

High-quality genome assembly and annotation of *Clonostachys chloroleuca* strain Cc878 using Oxford Nanopore long-read sequencing

Xin Zhang,¹ Xinyuan Long,¹ Xiaoxing Xing,¹ Jia Tai,¹ Guanghui Wang,¹ Ming Xu,¹ Huiquan Liu¹

AUTHOR AFFILIATION See affiliation list on p. 3.

ABSTRACT As a noteworthy biocontrol fungus, *Clonostachys chloroleuca* currently lacks a high-quality reference genome. Here, we present the first high-quality genome assembly of *C. chloroleuca* strain Cc878 achieved through Oxford Nanopore Long-Read sequencing. The nuclear genome of Cc878 was assembled into four contigs, totaling 59.38 Mb.

KEYWORDS *Clonostachys*, genome assembly, gene annotation

Clonostachys spp. are biocontrol fungi with saprotrophic, mycotrophic, and plant endophytic lifestyles (1). These fungi can directly antagonize plant pathogens, promote plant growth, and induce resistance to protect plants from disease (2–4). While a high-quality genome assembly is available for the well-known species *C. rosea*, it is rare for other species in *Clonostachys*. Currently, the genomic assemblies for *C. chloroleuca* are limited to strains 67-1 and CBS 570.77, which were generated using Illumina technology (5, 6). These genome assemblies contain hundreds of contigs, diminishing their value as high-quality reference genomes for studying the molecular level of biocontrol mechanisms and conducting comparative genomics.

The *C. chloroleuca* strain Cc878 was originally isolated from soybean rhizosphere soil in Harbin City, Heilongjiang Province, China. The confirmation of Cc878 as *C. chloroleuca* was established through phylogenetic analysis of the ATP citrate lyase gene (GenBank: [KX184855.1](#)).

To provide a high-quality reference genome, we conducted *de novo* sequencing and assembly of the Cc878 genome using Oxford Nanopore Technology (Fig. 1). Cc878 was cultured on potato dextrose agar (PDA) medium at 25°C for 10 days. Spores were harvested by washing and filtering with sterile water. The spores were then used to inoculate yeast extract peptone dextrose (YEPD) medium to grow the mycelium. Genomic DNA was extracted using the SDS method (7), and its quality and quantity were assessed by agarose gel electrophoresis and Qubit 2.0 Fluorometer (Thermo Scientific), respectively.

Libraries were constructed with an insert size of ≥ 20 kb using a 1D ligation sequencing kit (catalog no. Q-SQK-LSK109) and sequenced on the PromethION platform. Base calling was performed using Guppy v6.2.7 (8), with the parameter “-c dna_r9.4.1_450bps_hac_prom.cfg.” A total of 1,352,886 filtered reads (10.31 Gb; 161× coverage) were obtained, with a mean length of 7,618 bp and an N₅₀ length of 10,471 bp.

The read correction, trimming, and genome assembly of Cc878 were conducted using Canu v2.2 (9) with default parameters. Mitochondrial DNA sequences were eliminated using BLASTn. Purge_dups v1.2.6 (10) was employed to remove redundant contigs in the primary assembly. To enhance assembly accuracy, long reads were initially aligned back

Editor Jennifer Geddes-McAlister, University of Guelph, Guelph, Canada

Address correspondence to Huiquan Liu, liuhuiquan@nwsuaf.edu.cn.

The authors declare no conflict of interest.

See the funding table on p. 3.

Received 5 April 2024

Accepted 24 May 2024

Published 20 June 2024

Copyright © 2024 Zhang et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](#).

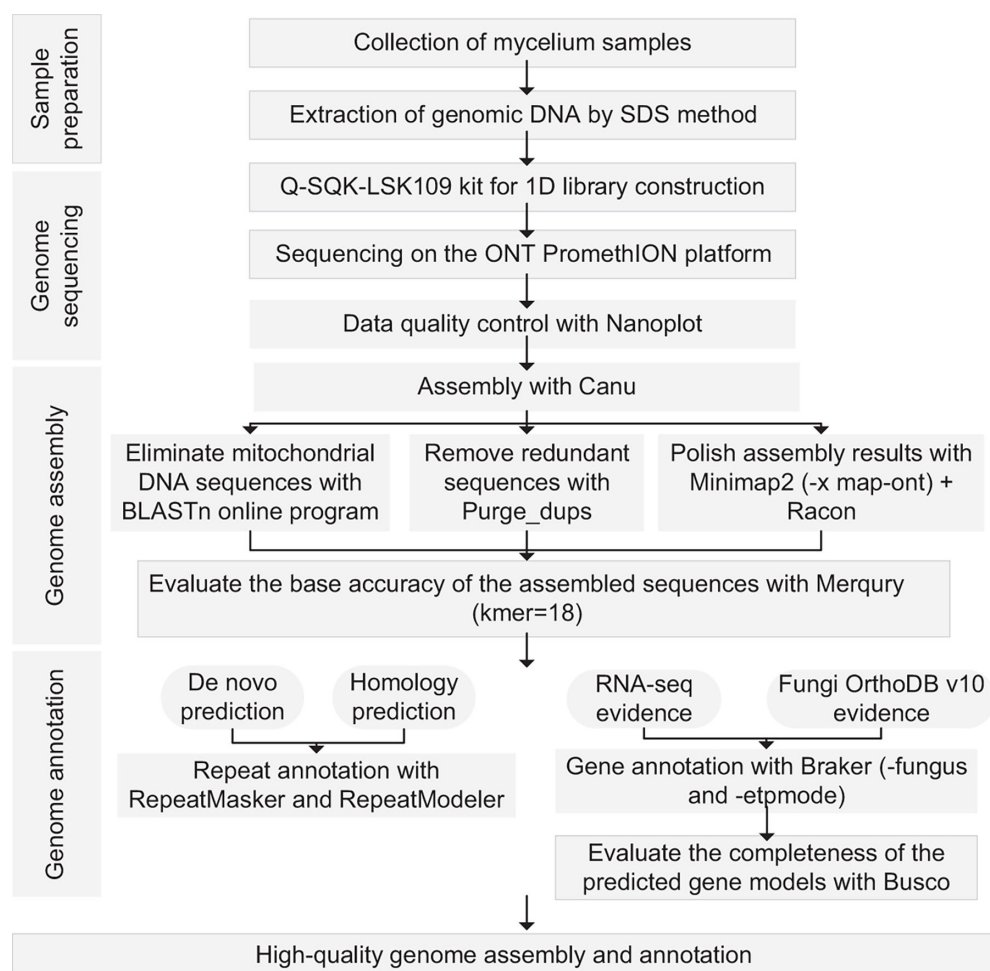


FIG 1 A schematic diagram that outlines the workflow for sample preparation, genome sequencing, assembly, annotation, and subsequent evaluation of the Cc878 genome.

to the assembly using Minimap2 v2.24 (11) with the parameter “-x map-ont” and the error correction was then performed using Racon v1.4.3 (12). Subsequently, the contigs were scaffolded using RagTag v2.1.0, and assembly quality was assessed using QUAST v5.0.2 (13) with default parameters.

The final genome assembly of Cc878 was 59.38 Mb, consisting of four contigs with an N_{50} of 37.47 Mb. The assembly had a GC content of 48.76% and a repeat content of 7.43%. According to Merqury V1.3 (14) assessment, the assembly had a consensus quality (QV) score of 45.75, corresponding to 99.99% accuracy at Q40, and a k-mer completeness score of 98.49%.

The BRAKER2 gene prediction pipeline was employed to predict genes using transcriptome evidence (RNA-seq data from *in vitro* vegetative growth and *in planta* saprophytic/endophytic growth) and homologous protein evidence (fungi OrthoDB v10 database) (15). Overall, we identified 21,504 protein-coding genes (>50 amino acids, Busco completeness score 98.94%) in the Cc878 genome.

ACKNOWLEDGMENTS

This work was supported by grants from the National Key R&D Program of China (2022YFD1400100).

AUTHOR AFFILIATION

¹State Key Laboratory for Crop Stress Resistance and High-Efficiency Production, College of Plant Protection, Northwest A&F University, Yangling, Shaanxi, China

AUTHOR ORCIDs

Huiquan Liu  <http://orcid.org/0000-0002-4723-845X>

FUNDING

Funder	Grant(s)	Author(s)
MOST National Key Research and Development Program of China (NKPs)	2022YFD1400100	Huiquan Liu

DATA AVAILABILITY

The whole-genome sequence has been deposited at DDBJ/ENA/GenBank under the accession number [JBANAY0000000000](https://doi.org/10.1128/JBANAY0000000000). The raw reads have been deposited in the SRA under the accession number [SRR27765574](https://doi.org/10.1128/SRR27765574).

REFERENCES

- Piombo E, Guaschino M, Jensen DF, Karlsson M, Dubey M. 2023. Insights into the ecological generalist lifestyle of *Clonostachys* fungi through analysis of their predicted secretomes. *Front Microbiol* 14:1112673. <https://doi.org/10.3389/fmicb.2023.1112673>
- Bahadoor A, Robinson KA, Loewen MC, Demissie ZA. 2023. *Clonostachys rosea* 'omics profiling: identification of putative metabolite-gene associations mediating its *in vitro* antagonism against *Fusarium graminearum*. *BMC Genomics* 24:352. <https://doi.org/10.1186/s12864-023-09463-6>
- Sun ZB, Li SD, Ren Q, Xu JL, Lu X, Sun MH. 2020. Biology and applications of *Clonostachys rosea*. *J Appl Microbiol* 129:486–495. <https://doi.org/10.1111/jam.14625>
- Mouekouba LDO, Zhang L, Guan X, Chen X, Chen H, Zhang J, Zhang J, Li J, Yang Y, Wang A. 2014. Analysis of *Clonostachys rosea*-induced resistance to tomato gray mold disease in tomato leaves. *PLoS One* 9:e102690. <https://doi.org/10.1371/journal.pone.0102690>
- Broberg M, Dubey M, Iqbal M, Gudmundsson M, Ihrmark K, Schroers H-J, Funck Jensen D, Brandström Durling M, Karlsson M. 2021. Comparative genomics highlights the importance of drug efflux transporters during evolution of mycoparasitism in *Clonostachys* subgenus *Bionectria* (fungi, ascomycota, hypocreales). *Evol Appl* 14:476–497. <https://doi.org/10.1111/eva.13134>
- Sun ZB, Sun MH, Li SD. 2015. Draft genome sequence of mycoparasite *Clonostachys rosea* strain 67-1. *Genome Announc* 3:e00546-15. <https://doi.org/10.1128/genomeA.00546-15>
- Lim HJ, Lee EH, Yoon Y, Chua B, Son A. 2016. Portable lysis apparatus for rapid single-step DNA extraction of *Bacillus subtilis*. *J Appl Microbiol* 120:379–387. <https://doi.org/10.1111/jam.13011>
- Wick RR, Judd LM, Holt KE. 2019. Performance of neural network basecalling tools for Oxford Nanopore sequencing. *Genome Biol* 20:129. <https://doi.org/10.1186/s13059-019-1727-y>
- Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res* 27:722–736. <https://doi.org/10.1101/gr.215087.116>
- Guan D, McCarthy SA, Wood J, Howe K, Wang Y, Durbin R. 2020. Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics* 36:2896–2898. <https://doi.org/10.1093/bioinformatics/btaa025>
- Li H. 2021. New strategies to improve minimap2 alignment accuracy. *Bioinformatics* 37:4572–4574. <https://doi.org/10.1093/bioinformatics/btab705>
- Vaser R, Sović I, Nagarajan N, Šikić M. 2017. Fast and accurate *de novo* genome assembly from long uncorrected reads. *Genome Res* 27:737–746. <https://doi.org/10.1101/gr.214270.116>
- 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>
- Rhie A, Walenz BP, Koren S, Phillippy AM. 2020. Merquy: reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biol* 21:245. <https://doi.org/10.1186/s13059-020-02134-9>
- Kriventseva EV, Kuznetsov D, Tegenfeldt F, Manni M, Dias R, Simão FA, Zdobnov EM. 2019. OrthoDB v10: sampling the diversity of animal, plant, fungal, Protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. *Nucleic Acids Res* 47:D807–D811. <https://doi.org/10.1093/nar/gky1053>