- with unsupervised training. *Genome Res* 18:1979–1990. <https://doi.org/10.1101/gr.081612.108>.
13. Contreras-Moreira B, Vinuesa P. 2013. GET\_HOMOLOGUES, a versatile software package for scalable and robust microbial pan-genome analysis. *Appl Environ Microbiol* 79:7696–7701. <https://doi.org/10.1128/AEM.02411-13>.
  14. Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21:3674–3676. <https://doi.org/10.1093/bioinformatics/bti610>.
  15. Yin Y, Mao X, Yang J, Chen X, Mao F, Xu Y. 2012. dbCAN: a Web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res* 40:W445–W451. <https://doi.org/10.1093/nar/gks479>.