



Improved Assembly of Reference Genome *Fusarium oxysporum* f. sp. *lycopersici* Strain Fol4287

Dilay Hazal Ayhan,^a Cristina López-Díaz,^b Antonio Di Pietro,^b  Li-Jun Ma^a

^aDepartment of Biochemistry and Molecular Biology, University of Massachusetts Amherst, Amherst, Massachusetts, USA

^bDepartment of Genetics, University of Córdoba, Córdoba, Spain

ABSTRACT *Fusarium oxysporum* is a pathogenic fungus that infects hundreds of plant species. This paper reports the improved genome assembly of a reference strain, *F. oxysporum* f. sp. *lycopersici* Fol4287, a tomato pathogen.

Fusarium oxysporum is a filamentous fungus that can infect hundreds of plant species, as well as immunocompromised human patients (1). The reference genome of *F. oxysporum* was first generated using Sanger sequencing with 6× coverage using a tomato-infecting strain, *F. oxysporum* f. sp. *lycopersici* Fol4287 (race 2, VCG 0030) (2). The strain was originally isolated from an infected tomato plant in Murcia by Javier Tello from the University of Almería, Spain (2, 3). It is available from the Fungal Genetics Stock Center (FGSC 9935), NCAUR/USDA (NRRL 34936), and CBS-KNAW (CBS 123668) collections. A comparative study of this reference genome with those of closely related species identified lineage-specific (LS) chromosomes that are rich in transposons and genes related to pathogenicity. The transfer of these LS chromosomes between strains of *F. oxysporum* was experimentally confirmed to convert a nonpathogenic strain into a pathogen (2, 4–6). These studies enabled the structural and functional partitioning of the *F. oxysporum* genome, which provides a novel means of dissecting fungal pathogenesis.

Unfortunately, this reference genome has a high level of single-nucleotide-level sequencing errors due to the low sequencing coverage. To improve its quality, we regenerated the whole-genome assembly with increased sequence coverage and combining Illumina and PacBio sequence technologies. Genomic DNA was extracted from the mycelium of Fol4287 (the same isolate that was sequenced before). The DNA library for short reads was prepared with an average 400-bp insert size. The DNA library was sequenced at 66× coverage, using the Illumina HiSeq 2500 platform, into 71-bp paired-end reads. The genomic DNA was sequenced using the PacBio RS II system with 10× subread coverage. FastQC (version 0.11.5) was used to check the quality of all reads. The average base quality of Illumina reads is 36.8, with a 71-bp read length. The PacBio reads have average and maximum read lengths of 6.8 kb and 52 kb, respectively.

The initial assembly was generated via SPAdes version 3.9.1 (7), combining raw Illumina and PacBio reads with default parameters. Quiver in SMRT Analysis (version 2.2.0) (8) was used to polish the assembly based on the PacBio reads. Further polishing was performed by mapping the Illumina reads to the assembly using BWA version 0.7.12 (9). FreeBayes v0.9.10-3-g47a713e (10) was used to identify base variants between the reads and the assembly. Highly confident variant sites were used to correct the assembly using a custom script (available at github.com/d-ayhan/tools). We also used structural variant (SV) callers, GRIDSS version 1.4.1 (11) and Sniffles version 1.0.8 (12), to identify the SVs in the initial assembly. All identified SVs were inspected manually to ensure accuracy. High-confidence merges/splits were integrated into the

Received 11 July 2018 Accepted 14 August 2018 Published 13 September 2018

Citation Ayhan DH, López-Díaz C, Di Pietro A, Ma L-J. 2018. Improved assembly of reference genome *Fusarium oxysporum* f. sp. *lycopersici* strain Fol4287. Microbiol Resour Announc 7:e00910-18. <https://doi.org/10.1128/MRA.00910-18>.

Editor Antonis Rokas, Vanderbilt University

Copyright © 2018 Ayhan et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license.

Address correspondence to Li-Jun Ma, lijun@biochem.umass.edu.

assembly. The improved assembly was then quality checked by remapping. This process was repeated until no future correction could be identified.

The final assembly is 53.9 Mb, with 499 contigs and an N_{50} value of 1.3 Mb. The largest contig size is 5.7 Mb. The GC content is 47.7%. The assembly includes a contig of 52,424 bp that captures the complete mitochondrial DNA and a contig of 7,875 bp of the complete ribosomal DNA sequence (2). In a comparison of the new assembly to the reference assembly of *F. oxysporum* (assembly [GCA_000149955](https://doi.org/10.1111/1462-2920.13445)) by BLAST, the contigs that belong to each chromosome were identified, ordered, and oriented within each chromosome. The contigs were divided into three categories, including 11 core chromosomes (C), 4 LS chromosomes (S), and some unmapped contigs (U). Except for chromosomes 1 and 2, each core chromosome was assembled into a single contig GenBank assembly number ([GCA_003315725](https://doi.org/10.1111/1462-2920.13445)).

Data availability. This whole-genome shotgun project has been deposited at DDBJ/ENA/GenBank under the accession number [QESU00000000](https://doi.org/10.1111/1462-2920.13445). The version described in this paper is version QESU01000000. The PacBio and Illumina reads are available in SRA under accession numbers [SRR7015920](https://doi.org/10.1111/1462-2920.13445) and [SRR7690004](https://doi.org/10.1111/1462-2920.13445), respectively.

ACKNOWLEDGMENTS

We thank the MGHPCC for providing high-performance computing capacity for the genome assembly process.

D.H.A. and L.-J.M. are supported by the Burroughs Wellcome Foundation (grant 1014893). C.L.-D. is supported by the Spanish MINECO (grant BES-2014-070450). This work was supported by the United States Department of Agriculture, National Institute of Food and Agriculture (grants 2011-35600-30379 and MASR-2009-04374) and the National Science Foundation (IOS-1652641), and MINECO (grant BIO2016-78923-R).

REFERENCES

- Ma L-J, Geiser DM, Proctor RH, Rooney AP, O'Donnell K, Trail F, Gardiner DM, Manners JM, Kazan K. 2013. *Fusarium* pathogenomics. *Annu Rev Microbiol* 67:399–416. <https://doi.org/10.1146/annurev-micro-092412-155650>.
- Ma L-J, van der Does HC, Borkovich KA, Coleman JJ, Daboussi M-J, Di Pietro A, Dufresne M, Freitag M, Grabherr M, Henrissat B, Houterman PM, Kang S, Shim W-B, Woloshuk C, Xie X, Xu J-R, Antoniw J, Baker SE, Bluhm BH, Breakspear A, Brown DW, Butchko RAE, Chapman S, Coulson R, Coutinho PM, Danchin EGJ, Diener A, Gale LR, Gardiner DM, Goff S, Hammond-Kosack KE, Hilburn K, Hua-Van A, Jonkers W, Kazan K, Kodira CD, Koehrsen M, Kumar L, Lee Y-H, Li L, Manners JM, Miranda-Saavedra D, Mukherjee M, Park G, Park J, Park S-Y, Proctor RH, Regev A, Ruiz-Roldan MC, Sain D, et al. 2010. Comparative genomics reveals mobile pathogenicity chromosomes in *Fusarium*. *Nature* 464:367–373. <https://doi.org/10.1038/nature08850>.
- Di Pietro A, Roncero MI. 1998. Cloning, expression, and role in pathogenicity of *pg1* encoding the major extracellular endopolygalacturonase of the vascular wilt pathogen *Fusarium oxysporum*. *Mol Plant Microbe Interact* 11:91–98. <https://doi.org/10.1094/MPMI.1998.11.2.91>.
- Kistler HC, Rep M, Ma L-J. 2013. Structural dynamics of *Fusarium* genomes, p 31–41. In Rown DW, Proctor RH (ed), *Fusarium: genomics, molecular and cellular biology*. Horizon Scientific Press, Norwich, United Kingdom.
- Ma L-J. 2014. Horizontal chromosome transfer and rational strategies to manage *Fusarium* vascular wilt diseases. *Mol Plant Pathol* 15:763–766. <https://doi.org/10.1111/mpp.12171>.
- van Dam P, Fokkens L, Schmidt SM, Linmans JHJ, Kistler HC, Ma L-J, Rep M. 2016. Effector profiles distinguish *formae speciales* of *Fusarium oxysporum*. *Environ Microbiol* 18:4087–4102. <https://doi.org/10.1111/1462-2920.13445>.
- Antipov D, Korobeynikov A, McLean JS, Pevzner PA. 2016. hybridSPAdes: an algorithm for hybrid assembly of short and long reads. *Bioinformatics* 32:1009–1015. <https://doi.org/10.1093/bioinformatics/btv688>.
- Chin C-S, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE, Turner SW, Korlach J. 2013. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat Methods* 10:563–569. <https://doi.org/10.1038/nmeth.2474>.
- Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>.
- Garrison E, Marth G. 2012. Haplotype-based variant detection from short-read sequencing. *arXiv arXiv:1207.3907v2 [q-bio.GN]*. <https://arxiv.org/abs/1207.3907>.
- Cameron DL, Schröder J, Penington JS, Do H, Molania R, Dobrovic A, Speed TP, Papenfuss AT. 2017. GRIDSS: sensitive and specific genomic rearrangement detection using positional de Bruijn graph assembly. *Genome Res* 27:2050–2060. <https://doi.org/10.1101/gr.222109.117>.
- Sedlazeck FJ, Rescheneder P, Smolka M, Fang H, Nattestad M, von Haeseler A, Schatz MC. 2018. Accurate detection of complex structural variations using single-molecule sequencing. *Nat Methods* 15:461–468. <https://doi.org/10.1038/s41592-018-0001-7>.