AMERICAN SOCIETY FOR MICROBIOLOGY | genomeAnnouncements

Genome Sequence and Annotation of *Colletotrichum higginsianum*, a Causal Agent of Crucifer Anthracnose Disease

Antonios Zampounis,[a] Sandrine Pigné,[a] Jean-Félix Dallery,[a] Alexander H. J. Wittenberg,[b] Shiguo Zhou,[c] David C. Schwartz,[c] Michael R. Thon,[d] Richard J. O'Connell[a]

UMR BIOGER, INRA, AgroParisTech, Université Paris-Saclay, Thiverval-Grignon, France[a]; KeyGene N.V., Wageningen, The Netherlands[b]; Laboratory for Molecular and Computational Genomics, Department of Chemistry, Laboratory of Genetics, University of Wisconsin–Madison, Madison, Wisconsin, USA[c]; Instituto Hispano-Luso de Investigaciones Agrarias (CIALE), Department of Microbiology and Genetics, University of Salamanca, Salamanca, Spain[d]

***Colletotrichum higginsianum* is an ascomycete fungus causing anthracnose disease on numerous cultivated plants in the family *Brassicaceae*, as well as the model plant *Arabidopsis thaliana*. We report an assembly of the nuclear genome and gene annotation of this pathogen, which was obtained using a combination of PacBio long-read sequencing and optical mapping.**

Received 17 June 2016 **Accepted** 23 June 2016 **Published** 18 August 2016

Citation Zampounis A, Pigné S, Dallery J-F, Wittenberg AHJ, Zhou S, Schwartz DC, Thon MR, O'Connell RJ. 2016. Genome sequence and annotation of *Colletotrichum higginsianum*, a causal agent of crucifer anthracnose disease. Genome Announc 4(4):e00821-16. doi:10.1128/genomeA.00821-16.

Address correspondence to Richard J. O'Connell, richard.oconnell@versailles.inra.fr.

Colletotrichum is a large genus of plant-pathogenic fungi that cause economically important anthracnose diseases on leaves and fruits of numerous monocot and dicot crops worldwide (1), although some species can grow as endophytes in symptomless plants (2). *C. higginsianum* attacks cultivated *Brassicaceae* and also *Arabidopsis thaliana*, providing a tractable model for analyzing fungal pathogenicity and plant responses (3). We previously sequenced the *C. higginsianum* genome using a combination of short reads from 454 FLX (350 bp) and Illumina (36 bp and 100 bp) platforms. This produced a highly fragmented assembly (GenBank accession number CACQ02000000) containing 10,269 contigs (N_{50} length = 6,150 bp) with a total length of 49.08 Mbp (4). In the resulting annotation of 16,172 protein-coding genes, many genes were truncated (~9%) or split between contigs (~7%), producing multiple gene calls. Here, we report a near-complete assembly of the nuclear genome of this fungus, which was obtained by combining long PacBio reads with the Optical Mapping System (5).

C. higginsianum strain IMI 349063 was originally isolated from leaves of *Brassica campestris* (3). High-molecular-weight genomic DNA was purified from mycelium using AXG100 columns (Macherey-Nagel) according to the manufacturer's instructions. A size-selected library (~20 kb) was prepared and sequenced using the PacBio RSII sequencer and 15 single-molecule real-time (SMRT) cells with the P5-C3 polymerase-chemistry combination and 240-min movie time. The filtered sequence reads (7.06 Gbp, ~132.13× average coverage) were assembled *de novo* using HGAP version 3.0 and SMRT analysis version 2.3.0 software with default settings. Assembled sequences were aligned to the previously reported chromosome optical maps (4) for manual ordering and orientation of the contigs. Overlapping contigs were merged using Minimus2 (6). The assembled nuclear genome comprises 25 contigs (N_{50} contig length = 5.20 Mbp, maximum contig length = 6.04 Mbp), with a total length of 50.72 Mbp and a 51.86% G+C content. The 12 largest contigs (the predicted number of chromosomes) represent 99.3% of the assembly, and 11 chromosomes are completely sequenced from telomere to telomere. An additional 13 contigs, containing the rDNA repeats, were too small for alignment to the optical map. Based on consensus-calling results, the accuracy of the assembly is high (≥99.9%). Using the MAKER2 pipeline (7) to annotate the nuclear genome, a total of 14,651 protein-coding gene models were predicted.

By combining the long reads obtained from SMRT sequencing with optical mapping, we obtained a near-complete assembly of the nuclear genome of *C. higginsianum*, allowing a more accurate gene annotation that will facilitate future studies on the infection biology of this important model pathogen. The large contiguous genomic regions will be especially valuable for studying structural rearrangements, large secondary metabolism gene clusters, and the chromosome distribution of repetitive elements, pathogenicity-related genes, and epigenetic marks.

Accession number(s). This whole-genome shotgun project has been deposited at DDBJ/ENA/GenBank under the accession number LTAN00000000. The version described in this paper is version LTAN01000000.

ACKNOWLEDGMENTS

We are grateful to Bruno Huettel (Max Planck Genome Centre, Cologne, Germany) for expert advice on the preparation of high molecular weight genomic DNA and Julie Vallet for excellent technical assistance.

FUNDING INFORMATION

This work, including the efforts of Richard O'Connell, was funded by Agence Nationale de la Recherche (ANR) (ANR-12-CHEX-0008-01).

REFERENCES

1. **Crouch J, O'Connell R, Gan P, Buiate E, Torres M, Beirn L, Shirasu K, Vaillancourt L.** 2014. The genomics of *Colletotrichum*, p. 69–102. In Dean RA, Lichens-Park A, Kole C (ed), Genomics of plant-associated fungi: monocot pathogens. Springer, Berlin. http://dx.doi.org/10.1007/978-3-662-44053-7_3.

2. **Hacquard S, Kracher B, Hiruma K, Münch PC, Garrido-Oter R, Thon MR, Weimann A, Damm U, Dallery J-F, Hainaut M, Henrissat B, Lespinet O, Sacristán S, van Themaat van Themaat E, Kemen E, McHardy AC, Schulze-Lefert P, O'Connell RJ.** 2016. Survival trade-offs in plant roots during colonization by closely related beneficial and pathogenic fungi. Nat Commun **7:**11362. http://dx.doi.org/10.1038/ncomms11362.
3. **O'Connell R, Herbert C, Sreenivasaprasad S, Khatib M, Esquerré-Tugayé M-T, Dumas B.** 2004. A novel *Arabidopsis-Colletotrichum* pathosystem for the molecular dissection of plant-fungal interactions. Mol Plant Microbe Interact **17:**272–282. http://dx.doi.org/10.1094/MPMI.2004.17.3.272.
4. **O'Connell RJ, Thon MR, Hacquard S, Amyotte SG, Kleemann J, Torres MF, Damm U, Buiate EA, Epstein L, Alkan N, Altmüller J, Alvarado-Balderrama L, Bauser CA, Becker C, Birren BW, Chen Z, Choi J, Crouch JA, Duvick JP, Farman MA, Gan P, Heiman D, Henrissat B, Howard RJ, Kabbage M, Koch C, Kracher B, Kubo Y, Law AD, Lebrun M-H, Lee Y-H, Miyara I, Moore N, Neumann U, Nordstroem K, Panaccione DG, Panstruga R, Place M, Proctor RH, Prusky D, Rech G, Reinhardt R, Rollins JA, Rounsley S, Schardl CL, Schwartz DC, Shenoy N, Shirasu K, Sikhakolli UR, Stueber K, Sukno SA, Sweigard JA, Takano Y, Takahara H, Trail F, van der Does HC, Voll LM, Will I, Young S, Zeng Q, Zhang J, Zhou S, Dickman MB, Schulze-Lefert P, van Themaat EL, Ma L-J, Vaillancourt LJ.** 2012. Lifestyle transitions in plant pathogenic *Colletotrichum* fungi deciphered by genome and transcriptome analyses. Nat Genet **44:**1060–1065. http://dx.doi.org/10.1038/ng.2372.
5. **Dimalanta ET, Lim A, Runnheim R, Lamers C, Churas C, Forrest DK, de Pablo JJ, Graham MD, Coppersmith SN, Schwartz DC.** 2004. A microfluidic system for large DNA molecule arrays. Anal Chem **76:**5293–5301. http://dx.doi.org/10.1021/ac0496401.
6. **Sommer DD, Delcher AL, Salzberg SL, Pop M.** 2007. Minimus: a fast, lightweight genome assembler. BMC Bioinformatics **8:**64. http://dx.doi.org/10.1186/1471-2105-8-64.
7. **Holt C, Yandell M.** 2011. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. BMC Bioinformatics **12:**491. http://dx.doi.org/10.1186/1471-2105-12-491.