Genome Sequence of *Fusarium graminearum* Strain CML3066, Isolated from a Wheat Spike in Southern Brazil

Ana K. Machado Wood,[a] Robert King,[a] Martin Urban,[a] Camila P. Nicolli,[b] Emerson M. Del Ponte,[b] Kim E. Hammond-Kosack[a]

[a]Rothamsted Research, Harpenden, Hertfordshire, United Kingdom
[b]Universidade Federal de Viçosa, Minas Gerais, Brazil

ABSTRACT *Fusarium graminearum* is a global fungal pathogen of wheat and other small grains, causing *Fusarium* head blight (FHB) disease, also known as wheat scab. We report here the annotated genome of a deoxynivalenol/15-acetyl-deoxynivalenol-producing Brazilian strain called CML3066, isolated from FHB-symptomatic wheat spikes collected in 2009.

The ascomycete fungus *Fusarium graminearum* is the main pathogen causing *Fusarium* head blight (FHB), an important cereal disease worldwide (1). The *F. graminearum* genome from a South American strain was not previously available. Here, we report an annotated assembly for strain CML3066 (DON/15-ADON), isolated in 2009 in Rio Grande do Sul state, Brazil (latitude, −28.327, longitude, −51.271). This strain was isolated from a symptomatic wheat spike with 21.5% FHB incidence (2).

Genomic DNA of *F. graminearum* CML3066 was extracted from mycelia grown for 3 days in potato dextrose broth (PDB) medium using a cetyltrimethylammonium bromide (CTAB) protocol (3) and quantified using a Qubit 2.0 fluorometer (Life Technologies). Libraries were prepared with TruSeq DNA high-throughput (HT) (Illumina) and SMRTbell (PacBio) kits. Genome sequencing was done both on an Illumina HiSeq 2000 platform, producing 100-bp paired-end reads (91× coverage) with no quality control required, and on a PacBio RS II platform with a postquality filter of minimum polymerase read quality of 0.80 and minimum subread length of 500 bp, resulting in 160× coverage with a subread total of 6,392,721,477 bp, 1,358,615 reads, and an N_{50} value of 5,621 bp. Default parameters were used for all software unless otherwise noted. The *de novo* assembly was carried out using SOAPdenovo2 v2.0.4 using the Illumina data with a range of k-mer values (61 to 99) and the SMRT analysis portal using the PacBio data. The PacBio assembly was manually gap filled and further scaffolded using Lastz v1.04.03 alignments with the complementary Illumina assembly, resulting in four complete chromosomes from telomere to telomere with no gaps or N bases. Reference sequence statistics were extracted from Geneious v8.1. The genome annotation of CML3066 was done using the MAKER v2.30 (4) annotation pipeline with RepeatMasker v4.50 (5). Gene calls were generated using both AUGUSTUS v2.7 (6) using the *F. graminearum* species model and GeneMark (7), which was trained using strain PH-1 (8, 9).

The CML3066 assembly is 36,908,675 bp long with a GC content of 47.9%. The CML3066 genome is predicted to contain 14,188 genes, 286 of which are not present in the PH-1 genome. Using the PH-1 reference, a minimum of 80% of the length of the gene was required to have mapped reads to be considered present. Single nucleotide polymorphism (SNP) calling was performed with SAMtools using default settings. SNP effects were predicted using SnpEff 4.2. Comparison of SNP frequencies along all four chromosomes of both CML3066 and PH-1 (8, 10) revealed that all telomere proximal regions displayed the

TABLE 1 Genome sequence assemblies for *F. graminearum* strains PH-1 and CML3066

Characteristic	Value for strain:	
	PH-1[a]	CML3066
Genome size (bp)[b]	36,663,736	36,908,675
No. of chromosomes	4	4
GC content (%)[c]	48.2	47.9
No. of spanned gaps	0	0
No. of predicted genes	14,145	14,188

[a] Reannotated genome (10).
[b] Including all scaffolds and the mitochondrial genome but excluding the large repetitive sequence at the carboxyl end of chromosome 4.
[c] Excluding the mitochondrial genome.

highest SNP density windows. In addition, chromosomes 1, 2, and 4 were found to have one or two large interstitial regions with a high SNP density. To predict secreted proteins, Blast2GO v3.2 was used to identify signal peptide and transmembrane domains. Prediction of glycosylphosphatidylinositol (GPI)-anchored membrane proteins, cellular protein localization, and effectors was performed using Big-Pi, WoLF PSort and ProtComp v9.0 (Softberry), and EffectorP v1.0 (11–14), respectively. The secretome was predicted to contain 874 genes. A genome comparison between CML3066 and the reference strain PH-1 is summarized in Table 1.

Data availability. The raw data and assembled/annotated sequences have been deposited in the European Nucleotide Archive (ENA). The study accession number is PRJEB12819. The accession numbers for the assembled chromosomes and mitochondrial genome are LT222053 to LT222057. The secretome and effector predictions can be found at https://github.com/ana321wood/Secretome_CML3066_Feb2020/blob/master/Secretome_CML3066_Feb2020_AMW.txt.

ACKNOWLEDGMENTS

REFERENCES

1. Dean R, Van Kan JAL, Pretorius ZA, Hammond-Kosack KE, Di Pietro A, Spanu PD, Rudd JJ, Dickman M, Kahmann R, Ellis J, Foster GD. 2012. The top 10 fungal pathogens in molecular plant pathology. Mol Plant Pathol 13:414–430. https://doi.org/10.1111/j.1364-3703.2011.00783.x.

2. Del Ponte EM, Spolti P, Ward TJ, Gomes LB, Nicolli CP, Kuhnem PR, Silva CN, Tessmann DJ. 2015. Regional and field-specific factors affect the composition of Fusarium head blight pathogens in subtropical no-till wheat agroecosystem of Brazil. Phytopathology 105:246–254. https://doi.org/10.1094/PHYTO-04-14-0102-R.

3. Brandfass C, Karlovsky P. 2008. Upscaled CTAB-based DNA extraction and real-time PCR assays for Fusarium culmorum and F. graminearum DNA in plant material with reduced sampling error. Int J Mol Sci 9:2306–2321. https://doi.org/10.3390/ijms9112306.

4. Cantarel BL, Korf I, Robb SMC, Parra G, Ross E, Moore B, Holt C, Alvarado AS, Yandell M. 2008. MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. Genome Res 18:188–196. https://doi.org/10.1101/gr.6743907.

5. Bedell JA, Korf I, Gish W. 2000. MaskerAid: a performance enhancement to RepeatMasker. Bioinformatics 16:1040–1041. https://doi.org/10.1093/bioinformatics/16.11.1040.

6. Stanke M, Steinkamp R, Waack S, Morgenstern B. 2004. AUGUSTUS: a Web server for gene finding in eukaryotes. Nucleic Acids Res 32:W309–W312. https://doi.org/10.1093/nar/gkh379.

7. Besemer J, Borodovsky M. 2005. GeneMark: Web software for gene finding in prokaryotes, eukaryotes and viruses. Nucleic Acids Res 33:W451–W454. https://doi.org/10.1093/nar/gki487.

8. Cuomo CA, Güldener U, Xu J-R, Trail F, Turgeon BG, Di Pietro A, Walton JD, Ma L-J, Baker SE, Rep M, Adam G, Antoniw J, Baldwin T, Calvo S, Chang Y-L, Decaprio D, Gale LR, Gnerre S, Goswami RS, Hammond-Kosack K, Harris LJ, Hilburn K, Kennell JC, Kroken S, Magnuson JK, Mannhaupt G, Mauceli E, Mewes H-W, Mitterbauer R, Muehlbauer G, Münsterkötter M, Nelson D, O'Donnell K, Ouellet T, Qi W, Quesneville H, Roncero MIG, Seong K-Y, Tetko IV, Urban M, Waalwijk C, Ward TJ, Yao J, Birren BW, Kistler HC. 2007. The Fusarium graminearum genome reveals a link between localized polymorphism and pathogen specialization. Science 317:1400–1402. https://doi.org/10.1126/science.1143708.

9. King R, Urban M, Lauder RP, Hawkins N, Evans M, Plummer A, Halsey K, Lovegrove A, Hammond-Kosack K, Rudd JJ. 2017. A conserved fungal glycosyltransferase facilitates pathogenesis of plants by enabling hyphal growth on solid surfaces. PLoS Pathog 13:e1006672. https://doi.org/10.1371/journal.ppat.1006672.

10. King R, Urban M, Hammond-Kosack MCU, Hassani-Pak K, Hammond-

Kosack KE. 2015. The completed genome sequence of the pathogenic ascomycete fungus *Fusarium graminearum*. BMC Genomics 16:544. https://doi.org/10.1186/s12864-015-1756-1.

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

12. Eisenhaber B, Schneider G, Wildpaner M, Eisenhaber F. 2004. A sensitive predictor for potential GPI lipid modification sites in fungal protein sequences and its application to genome-wide studies for *Aspergillus nidulans*, *Candida albicans*, *Neurospora crassa*, *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*. J Mol Biol 337:243–253. https://doi.org/10.1016/j.jmb.2004.01.025.

13. Sperschneider J, Gardiner DM, Dodds PN, Tini F, Covarelli L, Singh KB, Manners JM, Taylor JM. 2016. EffectorP: predicting fungal effector proteins from secretomes using machine learning. New Phytol 210:743–761. https://doi.org/10.1111/nph.13794.

14. Brown NA, Antoniw J, Hammond-Kosack KE. 2012. The predicted secretome of the plant pathogenic fungus *Fusarium graminearum*: a refined comparative analysis. PLoS One 7:e33731. https://doi.org/10.1371/journal.pone.0033731.