Whole-Genome Sequence Resource of *Calonectria ilicicola*, the Casual Pathogen of Soybean Red Crown Rot

Hsien-Hao Liu,¹ Jie Wang,² Ping-Hu Wu,¹ Mei-Yeh Jade Lu,³ Jeng-Yi Li,³ Yuan-Min Shen,⁴ Min-Nan Tzeng,⁵ Chang-Hsin Kuo,⁶ Ying-Hong Lin,^{7,8} and Hao-Xun Chang^{1,†}

¹ Department of Plant Pathology and Microbiology, National Taiwan University, Taipei City 10617, Taiwan

² Center for Genomics-Enabled Plant Science, Michigan State University, East Lansing, MI, U.S.A.

³ NGS High Throughput Genomics Core, Biodiversity Research Center, Academia Sinica, Taipei City 11529, Taiwan

⁴ Taichung District Agricultural Research and Extension Station, Council of Agriculture, Changhua County 51544, Taiwan

⁵ Kaohsiung District Agricultural Research and Extension Station, Council of Agriculture, Pingtung County 90846, Taiwan

⁶ Department of Plant Medicine, National Chiayi University, Chiayi City 60004, Taiwan

⁷ Department of Plant Medicine, National Pingtung University of Science and Technology, Pingtung 912301, Taiwan

⁸ Plant Medicine Teaching Hospital, General Research Service Center, National Pingtung University of Science and Technology, Pingtung, Taiwan

Funding

This work was supported by the Taiwan Ministry of Education YuShan Young Scholar Program and the National Taiwan University to H.-X. Chang.

Keywords

Calonectria ilicicola, Oxford Nanopore, red crown rot, soybean (*Glycine max*), whole-genome sequencing

Abstract

Calonectria ilicicola (anamorph: *Cylindrocladium parasiticum*) is a soilborne plant-pathogenic fungus with a broad host range, and it can cause red crown rot of soybean and *Cylindrocladium* black rot of peanut, which has become an emerging threat to crop production worldwide. Limited molecular studies have focused on *Calonectria ilicicola* and one of the possible difficulties is the lack of genomic resources. This study presents the first high quality and near-completed genome of *C. ilicicola*, using the Oxford Nanopore GridION sequencing platform. A total of 16 contigs were assembled and the genome of *C. ilicicola* isolate F018 was estimated to have 11 chromosomes. Currently, the *C. ilicicola* F018 genome represents the most contiguous assembly, which has the lowest contig number and the highest contig N₅₀ among all *Calonectria* genome resources. Putative protein-coding sequences and secretory proteins were estimated to be 17,308 and 1,930 in the *C. ilicicola* F018 genome, respectively; and the prediction was close to other plant-pathogenic fungi, such as *Fusarium* species, within the Nectriaceae family. The availability of this high-quality genome resource is expected to facilitate research on fungal biology and genetics of *C. ilicicola* and to support advanced understanding of pathogen virulence and disease management.

†Corresponding author: H.-X. Chang; hxchang@ntu.edu.tw

Hsien-Hao Liu and Jie Wang contributed equally to the project.

Current address for Jie Wang: Department of Energy Joint Genome Institute, Berkeley, CA, U.S.A.

The author(s) declare no conflict of interest.

*The e-Xtra logo stands for “electronic extra” and indicates there are supplementary materials published online.

Accepted for publication 24 February 2021.

Copyright © 2021 The Author(s). This is an open access article distributed under the CC BY-NC-ND 4.0 International license.

Genome Announcement

The genus *Calonectria* was first documented in 1867 (Rossman 1979) and more than 400 *Calonectria* species have been recorded in the MycoBank database. The species *Calonectria illicicola* was first defined in 1950 (Boedijn and Reitsma 1950), and this species had a synonym as *Cercospora theae* var. *crotalariae*, which was first described as a plant pathogen to cause collar rot on crotalaris (*Crotalaria anagyroides*) and leaf spot on tea (*Thea sinensis*) (Boedijn and Reitsma 1950; Loos 1950). The pathogen was also described as *Calonectria crotalariae*, which causes root infection and Cylindrocladium black rot on peanut (*Arachis hypogaea*) (Bell and Sobers 1966). Due to the morphological similarities and the criteria of fungal nomenclature (Crous et al. 1993), the teleomorph *Calonectria illicicola* and the anamorph *Cylindrocladium parasiticum* have become the accepted names for this pathogen.

While *C. illicicola* was first found to cause Cylindrocladium black rot on peanut in southwest Georgia in the United States in 1965 (Bell and Sobers 1966), the case of infecting soybean and causing red crown rot was first reported in Japan in 1968 (Krigsvold et al. 1977). Later, soybean red crown rot was discovered in Louisiana in 1976 (Berner et al. 1986) and the yield losses on soybean and peanut were estimated to be more than 50% when the susceptible cultivars were cultivated under a moist environment or flooding situation (Bell et al. 1973; Berner et al. 1988; Pataky et al. 1983). Recently, soybean red crown rot has been distributed to other soybean production areas, such as Illinois (Kleczewski et al. 2019), Taiwan (Liu et al. 2020), and other regions of southern China (Gai et al. 2017). Isolates of *C. illicicola* from different hosts were found to express varied levels of virulence, suggesting potential virulence variability existed in the *C. illicicola* population (Gai et al. 2017). However, molecular tools and genomic resources are limited for studying *C. illicicola*, and no study has been focused on the molecular interactions between *C. illicicola* and soybean. With the emerging threat of *C. illicicola* on soybean and peanut production worldwide, a high-quality *C. illicicola* genome is desired to assist molecular and genetic research on the virulence mechanism of *C. illicicola*. Accordingly, this study aims to present a high-quality whole-genome sequencing resource for *C. illicicola* to the community of interest.

The *C. illicicola* isolate F018 was isolated from diseased soybean plants in Wandan, Pingtung County, Taiwan (Liu et al. 2020). This isolate was grown in potato dextrose broth (PDB) at 25°C with 150 rpm shaking for 7 days. The mycelia were harvested by filtering through double-layer gauze and the mycelial were ground into powder with liquid nitrogen. Mycelial DNA was isolated by a modified CTAB method (Lee et al. 1988). Briefly, the powder of mycelia was added to CTAB extraction buffer and incubated at 65°C for 30 min, and the DNA was purified by chloroform and precipitated by isopropanol. After 75% ethanol washing twice, the mycelial DNA was dissolved in 50 µl of pure water. The carryover of degraded RNA was first removed from DNA using the ProNex size-selective purification system (Promega). Subsequently, high-molecular weight genomic DNA (gDNA) was obtained by removing fragments shorter than 3 Kbp, using Agencourt AMPure XP (Beckman Coulter Inc.) (Supplementary Fig. S1), which resulted in a high-molecular weight gDNA containing 24.9% fragments in 3 to 10 Kbp and 72.6% fragments in 10 to 100 Kbp. The high-molecular weight gDNA was repaired using the PreCR repair mix (New England Biolabs) before being subjected to library construction using an Oxford Nanopore Technologies gDNA library preparation kit. The high-molecular weight gDNA was sequenced in one flow cell (FLO-MIN106) on the GridION sequencing platform (Oxford Nanopore Technologies). Raw sequencing signals were processed to convert into the FASTQ format using Guppy basecaller version 4.0.11 (Wick et al. 2019). PoreChop version 0.2.4 was used to remove sequencing adapter and index sequences from the raw reads. The FilTlong version 0.2.0 was used to remove reads shorter than 10 kb and reads with low quality. To extract mitochondrial sequences, BBTools version 38.86 (Bushnell et al. 2017) was used to separate the mitochondrial reads from the raw reads, using the *k*-mer matching method against the National Center for Biotechnology Information (NCBI) RefSeq mitochondrial database. The filtered main chromosome and mitochondrial reads were used for main genome and mitochondrial genome assembly, respectively, using assembler Flye version 2.81 (Kolmogorov et al. 2019). The initial draft genome and mitochondrial genome assemblies were polished with three rounds Racon version 1.4.3 (Vaser et al. 2017). In addition, Medaka version 1.1.2 was used to further polish both

Table 1. Nanopore sequencing and genome assembly statistics

Genome features		Statistics		
Nanopore-generated reads		2.76 M		
QC passed bases		20.1 Gbp		
Genome size		68,974,662 bp		
GC content (%)		46.58		
Contigs >50 Kbp		15		
BUSCO completeness		98.6%		
Augustus v3.2.3 protein-coding sequences		17,308		
SignalP v5.0 secretory proteins		1,930		
Gene	Length (bp)	GC (%)	Telomere	Chromosomal level
Calil-F018_Ch01	10,133,255	49.46	Both ends	Completed
Calil-F018_Ch02	8,016,852	46.68	Both ends	Completed
Calil-F018_Ch03	8,005,679	48.73	Both ends	Completed
Calil-F018_Ch04	7,244,835	48.49	Both ends	Completed
Calil-F018_Ch05	6,053,737	45.76	Both ends	Completed
Calil-F018_Ch06	5,673,255	42.16	Both ends	Completed
Calil-F018_Ch07	5,253,079	46.56	Both ends	Completed
Calil-F018_Contig08	5,190,824	47.22	Single end	Near-completed
Calil-F018_Ch09	4,131,236	45.86	Both ends	Completed
Calil-F018_Ch010	4,035,728	44.14	Both ends	Completed
Calil-F018_Contig11	2,948,775	41.18	Single end	Near-completed
Calil-F018_Contig12	1,333,656	45.33	Single end	Near-completed
Calil-F018_Contig13	871,732	44.34	Single end	Near-completed
Calil-F018_Contig14	59,341	20.35	No	Centromere
Calil-F018_Contig15	17,578	27.19	No	Accessory or incomplete
Calil-F018_Contig16	5,100	54.57	No	Accessory or incomplete

assemblies, using the pretrained model r941_prom_high_g4011. To evaluate genome assembly quality, QUAST version 5.0.2 (Gurevich et al. 2013) was used to generate assembly summary statistics and BUSCO version 4.1.4 (Simão et al. 2015) was used to evaluate genome assembly completeness, using the *Hypocreales_odb10* lineage gene dataset. The protein-coding sequences were predicted by Augustus version 3.2.3, using the *ab initio* method with *Fusarium graminearum* as a model (Stanke and Morgenstern 2005), and the putative secretory proteins were predicted by SignalP version 5.0 (Almagro Armenteros et al. 2019).

A total of 16 contigs were assembled with a total genome size of 68.97 Mbp and N_{50} contig size of 6.05 Mbp (Table 1), which makes *C. ilicicola* F018 genome the most contiguous assembly compare with other isolates sequenced by Illumina platform (Supplementary Tables S1 and S2). Telomere (TTAGGG) repetitive sequences were detected in nine contigs on both ends and four contigs on a single end. Based on the assembly graph (Supplementary Fig. S2), the four contigs (Calil-F018_Contig08, Calil-F018_Contig11, Calil-F018_Contig12, and Calil-F018_Contig13) with telomere sequence on one end were connected with a low-GC (Calil-F018_Contig14, GC = 20.35%) centromere sequence. While the accurate chromosomal structure requires additional high-quality long reads or a genetic map to resolve, these four contigs are predicted to form two chromosomes in three possible combinations: i) Calil-F018_Contig08-Contig14-Contig11 and Calil-F018_Contig12-Contig14-Contig13, or ii) Calil-F018_Contig08-Contig14-Contig12 and Calil-F018_Contig11-Contig14-Contig13, or iii) Calil-F018_Contig08-Contig14-Contig13 and Calil-F018_Contig11-Contig14-Contig12. Therefore, there are a total of 11 chromosomes estimated in the *C. ilicicola* F018 genome and the current assembly contains 16 contigs.

To evaluate the completeness of this genome assembly, the BUSCO (benchmarking universal single-copy orthologs) score of the *Hypocreales_odb10* gene dataset was determined to be 98.6% completed BUSCOs, 0.4% fragmented BUSCOs, and 1.0% missing BUSCOs. The repetitive ribosomal DNA elements were detected on the terminal of Calil-F018_Ch02. The putative protein-coding sequences and secretory proteins in the *C. ilicicola* F018 genome (17,308 and 1,930, respectively) were slightly fewer than *Fusarium oxysporum* f. sp. *lycopersici* 4287 v2.0 (27,347 and 2,311, respectively) and *Fusarium verticillioides*

7600 v2.0 (20,533 and 1,891, respectively) but were more than *F. graminearum* (13,332 and 1,424, respectively) and *Nectria hematococca* MPVI v2.0 (15,707 and 1,475, respectively) (Chang et al. 2018). Based on the BUSCOs and gene predictions, the genome assembly quality of *C. ilicicola* F018 is at the near-completed level within the Hypocreales order and Nectriaceae family. The availability of this high-quality genome resource will accelerate studies on *C. ilicicola* regarding fungal genomics or transcriptomics, fungal knockout and genome editing, as well as evolution and population genetics of *C. ilicicola*. This advanced understanding on *C. ilicicola* will further improve strategies on disease management. The genome sequences have been deposited in the NCBI BioProject PRJNA672291 and BioSample SAMN16562416.

Acknowledgments

We thank C.-H. Wu in the Institute of Plant and Microbial Biology, Academia Sinica, for assisting with Oxford Nanopore genome sequencing.

Author-Recommended Internet Resources

Mycobank database: <https://www.mycobank.org>

Porechop: <https://github.com/rwrick/Porechop>

Filtlong: <https://github.com/rwrick/Filtlong>

Medaka: <https://github.com/nanoporetech/medaka>

Literature Cited

- Almagro Armenteros, J. J., Tsirigos, K. D., Sønderby, C. K., Petersen, T. N., Winther, O., Brunak, S., von Heijne, G., and Nielsen, H. 2019. SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat. Biotechnol.* 37:420-423.
- Bell, D., Locke, B. J., and Thompson, S. S. 1973. The status of *Cylindrocladium* black rot of peanut in Georgia since its discovery in 1965. *Plant Dis. Rep.* 57:90-94.
- Bell, D. K., and Sobers, E. K. 1966. A peg, pod, and root necrosis of peanuts caused by a species of *Calonectria*. *Phytopathology* 56:1361-1364.
- Berner, D. K., Berggren, G. T., Pace, M. E., White, E. P., Gershay, J. A., Freedman, J. A., and Snow, J. P. 1986. Red crown rot: Now a major disease of soybeans. *Louisiana Agric.* 29:4-5.
- Berner, D. K., Berggren, G. T., Snow, J. P., and White, E. P. 1988. Distribution and management of red crown rot of soybean in Louisiana. *Louisiana Appl. Agric. Res.* 3:160-166.
- Boedijn, K. B., and Reitsma, J. 1950. Notes on the genus *Cylindrocladium*. *Reinwardtia* 5:51-60.
- Bushnell, B., Rood, J., and Singer, E. 2017. BBMerge—Accurate paired shotgun read merging via overlap. *PLoS One* 12:e0185056.
- Chang, H.-X., Noel, Z. A., Sang, H., and Chilvers, M. I. 2018. Annotation resource of tandem repeat-containing secretory proteins in sixty fungi. *Fungal Genet. Biol.* 119:7-19.
- Crous, P. W., Wingfield, M. J., and Alfenas, A. 1993. *Cylindrocladium parasiticum* sp. nov., a new name for *C. crotalariae*. *Mycol. Res.* 97:889-896.
- Gai, Y., Deng, Q., Chen, X., Guan, M., Xiao, X., Xu, D., Deng, M., and Pan, R. 2017. Phylogenetic diversity of *Calonectria ilicicola* causing *Cylindrocladium* black rot of peanut and red crown rot of soybean in southern China. *J. Gen. Plant Pathol.* 83:273-282.
- Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. 2013. QUAST: Quality assessment tool for genome assemblies. *Bioinformatics* 29:1072-1075.
- Kleczewski, N., Plewa, D., Kangas, C., Phillippi, E., and Kleczewski, V. 2019. First report of red crown rot of soybeans caused by *Calonectria ilicicola* (anamorph: *Cylindrocladium parasiticum*) in Illinois. *Plant Dis.* 103:1777.
- Kolmogorov, M., Yuan, J., Lin, Y., and Pevzner, P. A. 2019. Assembly of long, error-prone reads using repeat graphs. *Nat. Biotechnol.* 37:540-546.
- Krigsvold, D. T., Garren, K. H., and Griffin, G. J. 1977. Importance of peanut field cultivation and soybean cropping in the spread of *Cylindrocladium crotalariae* within and among peanut fields. *Plant Dis. Rep.* 61:495-499.
- Lee, S., Milgroom, M., and Taylor, J. 1988. A rapid, high yield mini-prep method for isolation of total genomic DNA from fungi. *Fungal Genet. Newsl.* 35:23-24.
- Liu, H. H., Shen, Y. M., Chang, H.-X., Tseng, M. N., and Lin, Y. H. 2020. First report of soybean red crown rot caused by *Calonectria ilicicola* in Taiwan. *Plant Dis.* 104:979.
- Loos, C. A. 1950. *Calonectria theae* n.sp.—The perfect stage of *Cercosporaella theae* Petch. *Trans. Br. Mycol. Soc.* 33:13-18.
- Pataky, J., Beute, M., Wynne, J. C., and Carlson, G. A. 1983. Peanut yield, market quality and value reductions due to *Cylindrocladium* black rot. *Peanut Sci.* 10:62-66.
- Rossman, A. Y. 1979. *Calonectria* and its type species, *Calonectria daldiniana*, a later synonym of *Calonectria pyrochroa*. *Mycotaxon* 8:321-328.
- Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., and Zdobnov, E. M. 2015. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31:3210-3212.
- Stanke, M., and Morgenstern, B. 2005. AUGUSTUS: A web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Res.* 33:W465-W467.
- Vaser, R., Sović, I., Nagarajan, N., and Šikić, M. 2017. Fast and accurate *de novo* genome assembly from long uncorrected reads. *Genome Res.* 27:737-746.
- Wick, R. R., Judd, L. M., and Holt, K. E. 2019. Performance of neural network basecalling tools for Oxford Nanopore sequencing. *Genome Biol.* 20:129.