

DATA NOTE

Open Access



Draft genome sequence and annotation of fungal pathogen *Cladosporium tenuissimum* SHK1100 causing postharvest sooty spot of fig fruit

Masahito Nakano^{1*}

Abstract

Objectives *Cladosporium tenuissimum*, a widespread filamentous fungus belonging to the *C. cladosporioides* species complex, has been identified as a causal agent of sooty spot disease on postharvest fig fruits. Despite its ecological and agricultural importance, genomic resources for this species remain limited. Here, I report a draft genome sequence of *C. tenuissimum* strain SHK1100, isolated from symptomatic fig fruits in Japan, to facilitate comparative and functional genomic studies.

Data description The draft genome was assembled using long-read sequencing data, resulting in a total assembly length of 32.8 Mb, composed of 58 contigs with an N50 of 1.92 Mb and L50 of 8. The genome completeness of the assembly reached 98.0% in Benchmarking Universal Single-Copy Orthologs analysis with no contamination. In total, 12,016 protein-coding genes were predicted in the assembled genome. Functional annotation assigned putative functions to a large proportion of predicted genes, including annotations for 77.1% of genes into Clusters of Orthologous Groups, 33.7% of genes using Gene Ontology, and 21.6% of genes as putative effectors. This draft genome provides a foundation for understanding the pathogenic mechanisms and secondary metabolite biosynthesis of *Cladosporium* species associated with plant diseases.

Keywords *Cladosporium tenuissimum*, *Ficus carica*, Fig fruit, Genome assembly, PacBio sequencing, Postharvest, Sooty spot

*Correspondence:

Masahito Nakano

masahito_nakano@shokuken.or.jp

¹Toyo Institute of Food Technology, 4-23-2 Minami-Hanayashiki, Kawanishi-shi, Hyogo 666-0026, Japan



© The Author(s) 2026. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Objective

Cladosporium, a monophyletic genus within the *Cladosporiaceae* family, is recognized as one of the most abundant fungi in indoor and outdoor environments [1]. The genus is divided into three major species complexes—*C. herbarum*, *C. sphaerospermum*, and *C. cladosporioides*—based on their morphological features and multilocus DNA sequences. Members of this genus include species of ecological and industrial importance, as well as potential allergenic species and plant pathogens. *Cladosporium* can produce a wide range of secondary metabolites that exhibit biological activities, such as cytotoxicity [2], antimicrobial activity [3], and enzyme inhibition [4].

C. tenuissimum, a member of the *C. cladosporioides* species complex, is a widely distributed filamentous fungus frequently isolated from indoor environments, soil, and plant surfaces [1]. *C. tenuissimum* has been reported to cause disease on economically important crops, such as peach [5] and citrus [6]. In a previous study, *C. tenuissimum* was identified as the causal agent of sooty spot disease on postharvest fig (*Ficus carica*) fruits [7]. The fungus was isolated from symptomatic fruits collected from commercial markets and an experimental orchard of the Toyo Institute of Food Technology (34.81926°N, 135.40315°E) in Japan. Among the isolates, strain SHK1100 (MAFF248147) was characterized based on morphological features and multilocus sequence analysis using ITS, TEF1, and ACT [7].

Recent genome and transcriptome analyses of *Cladosporium* species have demonstrated their practical utility by revealing gene clusters and enzyme repertoires associated with plant growth promotion and organic pollutant degradation [8–10]. Despite its widespread occurrence, genomic information for *C. tenuissimum* remains limited, hindering comparative and functional genomic studies within this genus. Here, I report the draft genome sequence of *C. tenuissimum* SHK1100 generated using long-read sequencing technology. This genetic resource provides a foundation for further studies

on the pathogenic mechanisms and secondary metabolite biosynthesis of *Cladosporium* species associated with postharvest fruit disease.

Data description

For genome sequencing, the *C. tenuissimum* SHK1100 was grown in 100 mL of potato dextrose broth containing 100 µg/mL chloramphenicol for three days at 28 °C with shaking at 150 rpm. Mycelia (100 mg) were harvested by gravity filtration using filter paper (Advantec, Tokyo, Japan) and frozen in liquid nitrogen [11]. Genomic DNA was extracted using the NucleoBond HMW DNA kit (Takara Bio, Shiga, Japan) and purified using DNA Clean Beads (MGI Tech, Shenzhen, China). The integrity of DNA was assessed using a Genomic DNA Screen-Tape (Agilent Technologies, Santa Clara, CA, USA) on an automated electrophoresis system TapeStation 4150 (Agilent Technologies), and the concentration of DNA was measured using a Qubit Fluorometer (Thermo Fisher Scientific, Waltham, MA, USA).

A PCR-free DNA library was prepared using the SMRTbell Prep Kit 3.0 (Pacific Biosciences, Menlo Park, CA, USA) after fragmenting the genomic DNA to 10–25 kb using a Megaruptor 3 (Diagenode, Liège, Belgium). The library was sequenced on a Revio system (Pacific Biosciences) by Bioengineering Lab. Co., Ltd. (Kanagawa, Japan). Raw reads were processed with SMRT Link v13.1.0.221970 to remove overhang adapter sequences and obtain HiFi reads (Q > 20), yielding 198,849 reads with an N50 of 7,430 bp (Table 1, Data set 1) [12]. HiFi reads longer than 5,000 bp were selected using SeqKit v2.10.0 [13] and subsequently assembled *de novo* using Hifiasm v0.21.0-r686 [14], resulting in 58 contigs with an N50 of 1,923,577 bp (Table 1, Data file 1, Data set 2) [15, 16]. Telomeric repeats were detected at both ends of 16 contigs using tidk v0.2.7 [17]. Assembly quality was assessed using QUAST-LG v5.3.0 [18]. Genome completeness was evaluated using Benchmarking Universal Single-Copy Orthologs (BUSCO) v5.7.0

Table 1 Overview of data files/data sets

Label	Name of data file/data set	File types (file extension)	Data repository and identifier (DOI or accession number)
Data file 1	Genome assembly and annotation summary of <i>Cladosporium tenuissimum</i> SHK1100	MS Word file (.docx)	Figshare, https://doi.org/10.6084/m9.figshare.30948959 [16]
Data file 2	Functional annotation of genes in <i>Cladosporium tenuissimum</i> SHK1100	MS Excel file (.xlsx)	Figshare, https://doi.org/10.6084/m9.figshare.30948908 [41]
Data set 1	Sequencing read of <i>Cladosporium tenuissimum</i> SHK1100	Fastq file (.fastq)	NCBI Sequence Read Archive, http://identifiers.org/insdc.sra:DRR700011 [12]
Data set 2	Genome assembly of <i>Cladosporium tenuissimum</i> SHK1100	Fasta file (.fasta)	NCBI Nucleotide, https://identifiers.org/nucleotide:BAJJPQ000000000 [15]
Data set 3	Predicted genes of <i>Cladosporium tenuissimum</i> SHK1100	Fasta file (.fasta)	Figshare, https://doi.org/10.6084/m9.figshare.30949160 [26]
Data set 4	Gene annotation of <i>Cladosporium tenuissimum</i> SHK1100	General Feature Format file (.gff3)	Figshare, https://doi.org/10.6084/m9.figshare.31437025 [31]

with the dothideomycetes_odb10 dataset [19]. Average nucleotide identity values calculated using FastANI v1.34 [20] showed that *C. tenuissimum* SHK1100 shared 96.23% identity with *C. tenuissimum* ACCC 37043 (GCA_046128905.1) and 88.38% identity with *C. cladosporioides* SYC63 (GCA_022457075.1).

For gene structure annotation, protein-coding genes were predicted using BRAKER v3.0.8 [21] based on GeneMark-ES/EP+ [22, 23] and AUGUSTUS v3.5.0 [24], incorporating fungal protein homology from OrthoDB v12 [25]. In total, 12,016 protein-coding genes were identified in the assembled genome (Table 1, Data file 1, Data set 3) [16, 26] and no contaminants were detected by OMArk v0.3.1 [27]. Non-coding genes, such as tRNA and rRNA, were predicted using tRNAscan-SE v2.0.12 [28] and Barrnap v0.9 [29]. Additionally, repeat family identification and modeling were conducted using RepeatModeler v2.0.7 [30] and RepeatMasker v4.2.1 (Table 1, Data set 4) [31].

For gene functional annotation, predicted protein-coding genes were annotated using eggNOG-mapper v2.1.12 [32] against orthology-based database eggNOG DB v5.0.2, including Clusters of Orthologous Groups (COG) [33], Gene Ontology (GO) [34], Kyoto Encyclopedia of Genes and Genomes (KEGG) [35], carbohydrate-active enzymes (CAZy) [36], and Pfam [37], with DIAMOND v2.1.12 [38]. Additionally, secreted proteins and putative effectors were predicted using SignalP v6.0 [39] and EffectorP v3.0 [40], respectively, identifying 1,209 secreted proteins and 2,600 putative effectors. A summary of functional annotation is presented in Data file 2 (Table 1) [41].

Limitations

This study presents a draft genome of *C. tenuissimum* generated using PacBio long-read sequencing. As the genome analysis was conducted on a single isolate, the results could not be generalized to the entire species and should be interpreted as strain-specific. Although the assembly shows high contiguity compared with short-read-based assemblies, it has not yet achieved complete chromosome-level resolution. A complete mitochondrial genome could not be identified in the assembly. Additionally, gene structure and functional annotations were generated using automated bioinformatics pipelines and public databases without experimental evidence; therefore, these findings remain preliminary. Further improvements, including scaffolding using Hi-C data and the integration of transcriptomic or proteomic evidence, would facilitate the generation of a more accurate chromosome-level genome assembly. Comparative genomic analyses with other *Cladosporium* species would further enhance our understanding of gene function and evolutionary relationships.

Abbreviations

BUSCO	Benchmarking Universal Single-Copy Orthologs
CAZy	Carbohydrate-active enzymes
COG	Clusters of Orthologous Groups
GO	Gene Ontology
KEGG	Kyoto Encyclopedia of Genes and Genomes

Acknowledgements

Not applicable.

Author contributions

Conceptualization: M.N.; Investigation: M.N.; Writing - original draft: M.N.; Writing - review & editing: M.N.

Funding

Not applicable.

Data availability

The data described in this Data note can be freely and openly accessed on NCBI Sequence Read Archive under BioProject PRJDB35511. Please see Table 1 and references [12, 15, 16, 26, 31, 41] for details and links to the data.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The author declares no competing interests.

Received: 3 February 2026 / Accepted: 15 April 2026

Published online: 21 April 2026

References

- Bensch K, Groenewald JZ, Meijer M, Dijksterhuis J, Jurjevic Z, Andersen B, Houbakken J, Crous PW, Samson RA. *Cladosporium* species in indoor environments. *Stud Mycol*. 2018;89:177–301.
- Liu Y, Kurtan T, Yun Wang C, Han Lin W, Orfali R, Muller WE, Daletos G, Proksch P. Cladosporinone, a new viriditoxin derivative from the hypersaline lake derived fungus *Cladosporium cladosporioides*. *J Antibiot (Tokyo)*. 2016;69(9):702–6.
- Wang X, Radwan MM, Tarawneh AH, Gao J, Wedge DE, Rosa LH, Cutler HG, Cutler SJ. Antifungal activity against plant pathogens of metabolites from the endophytic fungus *Cladosporium cladosporioides*. *J Agric Food Chem*. 2013;61(19):4551–5.
- Rotinsulu H, Yamazaki H, Sugai S, Iwakura N, Wewengkang DS, Sumilat DA, Namikoshi M. Cladosporamide A, a new protein tyrosine phosphatase 1B inhibitor, produced by an Indonesian marine sponge-derived *Cladosporium* sp. *J Nat Med*. 2018;72(3):779–83.
- Nativitas-Lima I, Leyva-Mir SG, Tovar-Pedraza JM, Leyva-Madrigal KY, Nativitas-Lima R, Cabrera-Hidalgo AJ, Camacho-Tapia M. First report of peach scab caused by *Cladosporium tenuissimum* in Mexico. *Plant Dis*. 2023;107(10).
- Xiao Y, Xie Y, Su Q, Huo G, Cui C. First report of brown spot of dekopon fruit caused by *Cladosporium tenuissimum* in China. *Plant Dis*. 2023;107(2).
- Nakano M. First report of *Cladosporium tenuissimum* causing postharvest sooty spot of fig fruit in Japan. *Plant Dis* 2026;110(2).
- Yang N, Zhang W, Wang D, Cao D, Cao Y, He W, Lin Z, Chen X, Ye G, Chen Z, et al. A novel endophytic fungus strain of *Cladosporium*: its identification, genomic analysis, and effects on plant growth. *Front Microbiol*. 2023;14:1287582.
- Nikolaivits E, Siaperas R, Agraftiotis A, Ouazzani J, Magoulas A, Gioti A, Topakas E. Functional and transcriptomic investigation of laccase activity in the presence of PCB29 identifies two novel enzymes and the multicopper oxidase repertoire of a marine-derived fungus. *Sci Total Environ*. 2021;775:145818.
- Gioti A, Siaperas R, Nikolaivits E, Le Goff G, Ouazzani J, Kotoulas G, Topakas E. Draft genome sequence of a *Cladosporium* species isolated from the

- Mesophotic Ascidian *Didemnum maculosum*. Microbiol Resour Announc. 2020;9(18).
11. Nakano M. Genome sequence of *Gilbertella persicaria* strain SHK1000 isolated from fig fruit. Microbiol Resour Announc. 2026:e0046725.
 12. NCBI Sequence Read. Archive 2026. . 2026. Available from: <https://www.ebi.ac.uk/ena/browser/view/DRR700011>.
 13. Shen W, Sipos B, Zhao L. SeqKit2: A Swiss army knife for sequence and alignment processing. iMeta. 2024;3(3):e191.
 14. Cheng H, Concepcion GT, Feng X, Zhang H, Li H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. Nat Methods. 2021;18(2):170–5.
 15. NCBI Nucleotide. 2026. Available from: <https://www.ncbi.nlm.nih.gov/nuccore/3153112477>.
 16. Nakano M. Genome assembly and annotation summary of *Cladosporium tenuissimum* SHK1100. Figshare. 2026. <https://doi.org/10.6084/m9.figshare.30948959>.
 17. Brown MR, Manuel G, de La Rosa P, Blaxter M. tidk: a toolkit to rapidly identify telomeric repeats from genomic datasets. Bioinformatics. 2025;41(2).
 18. Mikheenko A, Pribelski A, Savelliev V, Antipov D, Gurevich A. Versatile genome assembly evaluation with QUAST-LG. Bioinformatics. 2018;34(13):i142–50.
 19. Manni M, Berkeley MR, Seppey M, Simao FA, Zdobnov EM. BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. Mol Biol Evol. 2021;38(10):4647–54.
 20. Jain C, Rodriguez RL, Phillippy AM, Konstantinidis KT, Aluru S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat Commun. 2018;9(1):5114.
 21. Gabriel L, Bruna T, Hoff KJ, Ebel M, Lomsadze A, Borodovsky M, Stanke M. BRAKER3: Fully automated genome annotation using RNA-seq and protein evidence with GeneMark-ETP, AUGUSTUS, and TSEBRA. Genome Res. 2024;34(5):769–77.
 22. Bruna T, Lomsadze A, Borodovsky M. GeneMark-EP+: eukaryotic gene prediction with self-training in the space of genes and proteins. NAR Genom Bioinform. 2020;2(2):lqaa026.
 23. Ter-Hovhannisyan V, Lomsadze A, Chernoff YO, Borodovsky M. Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training. Genome Res. 2008;18(12):1979–90.
 24. Stanke M, Diekhans M, Baertsch R, Haussler D. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. Bioinformatics. 2008;24(5):637–44.
 25. Tegenfeldt F, Kuznetsov D, Manni M, Berkeley M, Zdobnov EM, Kriventseva EV. OrthoDB and BUSCO update: annotation of orthologs with wider sampling of genomes. Nucleic Acids Res. 2025;53(D1):D516–22.
 26. Nakano M. Predicted genes of *Cladosporium tenuissimum* SHK1100. Figshare. 2026. <https://doi.org/10.6084/m9.figshare.30949160>.
 27. Nevers Y, Warwick Vesztrocy A, Rossier V, Train CM, Altenhoff A, Dessimoz C, Glover NM. Quality assessment of gene repertoire annotations with OMArk. Nat Biotechnol. 2025;43(1):124–33.
 28. Chan PP, Lin BY, Mak AJ, Lowe TM. tRNAscan-SE 2.0: improved detection and functional classification of transfer RNA genes. Nucleic Acids Res. 2021;49(16):9077–96.
 29. Seemann T, Barnap. 2018. Available from: <https://github.com/tseemann/barrnap>.
 30. Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, Smit AF. RepeatModeler2 for automated genomic discovery of transposable element families. Proc Natl Acad Sci U S A. 2020;117(17):9451–7.
 31. Nakano M. Gene annotation of *Cladosporium tenuissimum* SHK1100. Figshare. 2026. <https://doi.org/10.6084/m9.figshare.31437025>.
 32. Cantalapiedra CP, Hernandez-Plaza A, Letunic I, Bork P, Huerta-Cepas J. eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. Mol Biol Evol. 2021;38(12):5825–9.
 33. Galperin MY, Makarova KS, Wolf YI, Koonin EV. Expanded microbial genome coverage and improved protein family annotation in the COG database. Nucleic Acids Res. 2015;43(Database issue):D261–269.
 34. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 2000;25(1):25–9.
 35. Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M. The KEGG resource for deciphering the genome. Nucleic Acids Res. 2004;32(Database issue):D277–280.
 36. Zheng J, Ge Q, Yan Y, Zhang X, Huang L, Yin Y. dbCAN3: automated carbohydrate-active enzyme and substrate annotation. Nucleic Acids Res. 2023;51(W1):W115–21.
 37. Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, Tosatto SCE, Paladin L, Raj S, Richardson LJ, et al. Pfam: The protein families database in 2021. Nucleic Acids Res. 2021;49(D1):D412–9.
 38. Buchfink B, Reuter K, Drost HG. Sensitive protein alignments at tree-of-life scale using DIAMOND. Nat Methods. 2021;18(4):366–8.
 39. Teufel F, Almagro Armenteros JJ, Johansen AR, Gislason MH, Pihl SI, Tsirigos KD, Winther O, Brunak S, von Heijne G, Nielsen H. SignalP 6.0 predicts all five types of signal peptides using protein language models. Nat Biotechnol. 2022;40(7):1023–5.
 40. Sperschneider J, Dodds PN. EffectorP 3.0: Prediction of Apoplastic and Cytoplasmic Effectors in Fungi and Oomycetes. Mol Plant Microbe Interact. 2022;35(2):146–56.
 41. Nakano M. Functional annotation of genes in *Cladosporium tenuissimum* SHK1100. Figshare. 2026. <https://doi.org/10.6084/m9.figshare.30948908>.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.