AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

GENOME SEQUENCES

Draft Genome Sequence of *Purpureocillium takamizusanense*, a Potential Bioinsecticide

Ngoc-Hung Nguyen,[a] Tomokazu Tamura,[a] Kiminori Shimizu[a,b]

[a]Department of Biological Science and Technology, Tokyo University of Science, Katsushika, Tokyo, Japan
[b]Medical Mycology Research Center, Chiba University, Chuo-ku, Chiba, Japan

ABSTRACT To investigate the biocontrol capability of the entomopathogenic fungus *Purpureocillium takamizusanense*, the genome of the wild-type strain isolated from synnemata on *Meimuna opalifera*, was sequenced using a combination of HiSeq and Nanopore technologies, and annotated using evidence from RNA sequences and protein sequences from its sister species *Purpureocillium lilacinum*.

Purpureocillium takamizusanense (phylum Ascomycota, class Sordariomycetes, order Hypocreales, family Ophiocordycipitaceae) is a species related to the better-known biocontrol agent *Purpureocillium lilacinum* (1). To investigate the capability of *P. takamizusanense* as a bioinsecticide, the complete genome of this species was sequenced, assembled, and annotated.

The wild-type strain (PT3) of *P. takamizusanense* was isolated from synnemata on *Meimuna opalifera* collected in Chiba, Chiba Prefecture, Japan (35°34′47.9″N, 140°13′39.0″E), on 7 August 2016. The isolated strain was grown on potato dextrose broth (BD Difco) at 25°C with shaking at 120 rpm for 5 days (for DNA extraction) and 7 days (for RNA extraction). Genomic DNA was extracted using the DNAeasy plant minikit (Qiagen) following the fungal DNA isolation protocol. For Illumina sequences, a library was constructed using the Nextera DNA library preparation kit and sequenced on a HiSeq X instrument using the paired-end 150-bp protocol. For the Nanopore platform, DNA was prepared using the ligation sequencing kit (SQK-LSK109) and sequenced on the PromethION platform using FLO-PRO002 flow cells. RNA was extracted using the RNA premium kit (FastGene), a sequencing library prepared using the TruSeq stranded mRNA kit, and sequenced using the Illumina NovaSeq instrument. Finally, we obtained 44,316,524 HiSeq reads (~3.51 Gb), 407,870 long reads (~4.09 Gb; mean, 10.92 kb; read N_{50}, 17,299 kb; and longest read, 55,422 kb), and 98,646,618 RNA reads (~14.896 Gb). The Nanopore reads were adapter trimmed using Porechop v0.2.4 (https://github.com/rrwick/Porechop), contamination filtered using NanoLyse v1.1.0 (2), and size filtered (>1,000 bp) using SeqKit v2.2.0 (3).

The draft genome was first assembled using the MaSuRCA pipeline (4) with both the HiSeq and Nanopore sequences. The assembly was polished by mapping the HiSeq sequences to the assembly using BWA-MEM v0.7.17 (5) and then implementing four rounds of polishing using Pilon v1.24 (6). A mitochondrial contig was identified by mapping the polished assembly against the mitochondrial genome of *P. lilacinum* (7) (GenBank accession no. MN635609) using Minimap2 v2.22 (8); it was extracted, and the overlapping ends of the circular sequence were trimmed using seq-circ-trim (9). The gene annotation was conducted using MFannot (10) and the MITOS (11) Web server (based on Genetic Code 4) and checked visually and combined manually using the Integrative Genomics Viewer (IGV) (12). Repeated sequences of the assembled genome were predicted using RepeatModeler v2.0.1 (13) and marked using RepeatMasker v4.1.1 (14). The RNA sequences were first adapter cut using Cutadapt v1.15 (15) and then mapped and indexed to repeat the masked assembly using HISAT2 (16) and SAMtools v1.15 (17). Gene prediction and annotation were performed using a protein

Editor Antonis Rokas, Vanderbilt University

Address correspondence to Kiminori Shimizu, shimizuk@rs.tus.ac.jp.

The authors declare no conflict of interest.

Received 29 March 2022
Accepted 14 June 2022
Published 28 June 2022

TABLE 1 Genome features of the *Purpureocillium takamizusanense* assembly and annotation

Characteristic	Value
Chromosome scaffolds	
Total assembly length (Mb)	35.6
No. of scaffolds	14
N_{50} (Mb)	3.09
L_{50}	4
Max length (Mb)	6.1
No. of complete BUSCOs[a] (%)	3,791 (99.3)
No. of incomplete BUSCOs (%)	12 (0.3)
No. of missing BUSCOs (%)	14 (0.4)
% repeats	4.6
%GC	57.76
No. of protein-coding genes	11,855
No. of tRNAs	85
No. of Phobius secretome genes	1,089
No. of Phobius transmembrane proteins	2,680
No. of antiSMASH biosynthetic gene clusters	36
Mitochondrial genome	
Length (bp)	33,113
No. of protein-coding genes	24
No. of rRNAs	2
No. of tRNAs	35

[a] BUSCOs, benchmarking universal single-copy orthologs.

homology search with the most closely related species, *P. lilacinum* (GCA_001653265), and RNA-seq using the Funannotate v1.8.1 pipeline (18), respectively. The completeness of the annotation was accessed using BUSCO v4.1.3 with 3,817 ortholog genes from the class Sordariomycetes (19). For the chromosome scaffolds, a total of 11,855 protein-coding genes were predicted (18) (Table 1). The complete mitochondrial genome of *P. takamizusanense* is a closed circular molecule of 33,113 bp thats contains 61 genes (Table 1). For the analyses in this study, default parameters were used except where otherwise noted.

Data availability. The *Purpureocillium takamizusanense* genome project, including the raw Illumina, Nanopore, and RNA data, has been deposited at DDBJ/EMBL/GenBank under BioProject accession number PRJNA685267. The assembled nuclear genome and assembled mitochondrial sequences are available under GenBank accession numbers GCA_022605165.1 and OK505612, respectively. The raw HiSeq, Nanopore, and RNA sequencing reads have been deposited at the Sequence Read Archive under accession numbers SRR16282004, SRR16282003, and SRR16888993, respectively.

ACKNOWLEDGMENTS

We are grateful to Lin Si-Min for lending us the computer server to conduct the bioinformatics analyses. This research was partly supported by funding from the Hokuto Foundation for Bioscience to K.S.

REFERENCES

1. Ban S, Azuma Y, Sato H, Suzuki KI, Nakagiri A. 2015. *Isaria takamizusanensis* is the anamorph of *Cordyceps ryogamimontana*, warranting a new combination, *Purpureocillium takamizusanense* comb. nov. Int J Syst Evol Microbiol 65:2459–2465. https://doi.org/10.1099/ijs.0.000284.
2. De Coster W, D'Hert S, Schultz DT, Cruts M, Van Broeckhoven C. 2018. NanoPack: visualizing and processing long-read sequencing data. Bioinformatics 34:2666–2669. https://doi.org/10.1093/bioinformatics/bty149.
3. Shen W, Le S, Li Y, Hu F. 2016. SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation. PLoS One 11:e0163962. https://doi.org/10.1371/journal.pone.0163962.
4. Zimin AV, Marçais G, Puiu D, Roberts M, Salzberg SL, Yorke JA. 2013. The MaSuRCA genome assembler. Bioinformatics 29:2669–2677. https://doi.org/10.1093/bioinformatics/btt476.
5. Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv 1303.3997 [q-bio.GN]. https://arxiv.org/abs/1303.3997.
6. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One 9:e112963. https://doi.org/10.1371/journal.pone.0112963.
7. Li J, Zhang G, Yu H, Huang L, Zeng W, Wang Y. 2019. Complete mitochondrial genome of the important bio-control fungus *Purpureocillium lilacinum* (Ophiocordycipitaceae, Hypocreales) and its phylogenetic analysis. Mitochondrial DNA B Resour 5:240–242. https://doi.org/10.1080/23802359.2019.1699466.
8. Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34:3094–3100. https://doi.org/10.1093/bioinformatics/bty191.
9. Hackl T. 2019. thackl/cr-genomes: cr-genomes-v1.9. Zenodo.

10. Valach M, Burger G, Gray MW, Lang BF. 2014. Widespread occurrence of organelle genome-encoded 5S rRNAs including permuted molecules. Nucleic Acids Res 42:13764–13777. https://doi.org/10.1093/nar/gku1266.
11. Bernt M, Donath A, Jühling F, Externbrink F, Florentz C, Fritzsch G, Pütz J, Middendorf M, Stadler PF. 2013. MITOS: improved de novo metazoan mitochondrial genome annotation. Mol Phylogenet Evol 69:313–319. https://doi.org/10.1016/j.ympev.2012.08.023.
12. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. 2011. Integrative genomics viewer. Nat Biotechnol 29:24–26. https://doi.org/10.1038/nbt.1754.
13. Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, Smit AF. 2020. RepeatModeler2 for automated genomic discovery of transposable element families. Proc Natl Acad Sci U S A 117:9451–9457. https://doi.org/10.1073/pnas.1921046117.
14. Smit A, Hubley R, Green P. 2020. RepeatMasker Open-4.0. http://www.repeatmasker.org/.
15. 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.
16. Kim D, Langmead B, Salzberg SL. 2015. HISAT: a fast spliced aligner with low memory requirements. Nat Methods 12:357–360. https://doi.org/10.1038/nmeth.3317.
17. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. Bioinformatics 25:2078–2079. The Sequence Alignment/Map format and SAMtools. https://doi.org/10.1093/bioinformatics/btp352.
18. Palmer J. 2020. Funannotate: fungal genome annotation scripts. https://github.com/nextgenusfs/funannotate.
19. Simão 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.