The Chromosome-Scale Genome Resource for Two Endophytic *Fusarium* species, *F. culmorum* and *F. pseudograminearum*

Xiaoliang Shan,^{1,2,3} Ying Zhu,⁴ Regina Redman,⁵ Russell J. Rodriguez,⁵ and Zhilin Yuan^{1,2,†}

¹ State Key Laboratory of Tree Genetics and Breeding, Chinese Academy of Forestry, Beijing 100081, China

² Nanjing Forestry University, Nanjing 210037, China

³ Research Institute of Subtropical Forestry, Chinese Academy of Forestry, Hangzhou 311400, China

⁴ Institute of Biology, Gansu Academy of Sciences, Lanzhou 730000, China

⁵ Adaptive Symbiotic Technologies, University of Washington, Seattle, WA 98195, U.S.A.

Abstract

Genus *Fusarium* (Ascomycota, Hypocreales, Nectriaceae) includes many economically important plant pathogens that cause devastating diseases of a wide range of crops and trees. Interestingly, there is increasing evidence that some *Fusarium* species also live as endophytes and benefit plant growth and stress tolerance. In this work, we sequence the whole genomes of endophytic *F. culmorum* and *F. pseudograminearum*, isolated from a coastal dunegrass (*Leymus mollis*), using long-read single-molecule real-time sequencing technology. Their genomes are assembled into four chromosomes and a mitochondrial genome with a total assembly size of 40.05 and 42.90 M, respectively. This resource should not only facilitate functional studies designed to better understand what makes the two *Fusarium* species such successful plant-beneficial fungi but should also reveal their genome evolution and adaptation.

Genome Announcement

Fusarium culmorum and *F. pseudograminearum* are the most common causal agents of *Fusarium* crown rot and *Fusarium* head blight worldwide, in particular on wheat and barley (Akinsanmi et al. 2004; Alahmad et al. 2018; Backhouse and Burgess 2002; Knight and Sutherland 2017; Scherm et al. 2013). Taxonomically, they belong to the *F. graminearum* subclade within the *F. sambucinum* species complex (Waalwijk et al. 2018). In the past few years, genome sequences of *F. culmorum* and *F. pseudograminearum* and other pathogenic *Fusarium* species (*F. graminearum*, *F. oxysporum*, and others) are available (Cuomo et al. 2007; Gardiner et al. 2018; Ma et al. 2010, 2013; Schmidt et al. 2018), providing an important genomic and genetic resource for understanding their evolution, pathogenesis, and lifestyle.

Interestingly, there is increasing evidence that some *Fusarium* species also live as endophytes and benefit plant growth and stress tolerance (Maciá-Vicente et al. 2008; Skiada et al. 2019), suggesting that our current knowledge of *Fusarium* biology and their strain-specific interactions with host plants is still incomplete. In previous investigations, *F. culmorum* and *F. pseudograminearum*, isolated from a coastal dunegrass (*Leymus mollis*), are

Funding

This work was founded by the Open Fund of State Key Laboratory of Tree Genetics and Breeding (TGB2018001), Fundamental Research Funds for the Central Nonprofit Research of Chinese Academy of Forestry (CAFYBB2019ZA001-3) and National Natural Science Foundation of China (number 31722014).

Keywords

fungal endophytes, long-read assembly, mitogenome, symbiosis

[†]Corresponding author: Z. Yuan; yuanzl@caf.ac.cn

The author(s) declare no conflict of interest.

Accepted for publication 17 January 2021.

shown to significantly improve plant growth and salinity resistance (Pan et al. 2018; Redman et al. 2011; Rodriguez et al. 2008). Thus, these two endophytic *Fusarium* species provide an excellent opportunity to not only better understand what makes these fungi such successful beneficial fungi associated with a wide range of host plants but, also, to reveal their genome evolution and adaptation. We surmise that there are possible extensive intragenomic variations related to reduced pathogenicity when compared with their conspecific pathogenic partners.

To this end, we used the single-molecule real-time (SMRT) sequencing developed by Pacific Biosciences (PacBio) to generate complete genome assemblies. Sequencing was performed at the Beijing Novogene Bioinformatics Technology Co., Ltd. For the PacBio library, 20-kb inserts for each strain were constructed using the SMRTbell template prep kit 1.0. One SMRTCell was used to sequence the DNA library with the PacBio Sequel II system. To polish the PacBio long reads-based genome assembly, a paired-end Illumina DNA library with a 500-bp insert size was sequenced on the Illumina HiSeq 2500. Based on the Illumina short reads data, we analyzed the K-mer distribution of the two genomes with Jellyfish (Marçais and Kingsford 2011) and estimated the genome size of the two genomes. We obtained 16.7 Gb of long-read data for *F. culmorum* Class2-1B, with a scaffold N₅₀ length of 9.63 M. For *F. pseudograminearum* Class2-1C, 19.7 Gb of long reads were collected, with a scaffold N₅₀ length of 9.15 M. Genome assembly and error correction in the long reads was conducted using MECAT2 (Xiao et al. 2017), and the initial assembly was further polished with Pilon (v. 1.22), using the Illumina short-reads data (Walker et al. 2014). The resulting assembly of Class2-1B had a total genome size of 40.05 Mb (nuclear genome size = 39.91 Mb and mitochondrial genomes (mitogenome) size = 139,406 bp) (Table 1) and produced a total of six scaffolds. We assembled them into four chromosomes of Class2-1B, two of which are telomere-to-telomere, while the other two have an identified telomere only at one end. We found the tandem repeat six-nucleotide sequences of TTAGGG (or the complementary DNA strand sequence AATCCC) at the end of scaffolds. This data indicates that almost all scaffolds approach the length of intact chromosomes (Aksenova and Mirkin 2019). The remaining two short scaffolds, representing the mitogenome, was identified by BLAST search, and overall GC content of 31.2%. Likewise, the resulting assembly of Class2-1C had a total genome size of 42.90 Mb (nuclear genome size = 42.76 Mb and mitogenome size = 136,045 bp) and produced a total of seven scaffolds, which we also assembled into four chromosomes, three of which are telomere-to-telomere, the other one carries only a telomere. We recovered three remaining short scaffolds (ranging from 10 to 115 kb) as a mitogenome, with overall GC content of 34.6%. The two mitogenomes both showed an average sequence identity of >98% when comparing their conspecific pathogenic partners (Kulik et al. 2020). Gene annotation and genome-assembly quality were evaluated using BUSCO (benchmarking universal single-copy orthologs) with the Fungi_odb10 dataset (v. 4.0.6), and the results showed 98.8 and 99.1% completeness (a total of 758 BUSCO groups were searched) for Class2-1B and Class2-1C respectively, indicating the reliability of our genome assembly and annotation (Simão et al. 2015) (Table 1).

Table 1. Genome characteristics and predicted features of the assembled *Fusarium culmorum* Class2-1B and *Fusarium pseudograminearum* (Class2-1C)

Characteristics	<i>F. culmorum</i>	<i>F. pseudograminearum</i>
Total genome size (Mb)	40.05	42.90
Nuclear genome size (Mb)	39.91	42.76
Mitogenome size (bp)	136,406	136,045
N ₅₀ scaffold length (Mb)	9.63	9.15
Chromosomes	4	4
Scaffolds	6	7
Genome coverage	443	519
G+C (%)	47.4	47.0
N ₅₀ scaffold average (Mb)	1.19	1.81
Total transposable elements (Mb)	1.55	2.04
Total number of gene	11,450	11,221
Average gene length (bp)	1,653	1,633
Genome BUSCO (%)	98.8	99.1

For annotation of transposable elements (TEs), RepeatMasker (v. 4.07) was used with the Repbase database (v. 23.06) (Bao et al. 2015) to identify the known TEs. Meanwhile, TEs were also modeled *ab initio*, using RepeatModeler (v. 1.0.11) and LTR_FINDER (Jurka et al. 2005) with default parameter. Approximately 1.55 and 2.04 Mb of TEs (4.13 and 5.37% of the total genome) were identified in the Class2-1B and Class2-1C, respectively. The structural annotation of protein-coding genes was performed by combining the results from *ab initio* and homolog-based prediction (Rigden 2017). A total of 11,450 and 112,21 intact protein-coding gene models were predicted as consensus gene sets in Class2-1B and Class2-1C, respectively (Table 1), using the MAKER (v. 2.31.9) annotation pipeline. In total, we found that 97.06 and 96.93% of genes in Class2-1B and Class2-1C, respectively, could be annotated with InterProScan, Gene Ontology, Kyoto Encyclopedia of Genes and Genomes, and the nonredundant database.

In summary, the high quality of genome sequences of *F. culmorum* Class2-1B and *F. pseudograminearum* Class2-1C will provide a valuable resource for future comparative genomic studies. Raw data and the assembled sequences have been submitted to the National Center for Biotechnology Information database, with BioProject accession number PRJNA643593 and BioSample accession numbers SAMN15354315 and SAMN15354316. Accession numbers for the assembled genomes are CP064747 to CP064752 (*Fusarium culmorum*) and CP064753 to CP064759 (*Fusarium pseudograminearum*). Sequence Read Archive numbers are SRR12095177 and SRR12095176.

Author-Recommended Internet Resources

Fungi odb10 dataset: <https://busco.ezlab.org/frames/fungi.htm>

Repbase database: <https://www.girinst.org/server/RepBase>

RepeatMasker: <http://repeatmasker.org>

RepeatModeler: <http://www.repeatmasker.org/RepeatModeler>

Literature Cited

- Akinsanmi, O. A., Mitter, V., Simpfendorfer, S., Backhouse, D., and Chakraborty, S. 2004. Identity and pathogenicity of *Fusarium* spp. isolated from wheat fields in Queensland and northern New South Wales. *Aust. J. Agric. Res.* 55:97-107.
- Aksenova, A. Y., and Mirkin, S. M. 2019. At the beginning of the end and in the middle of the beginning: Structure and maintenance of telomeric DNA repeats and interstitial telomeric sequences. *Genes (Basel)* 10:118.
- Alahmad, S., Simpfendorfer, S., Bentley, A. R., and Hickey, L. T. 2018. Crown rot of wheat in Australia: *Fusarium pseudograminearum* taxonomy, population biology and disease management. *Australas. Plant Pathol.* 47:285-299.
- Backhouse, D., and Burgess, L. W. 2002. Climatic analysis of the distribution of *Fusarium graminearum*, *F. pseudograminearum* and *F. culmorum* on cereals in Australia. *Australas. Plant Pathol.* 31:321-327.
- Bao, W., Kojima, K. K., and Kohany, O. 2015. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob. DNA* 6:11.
- Cuomo, C. A., Gildener, U., Xu, J. R., Trail, F., Turgeon, B. G., Di Pietro, A., Walton, J. D., Ma, L. J., Baker, S. E., Rep, M., Adam, G., Antoniw, J., Baldwin, T., Calvo, S., Chang, Y. L., Decaprio, D., Gale, L. R., Gnerre, S., Goswami, R. S., Hammond-Kosack, K., Harris, L. J., Hilburn, K., Kennell, J. C., Kroken, S., Magnuson, J. K., 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, M. I. G., Seong, K. Y., Tetko, I. V., Urban, M., Waalwijk, C., Ward, T. J., Yao, J., Birren, B. W., and Kistler, H. C. 2007. The *Fusarium graminearum* genome reveals a link between localized polymorphism and pathogen specialization. *Science* 317:1400-1402.
- Gardiner, D. M., Benfield, A. H., Stiller, J., Stephen, S., Aitken, K., Liu, C., and Kazan, K. 2018. A high-resolution genetic map of the cereal crown rot pathogen *Fusarium pseudograminearum* provides a near-complete genome assembly. *Mol. Plant Pathol.* 19:217-226.
- Jurka, J., Kapitonov, V. V., Pavlicek, A., Klonowski, P., Kohany, O., and Walichiewicz, J. 2005. Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet. Genome Res.* 110:462-467.
- Knight, N. L., and Sutherland, M. W. 2017. Assessment of *Fusarium pseudograminearum* and *F. culmorum* biomass in seedlings of potential host cereal species. *Plant Dis.* 101:2116-2122.
- Kulik, T., Brankovics, B., van Diepeningen, A. D., Bilska, K., Żelechowski, M., Myszczyński, K., Molcan, T., Stakheev, A., Stenglein, S., Beyer, M., Pasquali, M., Sawicki, J., Wyrębek, J., and Batur-Cieśniewska, A. 2020. Diversity of mobile genetic elements in the mitogenomes of closely related *Fusarium culmorum* and *F. graminearum* sensu stricto strains and its implication for diagnostic purposes. *Front. Microbiol.* 11:1002.
- Ma, L.-J., Geiser, D. M., Proctor, R. H., Rooney, A. P., O'Donnell, K., Trail, F., Gardiner, D. M., Manners, J. M., and Kazan, K. 2013. *Fusarium* pathogenomics. *Annu. Rev. Microbiol.* 67:399-416.
- Ma, L. J., van der Does, H. C., Borkovich, K. A., Coleman, J. J., Daboussi, M. J., Di Pietro, A., Dufresne, M., Freitag, M., Grabherr, M., Henrissat, B., Houterman, P. M., Kang, S., Shim, W. B., Woloshuk, C., Xie, X., Xu, J. R., Antoniw, J., Baker, S. E., Bluhm, B. H., Breakspear, A., Brown, D. W., Butchko, R. A. E., Chapman, S., Coulson, R., Coutinho, P. M., Danchin, E. G. J., Diener, A., Gale, L. R., Gardiner, D. M., Goff, S., Hammond-Kosack, K. E., Hilburn, K., Hua-Van, A., Jonkers, W., Kazan, K., Kodira, C. D., Koehrsen, M., Kumar, L., Lee, Y. H., Li, L., Manners, J. M., Miranda-Saavedra, D., Mukherjee, M., Park, G., Park, J., Park, S. Y., Proctor, R. H., Regev, A., Ruiz-Roldan, M. C., Sain, D., Sakthikumar, S., Sykes, S., Schwartz, D. C., Turgeon, B. G., Wapinski, I., Yoder, O., Young, S., Zeng, Q., Zhou, S., Galagan, J., Cuomo, C. A., Kistler, H. C., and Rep, M. 2010. Comparative genomics reveals mobile pathogenicity chromosomes in *Fusarium*. *Nature* 464:367-373.
- Maciá-Vicente, J. G., Jansson, H. B., Abdullah, S. K., Descals, E., Salinas, J., and Lopez-Llorca, L. V. 2008. Fungal root endophytes from natural vegetation in Mediterranean environments with special reference to *Fusarium* spp. *FEMS Microbiol. Ecol.* 64:90-105.
- Marçais, G., and Kingsford, C. 2011. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* 27:764-770.
- Pan, X. Y., Sun, H. J., and Yuan, Z. L. 2018. Toxin accumulation of three *Leymus mollis*-associated endophytic *Fusarium* isolates and their effects on growth and salt tolerance of *Liquidambar styraciflua* seedlings. *Forest Res.* 31:64-73.

- Redman, R. S., Kim, Y. O., Woodward, C. J. D. A., Greer, C., Espino, L., Doty, S. L., and Rodriguez, R. J. 2011. Increased fitness of rice plants to abiotic stress via habitat adapted symbiosis: A strategy for mitigating impacts of climate change. *PLoS One* 6:e14823.
- Rigden, D. J. 2017. *From protein structure to function with bioinformatics*, 2nd Ed. Springer, Dordrecht, The Netherlands.
- Rodriguez, R. J., Henson, J., Van Volkenburgh, E., Hoy, M., Wright, L., Beckwith, F., Kim, Y. O., and Redman, R. S. 2008. Stress tolerance in plants via habitat-adapted symbiosis. *ISME J.* 2:404-416.
- Scherm, B., Balmas, V., Spanu, F., Pani, G., Delogu, G., Pasquali, M., and Migheli, Q. 2013. *Fusarium culmorum*: Causal agent of foot and root rot and head blight on wheat. *Mol. Plant Pathol.* 14:323-341.
- Schmidt, R., Durling, M. B., de Jager, V., Menezes, R. C., Nordkvist, E., Svatoš, A., Dubey, M., Lauterbach, L., Dickschat, J. S., Karlsson, M., and Garbeva, P. 2018. Deciphering the genome and secondary metabolome of the plant pathogen *Fusarium culmorum*. *FEMS Microbiol. Ecol.* 94:fiy078.
- Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., and Zdobnov, E. M. 2015. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31:3210-3212.
- Skiada, V., Faccio, A., Kavroulakis, N., Genre, A., Bonfante, P., and Papadopoulos, K. K. 2019. Colonization of legumes by an endophytic *Fusarium solani* strain FsK reveals common features to symbionts or pathogens. *Fungal Genet. Biol.* 127:60-74.
- Waalwijk, C., Taga, M., Zheng, S. L., Proctor, R. H., Vaughan, M. M., and O'Donnell, K. 2018. Karyotype evolution in *Fusarium*. *IMA Fungus* 9:13-26.
- Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C. A., Zeng, Q., Wortman, J., Young, S. K., and Earl, A. M. 2014. Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9:e112963.
- Xiao, C. L., Chen, Y., Xie, S. Q., Chen, K. N., Wang, Y., Han, Y., Luo, F., and Xie, Z. 2017. MECAT: Fast mapping, error correction, and de novo assembly for single-molecule sequencing reads. *Nat. Methods* 14:1072-1074.