Microbiology Resource Announcements

Genomics and Proteomics | Announcement

The draft genome sequence of the ascomycete fungus *Colletotrichum acutatum* obtained from strawberry roots

Moytri RoyChowdhury,[1,2] Jake Sternhagen,[3] Yookyeong Kim[4]

AUTHOR AFFILIATIONS See affiliation list on p. 2.

ABSTRACT Here we report the genome sequence of the ascomycete pathogenic fungus *Colletotrichum acutatum* isolate #83 obtained from the rootstock of strawberries in Santa Maria, CA.

KEYWORDS fungi, genome analysis, genome sequencing, ascomycetes

Colletotrichum acutatum causes severe economic losses in the production of strawberries in California (1). The absence of genomic sequences of *C. acutatum* is a major hindrance in understanding the development and pathogenicity of this fungus, so this sequence information will facilitate future research into this key pathogen. Here we present the whole genome sequence of the *C. acutatum* isolate #83 obtained from the rootstock (underground stem/rhizome that can produce aboveground growth) of strawberries in Santa Maria, CA.

The fungus was grown from a pure-culture, strawberry rootstock-derived isolate #83 of *C. acutatum* was provided by Plant Sciences, Inc., Watsonville, CA, USA. A 0.6 cm diameter plug of developed hyphae was placed in the center of 3.9% potato dextrose agar (manufactured by Millipore Sigma P-2182) plates and grown in the dark at 27°C for 15 days.

In the DNA extraction, 15-day-old mycelium was used to extract DNA from the fungal isolates using the E.Z.N.A. Universal Pathogen Kit (Omega Biotek) with additional modifications for a more robust mechanical lysis with the OMNI disruptor, three cycles of freeze and thaw, and a hot phenol extraction. DNA was quantified using PicoGreen, obtaining 400 µL at approximately 10 ng/µL for a total of 4 µg of DNA. DNA concentration was measured using the QuantiFluor dsDNA System on a Quantus Fluorometer (Promega, Madison, WI, USA).

For the Illumina library prep, a Kapa Biosystems Hyper Prep kit (Kapa Biosystems, Wilmington, MA, USA) was used for whole genome library construction. One microgram of genomic DNA was fragmented using a Bioruptor sonicator (Diagenode, Denville, NJ, USA). DNA fragment ends were repaired, 3′ adenylated, and ligated to adapters. The resulting adapter-ligated libraries were PCR-amplified to create enough genetic material for sequencing. Illumina indexes were added and pooled for multiplexed sequencing on an Illumina HiSeq 2500 sequencer (Illumina, San Diego, CA, USA) using the paired-end 100 bp run format.

Sequencing statistics showed that a total of 12,447,195 paired reads were generated by Illumina sequencing that was estimated as a 17× fold coverage of the genome. FastQC was used to determine read quality, and it did not flag any sequencing reads as poor quality and estimated the GC content as 50%. The reads were filtered using Trimmomatic (v.0.30) (2) to remove primers, adapter sequences, and low-quality bases using default settings. The genome assembly was performed with SPAdes (v.3.10.0, 27 January 2017) (JAJKBH000000000). The genome assembly length was 49,743,065 bp and consisted of 1,998 contigs with an NGA50 of 57,457 bp (Table 1), as computed by QUAST-LG (3, 4). The

Editor Jason E. Stajich, University of California Riverside, Riverside, California, USA

Address correspondence to Moytri RoyChowdhury, moytrirc@gmail.com.

The authors declare no conflict of interest.

Received 14 January 2023
Accepted 29 October 2025
Published 1 April 2026

TABLE 1 Genome sequence of *C. acutatum* W537[a]

Statistics	Results
Genome statistics (contigs)	
Genome fraction (%)	77.192
Duplication ratio	1.001
Largest alignment	342,153
Total aligned length	40,233,184
NGA50	57,457
LGA50	232
Misassemblies	
No. of misassemblies	264
Length of misassembled contigs	24,608,249
Mismatches	
No. of mismatches/100 kbp	2,646.08
No. of indels/100 kbp	236.71
No. of N's/100 kbp	0
Statistics without reference	
No. of contigs	1,172
Largest contig	909,718
Total length	49,743,065
Total length (≥1,000 bp)	49,530,365
Total length (≥10,000 bp)	48,498,476
Total length (≥50,000 bp)	42,041,346

[a]Regarding all software, default parameters were used except where otherwise noted.

reference genome that was compared to the alignment as part of the QUAST-LG run was *Colletotrichum scovillei* strain (PRJNA314171) (5).

NGA50 are analogous to N50 and NG50 where the contigs are replaced by blocks that can be aligned to the reference. Correctness is measured by detecting misassemblies such as mismatches, indels, and misjoins (6). Our analysis pipeline at the time of sequencing this sample did not include BUSCO assessments.

ACKNOWLEDGMENTS

This work was supported by the INBRE Program NIH (grant number P20 GM103408) and SBOE/ISU STEM (grant number AHRC38).

AUTHOR AFFILIATIONS

[1]California Department of Public Health, Infectious Diseases Program, Richmond, California, USA
[2]Department of Biological Sciences, Idaho State University, Pocatello, Idaho, USA
[3]Riverside School of Medicine; SOM Education Building, University of California, Riverside, California, USA
[4]UC Berkeley Extensions, Berkeley, California, USA

AUTHOR ORCIDs

Jake Sternhagen http://orcid.org/0009-0004-3006-6550
Yookyeong Kim http://orcid.org/0009-0007-4537-5729

AUTHOR CONTRIBUTIONS

Moytri RoyChowdhury, Conceptualization, Data curation, Formal analysis, Funding acquisition, Project administration, Resources, Supervision, Validation, Visualization, Writing – original draft | Jake Sternhagen, Data curation, Formal analysis, Funding acquisition, Investigation, Resources, Software, Supervision, Validation, Visualization,

Writing – original draft | Yookyeong Kim, Resources, Visualization, Writing – review and editing

DATA AVAILABILITY

This Whole Genome Shotgun project has been deposited in GenBank (accession no. JAJKBH000000000 and SRA accession no. SRR16974093). The version described in this paper is the first version (JAJKBH010000000, BioProject no. PRJNA779114). The data of the reference genome have been deposited with links to BioProject accession number PRJNA314171 in the NCBI BioProject database (https://www.ncbi.nlm.nih.gov/bioproject/).

REFERENCES

1. Gaines C. 2005. Anthracnose on strawberry. Amer. Soc. Hort. Sci., Video Workshop Series, Disk 2, An effort to get the attention of researchers who work with *Colletotrichum acutatum*. ASHS, Alexandria, VA.
2. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170
3. Gurevich A, Mikheenko A, The QUAST team. 2024. QUAST: quality assessment tool for genome assemblies (Version 5.3.0) [Computer software]. Zenodo. https://zenodo.org/records/14062419.
4. 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
5. Han J-H, Chon J-K, Ahn J-H, Choi I-Y, Lee Y-H, Kim KS. 2016. Whole genome sequence and genome annotation of *Colletotrichum acutatum*, causal agent of anthracnose in pepper plants in South Korea. Genom Data 8:45–46. https://doi.org/10.1016/j.gdata.2016.03.007
6. Alhakami H, Mirebrahim H, Lonardi S. 2017. A comparative evaluation of genome assembly reconciliation tools. Genome Biol 18:93. https://doi.org/10.1186/s13059-017-1213-3