



First *De Novo* Draft Genome Sequence of the Pathogenic Fungus *Fusarium udum* F02845, Associated with Pigeonpea (*Cajanus cajan* L. Millspaugh) Wilt

Alok Kumar Srivastava,^a Prem Lal Kashyap,^{a,b} Hillol Chakdar,^a Murugan Kumar,^a Anchal Kumar Srivastava,^a Jagriti Yadav,^a Hena Jamali,^a Ruchi Srivastava,^a Anjney Sharma,^a Praveen Tiwari,^a Alok Singh,^a Anil Kumar Saxena^a

^aICAR-National Bureau of Agriculturally Important Microorganisms (NBAIM), Mau, Uttar Pradesh, India

^bICAR-Indian Institute of Wheat and Barley Research (IIWBR), Karnal, Haryana, India

ABSTRACT *Fusarium udum* F02845 is a destructive fungal pathogen which causes pigeonpea (*Cajanus cajan* L. Millspaugh) wilt. Here we report the first *de novo* draft assembly of *Fusarium udum* F02845, isolated from an infected pigeonpea stem. The genome was determined to be 56.38 Mb in size, with a G+C content of 42.44%, and predicted to have 712 scaffolds with a total number of 11,829 genes.

Wilt caused by *Fusarium udum* F02845 is one of the most important diseases of pigeonpea. The disease has been reported from several countries, including India, Bangladesh, Mauritius, Ghana, Kenya, Malawi, Tanzania, Uganda, Indonesia, Thailand, and Trinidad (1). The fungus causing wilt can survive on infected plant debris in soil for about 2 to 3 years, and it is responsible for causing 16 to 47% yield loss under favorable environmental conditions (2). At present, chemical and biological disease management strategies for containing this fungus are not very effective; therefore, genome sequencing can divulge virulence-related genes for better understanding of host-pathogen interactions. So far, molecular investigations on the pathogenicity of *F. udum* have not been performed. Here, we describe the first draft genome sequence to assist further genome-based examination of *F. udum* and its host-pathogen interactions.

The isolate of *F. udum* (F02845) was collected from a pigeonpea stem displaying pronounced symptoms of wilt disease in the year 2010 from Bahraich (27°34'31.4" N, 81°35'33.6" E), Uttar Pradesh, India. The fungal isolate was cultivated on potato dextrose agar (PDA), and total genomic DNA was extracted with cetyl-trimethylammonium bromide (CTAB), as described by Kumar et al. (3).

The draft genome was sequenced with the Illumina NextSeq sequencer, using a HiSeq 2000 platform for paired-end reads. A TruSeq Nano DNA library kit (Illumina) was used for sequencing-library preparation. A library of 28.08 million paired-end reads (read length, 101 bp; insert size, 433 bp) of 28.36 Gb total size was generated. The Next-Generation Sequencing Quality Control (NGS QC) toolkit version 2.3 (4) was used to filter high-quality data (at a Phred score of 20), and 24.99 million high-quality reads were obtained. Primary genome assembly was done using the program Velvet version 1.2.10 (5) with a kmer length of 81. The assembled genome was 56,750,279 bp in length, with an N_{50} value of ~0.08 Mb, resulting in 10,427 contigs. The scaffolding of primarily assembled data was done using SSPACE version 3.0 (6), resulting in 2,634 scaffolds at a maximum link ratio of ≥ 0.5 . The maximum and average scaffold lengths were ~0.7 Mb and ~0.21 Mb, respectively. After removing scaffolds that were less than 200 bp by using CONTIGuator 2.7 (7), the final assembly consisted of 712 scaffolds with a genome size of 56,381,318 bp (42.44% G+C content). Interspersed repetitive elements and low-complexity DNA sequences were masked using RepeatMasker version

Received 16 July 2018 **Accepted** 8

September 2018 **Published** 4 October 2018

Citation Srivastava AK, Kashyap PL, Chakdar H, Kumar M, Srivastava AK, Yadav J, Jamali H, Srivastava R, Sharma A, Tiwari P, Singh A, Saxena AK. 2018. First *de novo* draft genome sequence of the pathogenic fungus *Fusarium udum* F02845, associated with pigeonpea (*Cajanus cajan* L. Millspaugh) wilt. *Microbiol Resour Announc* 7:e01001-18. <https://doi.org/10.1128/MRA.01001-18>.

Editor Antonis Rokas, Vanderbilt University

Copyright © 2018 Srivastava et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Alok Kumar Srivastava, aloksrivastva@gmail.com, or Prem Lal Kashyap, plkashyap@gmail.com.

4 (8), followed by rRNA and tRNA prediction using RNAmmer v1.2 (9) and tRNAscan-SE v1.3.1 (10), respectively. Prediction of protein-coding genes was performed using GeneMark-ES fungal version 2 (11). Functional classification of the predicted proteins was done using the EuKaryotic Orthologous Groups of proteins database (KOG) (12), while motifs and protein domains were predicted with InterProScan v5 (13). The dbCAN database was used to predict carbohydrate-active enzymes (14). Overall, the whole genome encompasses 11,829 protein-coding genes, 296 tRNAs, and 53 rRNAs. A total of 8,928 genes were categorized into functional groups using the KOG database (11). Furthermore, the genome contained 1,439 signal peptides, 15,649 transmembrane helices, 2,858 carbohydrate-active enzymes (CAZy), 3,682 transporter genes, and 1,060 putative pathogenicity genes.

Data availability. The present draft genome assembly has been deposited in the NCBI repository under GenBank accession number [NIFK00000000](https://ncbi.nlm.nih.gov/GenBank/entry/1111111111) and assembly accession number [GCA_002194535](https://ncbi.nlm.nih.gov/BioProject/entry/1111111111) (BioProject number [PRJNA385264](https://ncbi.nlm.nih.gov/BioProject/entry/1111111111)). Short-read data have been submitted to the SRA under NCBI accession number [SRP157084](https://ncbi.nlm.nih.gov/SRA/entry/1111111111).

ACKNOWLEDGMENTS

This work was supported by grants from the Indian Council of Agricultural Research (ICAR) under the Consortium Research Platform (CRP) on genomics.

We thank the staff at Bionivid Technology Private Limited, Bangalore, India, for their advice and excellent genome-sequencing technical assistance.

REFERENCES

- Singh D, Sinha B, Rai VP, Singh MN, Singh DK, Kumar R, Singh AK. 2016. Genetics of *Fusarium* wilt resistance in pigeonpea (*Cajanus cajan*) and efficacy of associated SSR markers. *Plant Pathol J* 32:95–101. <https://doi.org/10.5423/PPJ.OA.09.2015.0182>.
- Kashyap PL, Rai S, Kumar S, Srivastava AK, Anandaraj M, Sharma AK. 2015. Mating type genes and genetic markers to decipher intraspecific variability among *Fusarium udum* isolates from pigeonpea. *J Basic Microbiol* 55:846–856. <https://doi.org/10.1002/jobm.201400483>.
- Kumar S, Rai S, Maurya DK, Kashyap PL, Srivastava AK, Anandaraj M. 2013. Cross-species transferability of microsatellite markers from *Fusarium oxysporum* for the assessment of genetic diversity in *Fusarium udum*. *Phytoparasitica* 41:615–622. <https://doi.org/10.1007/s12600-013-0324-y>.
- Patel RK, Jain M. 2012. NGS QC toolkit: a toolkit for quality control of next generation sequencing data. *PLoS One* 7:e30619. <https://doi.org/10.1371/journal.pone.0030619>.
- Zerbino DR, Birney E. 2008. Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Res* 18:821–829. <https://doi.org/10.1101/gr.074492.107>.
- Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. 2011. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* 27:578–579. <https://doi.org/10.1093/bioinformatics/btq683>.
- Galardini M, Biondi EG, Bazzicalupo M, Mengoni A. 2011. CONTIGuator: a bacterial genomes finishing tool for structural insights on draft genomes. *Source Code Biol Med* 6:11. <https://doi.org/10.1186/1751-0473-6-11>.
- Smit AFA, Hubley R, Green P. 1996. RepeatMasker Open-3.0. <http://www.repeatmasker.org>.
- Lagesen K, Hallin P, Rødland EA, Staerfeldt H-H, Rognes T, Ussery DW. 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res* 35:3100–3108. <https://doi.org/10.1093/nar/gkm160>.
- Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* 25:955–964.
- Ter-Hovhannisyan V, Lomsadze A, Chernoff Y, Borodovsky M. 2008. Gene prediction in novel fungal genomes using an *ab initio* algorithm with unsupervised training. *Genome Res* 18:1979–1990. <https://doi.org/10.1101/gr.081612.108>.
- Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, Krylov DM, Mazumder R, Mekhedov SL, Nikolskaya AN, Rao BS, Smirnov S, Sverdlov AV, Vasudevan S, Wolf YI, Yin JJ, Natale DA. 2003. The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* 4:41. <https://doi.org/10.1186/1471-2105-4-41>.
- Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong S-Y, Lopez R, Hunter S. 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30:1236–1240. <https://doi.org/10.1093/bioinformatics/btu031>.
- Yin Y, Mao X, Yang JC, Chen X, Mao F, Xu Y. 2012. dbCAN: a Web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res* 40:W445–W451. <https://doi.org/10.1093/nar/gks479>.