MPMI Vol. 33, No. 7, 2020, pp. 880–883. https://doi.org/10.1094/MPMI-03-20-0060-A

RESOURCE ANNOUNCEMENT

A Chromosome-Scale Genome Assembly Resource for *Myriosclerotinia sulcatula* Infecting Sedge Grass (*Carex* sp.)

Stefan Kusch,[1,†] Heba M. M. Ibrahim,[1,2] Catherine Zanchetta,[3] Celine Lopez-Roques,[3] Cecile Donnadieu,[3] and Sylvain Raffaele[1,†]

[1] Laboratoire des Interactions Plantes-Microorganismes (LIPM), INRA, CNRS, 24 Chemin de Borde Rouge - Auzeville, CS52627, F31326 Castanet Tolosan Cedex, France
[2] Department of Genetics, Faculty of Agriculture, Cairo University, Gammaa St., 12613 Giza, Egypt
[3] INRAE, US 1426, GeT-PlaGe, Genotoul, Castanet-Tolosan, France

Abstract

The fungus *Myriosclerotinia sulcatula* is a close relative of the notorious polyphagous plant pathogens *Botrytis cinerea* and *Sclerotinia sclerotiorum* but exhibits a host range restricted to plants from the *Carex* genus (*Cyperaceae* family). To date, there are no genomic resources available for fungi in the *Myriosclerotinia* genus. Here, we present a chromosome-scale reference genome assembly for *M. sulcatula*. The assembly contains 24 contigs with a total length of 43.53 Mbp, with scaffold N_{50} of 2,649.7 kbp and N_{90} of 1,133.1 kbp. BRAKER-predicted gene models were manually curated using WebApollo, resulting in 11,275 protein-coding genes that we functionally annotated. We provide a high-quality reference genome assembly and annotation for *M. sulcatula* as a resource for studying evolution and pathogenicity in fungi from the *Sclerotiniaceae* family.

Funding

This work was supported by a Starting grant from the European Research Council (ERC-StG-336808) to S. Raffaele and the French Laboratory of Excellence project TULIP (ANR-10-LABX-41 and ANR-11-IDEX-0002-02) and by France Génomique National infrastructure, funded as part of "Investissement d'avenir" program managed by Agence Nationale pour la Recherche (contract ANR-10-INBS-09).

Keywords

fungal pathogen, grass plants, genome, host specific, long read sequencing, *Myriosclerotinia sulcatula*, proteomics, Sclerotinia

Genome Announcement

Fungal plant pathogens are a major threat to global food security. The notorious necrotrophic plant pathogens *Sclerotinia sclerotiorum*, the causal agent of the white mold disease, and the gray mold disease-causing fungus *Botrytis cinerea* are the two most devastating pathogens of the family *Sclerotiniaceae*; both infect more than 600 plant species, including important oil and vegetable crops (Boland and Hall 1994). However, the *Sclerotiniaceae* family also includes species restricted to a narrow range of host plants, saprobic species, and biotrophic lineages (Navaud et al. 2018). Species from the genus *Myriosclerotinia* have a host range restricted to plants from the *Cyperaceae* and *Juncaceae* families of grasses (Schumacher and Kohn 1985). The nutritional habit of *Myriosclerotinia* species was proposed to be biotrophic but remains poorly documented (Schumacher and Kohn 1985). The *Myriosclerotinia* genus forms a monophyletic clade separate from the *Botrytis* and *Sclerotinia* genera within the *Sclerotiniaceae* family (Navaud et al. 2018).

[†]Corresponding authors: S. Kusch; skusch@bio1.rwth-aachen.de and S. Raffaele; sylvain.raffaele@inra.fr

Current address for Stefan Kusch: Unit of Plant Molecular Cell Biology, Institute for Biology I, RWTH Aachen University, Worringerweg 1, 52056 Aachen, Germany

Current address for Heba M. M. Ibrahim: Division of Plant Biotechnics, Department of Biosystems, Faculty of Bioscience Engineering, Katholieke Universiteit Leuven, Willem de Croylan 42, 3001 Leuven, Belgium

The author(s) declare no conflict of interest.

Accepted for publication 16 April 2020.

M. sulcatula Schumacher & Kohn is a facultative pathogen colonizing culms of sedge plants (*Carex* and *Shoenus* spp.) and are recorded across North America and Europe (Schumacher and Kohn 1985). As a close relative of *B. cinerea* and *S. sclerotiorum*, *M. sulcatula* is particularly interesting for exhibiting a very distinctive lifestyle and host range. Here, we present a chromosome-scale reference assembly and a manually curated reference gene annotation for *M. sulcatula* isolate MySu01 isolated from *Carex* culm.

The fungal strain *Myriosclerotinia sulcatula* MySu01, collected from *Carex* culm in the Naturpark Bayerischer Wald, Zwiesel (Germany) in 1989, was obtained from the Center for Biological Sequence Analysis Fungal Biodiversity Center (CBS-KNAW) microbial collection (strain CBS 145.92). The fungus was cultivated on potato dextrose agar (PDA) at 23°C or was stored on PDA at 4°C. For high–molecular weight DNA isolation and Oxford Nanopore Technologies sequencing, *M. sulcatula* was grown in potato dextrose broth at 23°C for 3 days. Next, the mycelium was incubated in 1.5% (wt/vol) Lysin enzyme mixture from *Trichoderma harzianum* (Sigma-Aldrich) in 50 ml of enzymatic digestion solution (0.8 M mannitol, 200 mM citric acid/trisodium citrate buffer, pH 5.5) for 3 h. Then, the suspension was filtered through 100-μm cell strainers (Falcon Corning) and was centrifuged at 8,000 × *g* for 2 min to collect fungal material. The pellet was resuspended in 500 μl of Tris-EDTA buffer (pH 8.0). DNA was extracted with the MagAttract HMW DNA kit (Qiagen), following instructions from the manufacturer with the following adjustments: 80 μl of Proteinase K, 20 μl of RNase A, and 600 μl of buffer AL for cell lysis were used. After lysis, the solution was filtered through Miracloth before adding 25 μl of MagAttract suspension G beads and 600 μl of buffer MB.

Library preparation and sequencing were performed at the GeT-PlaGe core facility, INRAE Toulouse, France, according to manufacturer instructions "1D gDNA selecting for long reads (SQK-LSK108)". At each step, DNA was quantified using the Qubit dsDNA HS assay kit (Life Technologies). DNA purity was tested using a Nanodrop (Thermo Fisher Scientific), and size distribution and degradation were assessed using the Fragment Analyzer (AATI) high sensitivity DNA fragment analysis kit. Purification steps were performed using AMPure XP beads (Beckman Coulter).

For one flow cell, 5 μg of DNA was purified and was then sheared at 10 kb, using the g-TUBE system (Covaris). A DNA damage repair step was performed on 2 μg of sample. Then an END-repair and dA-tailing of double-stranded DNA fragments was performed and adapters were ligated to the library. Library was loaded onto one R9.4 flow cell and was sequenced on a GridION instrument at 0.017 pmol within 48 h.

The *M. sulcatula* sample for RNA sequencing was collected from mycelium (30 mm diameter) on PDA plates by peeling 5 mm of the mycelial front from the agar and flash-freezing the material in liquid nitrogen. The frozen samples were ground with metal beads (2.5 mm diameter) in a Retschmill apparatus (24 Hz, 2× for 1 min). RNA was isolated using the RNeasy mini RNA extraction kit (Qiagen), following the manufacturer instructions. Genomic DNA was removed via DNase treatment (TURBO DNase; Ambion, Thermo Fisher Scientific). Quality and concentrations of RNA were assessed with Agilent bioanalyzer nano chips. RNA sequencing (RNA-seq) was conducted by Fasteris on a HiSeq2500 sequencer and yielded 125 bp paired-end reads. The reads were quality- and adapter-trimmed using Trimmomatic v0.36, applying the following settings: ILLUMINACLIP:TruSeq2-PE.fa:5:30:10 SLIDINGWINDOW: 3:18 LEADING:6 TRAILING:6 MINLEN:90 (Bolger et al. 2014). The total number of raw read pairs was 27,860,646; 14,680,505 read pairs remained after quality trimming. The quality of the RNA-seq data were assessed via FastQC v0.11.5 (Babraham Bioinformatics). The reads were mapped, to the new reference assembly of *M. sulcatula*, using HISAT2 (Kim et al. 2015) with –max-intronlen 500 -k 1; output files were parsed with samtools v1.7 (Li et al. 2009) and bedtools v2.25.0 (Quinlan and Hall 2010) to obtain BAM and coverage files.

We conducted the genome assembly using the long reads from Oxford Nanopore Technologies sequencing with Canu v1.6 (Koren et al. 2017); there were 539,594 reads (3,983,977,882 bp), 140,671 reads (1,571,817,355 bp) remained after filtering. Then, we performed three rounds of polishing of the genome assembly using Illumina short read DNA sequencing data (125-bp paired-end, containing 22,571,551 raw read pairs and 19,973,645 read pairs after trimming; approximately 35.8-fold coverage) with Pilon v1.22 (Walker et al. 2014). We assessed the completeness of the genome assembly with BUSCO v3 (Simão et al. 2015) against the ascomycete_odb9 database. RepeatMasker v4.0.7 (Smit et al. 2016) with the repeat database version RepBase-20170127 was used to quantify the extent of repetitive sequences and transposable elements (TEs) and mask these sequences in the assembly.

Table 1. *Myriosclerotinia sulcatula* genome assembly statistics[a]

Feature	*Sclerotinia sclerotiorum*	*Myriosclerotinia sulcatula*
Genome (bp)	38,906,597	43,533,135
Estimated size (kmer)	–	45.2 Mb
Genomic contigs	–	24
Mitochondrial contigs	1	2
Mitochondrial genome (bp)	132,532	330,647
Full chromosomes	16	6
Number chromosomes	16	14[b]
Largest contig (bp)	3,951,982	5,732,858
Average length (bp)	2,431,662	1,813,881
GC content	41.6%	39.8%
N_{50}	2,434,682	2,649,720
N_{90}	1,815,632	1,133,061
BUSCO[b]	–	S:97.6% F:1.0%
Repeat elements	6.2%	3.6%
Genes (BRAKER2)	–	11,631
Genes (curated)	11,130	11,275
BUSCO[c] annotation	99.3%	S:98.1% F:1.2%

[a] Statistics for the *S. sclerotiorum* 1980 complete genome assembly (Derbyshire et al. 2017) are provided for comparison purposes.

[b] Estimated by number of telomeric repeats in the assembly.

[c] Using the database for ascomycete core genes (ascomycota_odb9). S, complete single-copy BUSCOs, F, fragmented BUSCOs.

We used BRAKER (Hoff et al. 2016) for unsupervised gene prediction, with Augustus v3.2.3 trained on the *S. sclerotiorum* 1980 annotation dataset (Derbyshire et al. 2017) and the RNA-seq data for *M. sulcatula* generated in this study. All gene models were inspected manually, using WebApollo (Lee et al. 2013). Functional annotation was performed using the hmmscan algorithm of HMMER v3.2.1 (Eddy 2011). Secreted proteins were identified by SignalP v4.1 (Nielsen 2017) and proteins with transmembrane domains were predicted with TMHMM v2.0 (Krogh et al. 2001).

The Canu assembly yielded 26 contigs between 5,732,858 and 13,241 bp in length and a total genomic length of 43,863,912 bp (Table 1). Two contigs (tig00000103, tig00000120) had BLASTN hits for mitochondrial DNA, totaling 330,647 bp of sequence. RNAweasel prediction with the genetic code for yeast mitochondria identified 42 transfer RNAs without introns, two rnpB, and the 5′ portion of rns on these contigs. The larger mitochondrial contig, tig00000103, has 94.3% identity with the *S. sclerotiorum* 1980 UF-70 mitochondrion (KT283062.1, query cover 29%, E value 0.0), while tig00000120 has 82.2% identity with only 3% query coverage (E value 7e-37); E values < E-50 are considered high quality and E values < E-5 are considered good hits. The *M. sulcatula* draft nuclear genome of 24 contigs had a length of 43,533,135 bp, with an N_{50} of 2,649,720, an N_{90} of 1,133,061, and a GC content of 39.7%. Genome size estimation, using jellyfish (Marçais and Kingsford 2011) with 30-mers, suggested a total genome size of 45.2 Mbp, which is very close to the size of our final assembly. Our assembly had a BUSCO score of 96.3% completeness (Simão et al. 2015), and approximately 3.4% of the genome consisted of repeat elements or TEs, according to RepeatMasker v4.0.7. Twenty-one of the 24 contigs had high similarity with *Sclerotiniaceae* after BLAST searches against the National Center for Biotechnology Information (NCBI) nucleotide (searches performed December 2019); only the three smallest contigs (tig00008974, tig00008975, and tig00008977) did not have a significant hit in the database. We found 28 cases of canonical telomeric repeat sequences (5′-TTAGGG-3′ hexamers), suggesting that *M. sulcatula* possesses 14 chromosomes, six of which may be complete in our assembly, since we found telomeric hexamers to both ends.

We then used BRAKER (Hoff et al. 2016) for *ab initio* gene prediction, using the RNA-seq sample obtained from mycelium grown on PDA as a guide. This approach predicted 11,628 gene models. We manually curated these gene models with an instance of WebApollo (Lee et al. 2013), resulting in 11,275 unique gene models encoding 13,666 transcripts. The resulting total proteome of *M. sulcatula* comprised 2,644 putative transmembrane proteins and 838 possible secreted proteins, of which 75 may be effectors, according to an EffectorP 2.0 search (Sperschneider et al. 2018). The most abundant PFAM domains were transporter domains (MFS1 and AAA-type), kinases, alcohol dehydrogenase, and Zn-cluster transcription factors.

Altogether, we generated a high-quality genome assembly and a manually curated and functionally annotated transcriptome and proteome of *Myriosclerotinia sulcatula*, which is a relative of the notorious necrotrophic plant pathogens *B. cinerea* and *S. sclerotiorum* but exhibits a different lifestyle. This dataset will be highly valuable for future comparative studies toward understanding the evolution of lifestyle and host range in plant-pathogenic fungi.

Data Availability

The sequencing datasets produced for this study are deposited at the EBI European Nucleotide Archive (ENA) under the project reference PRJEB36718. The *Myriosclerotinia sulcatula* isolate MySu01 annotated nuclear genome assembly is deposited under the reference GCA_902810775. In addition, the fasta files for the genome, CDS and protein sequences and the GFF file are available at Zenodo. The codes used for this study are deposited at GitHub.

Acknowledgments

We thank O. Navaud for initial contributions to this project. We are grateful to the genotoul bioinformatics platform Toulouse Midi-Pyrenees (Bioinfo Genotoul) for providing help and computing and storage resources. This work was performed in collaboration with the GeT core facility, Toulouse, France.

Author-Recommended Internet Resources

Bioinfo Genotoul: http://bioinfo.genotoul.fr
CBS-KNAW microbial collection: http://www.wi.knaw.nl
EBI European Nucleotide Archive (ENA) https://www.ebi.ac.uk/ena/
GeT core facility: https://get.genotoul.fr
GitHub: https://github.com/stefankusch/genome_Myriosclerotinia_2020
NCBI BLAST: https://blast.ncbi.nlm.nih.gov/Blast.cgi
RNAweasel: https://megasun.bch.umontreal.ca/RNAweasel
Zenodo: https://zenodo.org/record/3726921#.XntotS-ZO1s

Literature Cited

Boland, G. J., and Hall, R. 1994. Index of plant hosts of *Sclerotinia sclerotiorum*. Can. J. Plant Pathol. 16:93-108.

Bolger, A. M., Lohse, M., and Usadel, B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114-2120.

Derbyshire, M., Denton-Giles, M., Hegedus, D., Seifbarghy, S., Rollins, J., van Kan, J., Seidl, M. F., Faino, L., Mbengue, M., Navaud, O., Raffaele, S., Hammond-Kosack, K., Heard, S., and Oliver, R. 2017. The complete genome sequence of the phytopathogenic fungus *Sclerotinia sclerotiorum* reveals insights into the genome architecture of broad host range pathogens. Genome Biol. Evol. 9: 593-618.

Eddy, S. R. 2011. Accelerated profile HMM searches. PLoS Comput. Biol. 7: e1002195.

Hoff, K. J., Lange, S., Lomsadze, A., Borodovsky, M., and Stanke, M. 2016. BRAKER1: Unsupervised RNA-Seq-based genome annotation with GeneMark-ET and AUGUSTUS. Bioinformatics 32:767-769.

Kim, D., Langmead, B., and Salzberg, S. L. 2015. HISAT: a fast spliced aligner with low memory requirements. Nat. Methods 12:357-360.

Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., and Phillippy, A. M. 2017. Canu: scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation. Genome Res. 27:722-736.

Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E. L. L. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J. Mol. Biol. 305:567-580.

Lee, E., Helt, G. A., Reese, J. T., Munoz-Torres, M. C., Childers, C. P., Buels, R. M., Stein, L., Holmes, I. H., Elsik, C. G., and Lewis, S. E. 2013. Web Apollo: a web-based genomic annotation editing platform. Genome Biol. 14:R93.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25: 2078-2079.

Marçais, G., and Kingsford, C. 2011. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics 27:764-770.

Navaud, O., Barbacci, A., Taylor, A., Clarkson, J. P., and Raffaele, S. 2018. Shifts in diversification rates and host jump frequencies shaped the diversity of host range among *Sclerotiniaceae* fungal plant pathogens. Mol. Ecol. 27:1309-1323.

Nielsen, H. 2017. Predicting secretory proteins with SignalP. Pages 59-73 in: Protein Function Prediction. Methods in Molecular Biology. D. Kihara, ed. Humana Press, New York.

Quinlan, A. R., and Hall, I. M. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26:841-842.

Schumacher, T., and Kohn, L. M. 1985. A monographic revision of the genus *Myriosclerotinia*. Can. J. Bot. 63:1610-1640.

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.

Smit, A. F. A., Hubley, R., and Green, P. 2016. Masker Open-4.0. 2013-2015.

Sperschneider, J., Dodds, P. N., Gardiner, D. M., Singh, K. B., and Taylor, J. M. 2018. Improved prediction of fungal effector proteins from secretomes with EffectorP 2.0. Mol. Plant Pathol. 19:2094-2110.

Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C. A., Zeng, Q., Wortman, J., Young, S. K., and Earl, A. M. 2014. Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One 9:e112963.