Draft Genome Sequence of *Colletotrichum sublineola*, a Destructive Pathogen of Cultivated Sorghum

Riccardo Baroncelli, José María Sanz-Martín, Gabriel E. Rech, Serenella A. Sukno, Michael R. Thon

Instituto Hispano-Luso de Investigaciones Agrarias (CIALE), Departamento de Microbiología y Genética, Universidad de Salamanca, Salamanca, Spain

Colletotrichum sublineola is a filamentous fungus that causes anthracnose disease on sorghum. We report a draft whole-genome shotgun sequence and gene annotation of the nuclear genome of this fungus using Illumina sequencing.

Address correspondence to Michael R. Thon, mthon@usal.es.

Sorghum (*Sorghum bicolor*) is an important cultivated crop and is used as a staple food, as animal fodder, and as a biofuels feedstock. One of the most important diseases that affects sorghum crops is sorghum anthracnose, caused by the ascomycete fungus *Colletotrichum sublineola*, which can cause a reduction in grain size as well as yield losses of more than 50% under epidemic conditions (1). *Colletotrichum sublineola* is also closely related to *C. graminicola* (2), a causal agent of maize anthracnose, and is therefore useful for comparative genomics of the two species to uncover the evolutionary mechanisms of speciation, host specificity, and pathogenicity.

Colletotrichum sublineola strain TX430BB was isolated from sorghum in College Station, TX, USA (3). Total genomic DNA was purified using the method of Baek and Kenerley (4) and sequenced using 100-bp paired-end reads on an Illumina HiSeq 2000, and the sequence reads (4.32 Gbp; average coverage, 91.93×) were assembled using Velvet version 1.2.07 (5). The draft genome of *C. sublineola* consists of 1,625 sequence scaffolds with a total length of 46.75 Mbp ($N_{50}=70,717$ bp, and $N_{90}=13,454$ bp), 52.70% G+C content, and a maximum scaffold size of 423,147 bp. The mitochondrial genome was identified by performing BLAST searches of the contigs in a database of mitochondrial genomes of other fungi, resulting in the identification and removal of 6 contigs. The completeness of the assembly was assessed using CEGMA v2.4 (6), which estimated the genome sequence to be 98.39% complete. The nuclear genome was annotated using the MAKER pipeline (7). Overall, 12,699 protein-coding gene models were predicted in the nuclear genome.

Using WoLF PSORT (8) we identified 1,820 proteins that are predicted to be extracellular (14.33% of the proteome). Based on BLAST searches (*e*-value threshold of $1e-3$) of the extracellular proteins, 168 (9.23% of the secretome) do not have any sequence similarity to proteins in *C. graminicola*, and of those 168 proteins, 70 (3.85% of the secretome) do not have any sequence similarity to proteins in other *Colletotrichum* species (2, 9–11). However, only 60 secreted proteins are unique to *C. sublineola* when compared to the nr database with BLAST. These species-specific extracellular proteins may be effectors, proteins that have important roles in modulating the plant's immune system and in host specialization.

In this study we present a draft genome sequence from a member of *C. sublineola*, a destructive pathogen of cultivated sorghum. The sequence represents a new resource that will be useful for further research into the biology, ecology, and evolution of this key pathogen.

Nucleotide sequence accession numbers. This whole-genome shotgun sequencing project has been deposited at GenBank under the accession no. JMSE00000000. The version described in this paper is JMSE00000000.1.

REFERENCES

1. Li L, Zhu F, Liu H, Chu A, Lo C. 2013. Isolation and expression analysis of defense-related genes in sorghum–*Colletotrichum sublineolum* interaction. Physiol. Mol. Plant Pathol. **84:**123–130. http://dx.doi.org/10.1016/j.pmpp.2013.08.005.
2. 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, Nordström 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, Stüber 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, Ver Loren van Themaat E, 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.
3. Cardwell KF, Hepperly PR, Frederiksen RA. 1989. Pathotypes of *Colletotrichum graminicola* and seed transmission of sorghum anthracnose. Plant Dis. **73:**255–257. http://dx.doi.org/10.1094/PD-73-0255.
4. Baek J-M, Kenerley CM. 1998. The *arg2* gene of *Trichoderma virens*: Cloning and development of a homologous transformation system. Fungal Genet. Biol. **23:**34–44. http://dx.doi.org/10.1006/fgbi.1997.1025.
5. Zerbino DR, Birney E. 2008. Velvet: algorithms for de novo short read

assembly using de Bruijn graphs. Genome Res. **18:**821–829. http://dx.doi.org/10.1101/gr.074492.107.

6. **Parra G, Bradnam K, Korf I.** 2007. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. Bioinformatics **23:** 1061–1067. http://dx.doi.org/10.1093/bioinformatics/btm071.

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.

8. **Horton P, Park K-J, Obayashi T, Fujita N, Harada H, Adams-Collier CJ, Nakai K.** 2007. WoLF PSORT: protein localization predictor. Nucleic Acids Res. **35:**W585–W587. http://dx.doi.org/10.1093/nar/gkm259.

9. **Baroncelli R, Sreenivasaprasad S, Sukno SA, Thon MR, Holub E.** 2014. Draft genome sequence of *Colletotrichum acutatum sensu lato* (*Colletotrichum fioriniae*). Genome Announc. **2**(2)**:**e00112-14. http://dx.doi.org/10.1128/genomeA.00112-14.

10. **Alkan N, Meng X, Friedlander G, Reuveni E, Sukno S, Sherman A, Thon M, Fluhr R, Prusky D.** 2013. Global aspects of pacC regulation of pathogenicity genes in *Colletotrichum gloeosporioides* as revealed by transcriptome analysis. Mol. Plant Microbe Interact. **26:**1345–1358. http://dx.doi.org/10.1094/MPMI-03-13-0080-R.

11. **Gan P, Ikeda K, Irieda H, Narusaka M, O'Connell RJ, Narusaka Y, Takano Y, Kubo Y, Shirasu K.** 2013. Comparative genomic and transcriptomic analyses reveal the hemibiotrophic stage shift of *Colletotrichum* fungi. New Phytol. **197:**1236–1249. http://dx.doi.org/10.1111/nph.12085.