High-Quality Genome Sequence Resource for *Fusarium andiyazi* Causing Pokkah Boeng Disease of Sugarcane in China

Yixue Bao,¹ Kaiyuan Pan,¹ Muhammad Tahir Khan,² Baoshan Chen,¹ and Muqing Zhang^{1,†}

¹ Guangxi Key Lab of Sugarcane Biology, State Key Lab for Conservation and Utilization of Sub-tropical Agri-Biological Resources, Guangxi University, Nanning, 530005, China

² Nuclear Institute of Agriculture (NIA), Tando Jam, 70060, Pakistan

Abstract

Sugarcane pokkah boeng disease (PBD) is emerging as a prevalent foliar disease in China. This airborne disease is caused by the *Fusarium* species complex. To investigate the diversity and evolution of *Fusarium* spp., we performed whole-genome sequencing of *Fusarium andiyazi* YN28 using a combination of Oxford Nanopore and Illumina technology. The *F. andiyazi* YN28 genome was sequenced, assembled, and annotated. A high-quality genome was assembled into 24 contigs with an N₅₀ of 2.80 Mb. The genome assembly generated a total size of 44.1 Mb with a GC content of 47.64%. In total, 15,508 genes were predicted, including 794 genes related to the carbohydrate-active enzymes, 397 noncoding RNA, 155 genes associated with transporter classification, 4,550 genes linked to pathogen–host interactions, and 269 genes involved in effector proteins. Collectively, our results will provide insight into the host–pathogen interactions and will facilitate the breeding of new varieties of sugarcane resistant to PBD.

Genome Announcement

Pokkah boeng disease (PBD) is a major constraint to sugarcane production globally and it is emerging as a predominant foliar disease in China (Vishwakarma et al. 2013; Wang et al. 2016). Typical symptoms include crumpled and twisted leaves, the formation of irregular red-dish streaks, and chlorotic tissue. The most serious infection occurs at the growing tip, which deforms the entire top (Mohamed Sidique et al. 2008; Sidique and Nordahliawate 2007). The PBD was reported to be caused by different species of *Fusarium* such as *Fusarium sacchari*, *F. verticillioides*, and *F. proliferatum* (Bao et al. 2020; Lin et al. 2014).

In 2019, we formally described and established 34 isolates as *F. andiyazi* associated with sugarcane PBD, based on morphological characteristics and molecular phylogenetic analyses of translation elongation factor 1- α (TEF-1 α , EF1/EF2), DNA-directed RNA polymerase II largest subunit (RPB1, Fa/G2R), and DNA-directed RNA polymerase II second largest subunit (RPB2, 5f2/7cr and 7cf/11ar), with the GenBank accession numbers MT911951 to MT911984, MK983446 to MK983479, and MK886811 to MK886828 and MK983384 to MK983399, respectively (Bao et al. 2020). Simultaneously, 32 (94%) *F. andiyazi* isolates

Funding

This work was supported by the China Agricultural Research System of MFA and MAFA (CARS170190) and the open Project of Guangxi Key Lab of Sugarcane Biology (GXKLSCB-202001).

Keywords

Fusarium andiyazi, genome, pathogen, pokkah boeng disease, sugarcane

†Corresponding author: M. Zhang; zmuqing@163.com

Y. Bao and K. Pan contributed equally to this article.

**The e-Xtra logo stands for “electronic extra” and indicates that a supplementary table is published online.

The author(s) declare no conflict of interest.

Accepted for publication 1 April 2021.

Copyright © 2021 The Author(s). This is an open access article distributed under the CC BY-NC-ND 4.0 International license.

Table 1. Genome assembly statistics of *Fusarium andiyazi* YN28

Variables ^a	Statistics
Genome size (bp)	44,104,723
Number of contigs	24
Number of scaffolds	12
Maximum scaffold length (bp)	6,367,790
GC content (%)	47.64
BUSCO completeness (%)	99.31
Number of genes	15,508
Number of secreted proteins	1,240
Number of effector proteins	269
Number of CAZymes	794

^aBUSCO = Benchmarking Universal Single-Copy Ortholog and CAZymes = carbohydrate-active enzymes.

were collected from Yunnan Province, located at the intersection zone of the subtropical monsoon and the Qinghai-Tibet alpine climates. The huge diurnal temperature difference and unique geographical environment may cause the variation and evolution of *F. andiyazi*.

Most of the other reported *Fusarium* sp. genome sequencing projects utilized first- and next-generation sequencing (Sanger, SOLiD, and 454), including *F. verticillioides* 7600 (AAIM000000000), *F. oxysporum* 4287 (AAXH000000000), *F. graminearum* PH-1 (AACM000000000), *F. fujikuroi* IMI58289 (GCA_900079805), and *F. circinatum* FSP34 (AYJV000000000) (Cuomo et al. 2007; Ma et al. 2010; Wiemann et al. 2013; Wingfield et al. 2012). The limitations of the earlier sequencing technologies offer hurdles in assembling a high-quality genome.

Our study is the first report of whole-genome sequencing of *F. andiyazi* using a combination of paired-end sequencing with the Illumina HiSeq4000 platform and Oxford Nanopore sequencing with the PromethION platform. We generated a high-quality genome assembly with a reduced mismatch rate in the sequencing process. For *F. andiyazi* YN28, to improve the genome assembly precision rate and carry out the genome size estimation, a paired-end library of 350 bp was generated with 4.18 Gb of data (94× fold coverage). Next, the 7.45 Gb of sequencing data were generated from 163-fold coverage of long reads (20 kb). The genome was assembled using Canu v.1.5 and WTDBG2 into 24 contigs with an N₅₀ length of 2.80 Mb; therewith, it was further improved by error correction through Pilon v.1.2 (Koren et al. 2017; Ruan and Li 2020; Walker et al. 2014). Twenty-four contigs, ranging in length from 56.2 kb to 4.92 Mb, were assembled into 12 scaffolds using Chromosomer software (Tamazian et al. 2016). Nine scaffolds had telomeric repeat sequences on both ends and two had telomeric repeats at the 5' end. Therefore, the genome contained at least nine complete chromosomes. The genome also had one integrated circular mitochondrial DNA that was approximately 80.3 kb. Two scaffolds without telomeric repeats on both ends might represent dispensable chromosomes and require further characterization (Supplementary Table S1). The genome assembly generated a total size of 44.1 Mb with a GC content of 47.64%. Furthermore, the quality of the genome assembly was assessed by using Benchmarking Universal Single-Copy Ortholog (BUSCO v.2.0) (eukaryota_odb9, 290 Lineage BUSCOs) (Simão et al. 2015). In total, 288 (99.31%) of the completed BUSCO sets were covered by the genome. Only 1 (0.34%) fragmented BUSCO and 1 (0.34%) missing BUSCO were identified in the YN28 genome (Table 1).

The total number of genes for YN28 was predicted to generate 15,508 genes using ab initio and homology-based methods (Haas et al. 2008; Keilwagen et al. 2016; Korf 2004; Majoros et al. 2004; Stanke and Waack 2003). Moreover, gene number was annotated using general databases, including Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG), nonredundant proteins (Nr); and special databases, including carbohydrate-active enzymes (CAZyme), pathogen-host interactions (PHI), and transporter classification (TCDB) (Ashburner et al. 2000; Cantarel et al. 2009; Deng et al. 2006; Kanehisa et al. 2004; Saier et al. 2006; Winnenburg et al. 2006). Among the annotated genes, 3,644 (23.50%) were annotated to the KEGG database, while 15,495 (99.92%) were annotated to the Nr database on the basis of the BLAST method. Furthermore, based on Blast2GO, 7,912 (51.02%) genes were mapped to the GO database (Conesa et al. 2005).

Special database analyses revealed that YN28 has 794 genes related to the CAZymes families, which play an important role in fungal pathogenicity. Additionally, 155 genes were annotated to the TCDB and 4,550 genes to the PHI database. In total, 1,240 genes were predicted to be secreted proteins, of which 269 effector proteins were predicted by the EffectorP program (Sperschneider et al. 2016). The effector proteins data can provide insight into the YN28 virulence and host–pathogen interaction. It is worth noting that FAND_06030 containing the LysM chitin-binding domain was observed in YN28 genome. This effector protein can inhibit the chitinase activity of the immune response (Mentlak et al. 2012). The activity of hydrolyzed lipids can be inhibited by lipase inhibitors encoded by FAND_00875 (Voigt et al. 2005). The relative abundance of repetitive elements was assessed using the Repeat Masker program (Chen 2004); the repetitive elements accounted for only 3.66% of the genome. The most abundant repetitive element in YN28 was a simple sequence repeat, which shared 2.10% of the assembly. Using tRNAscan-SE and Infernal 1.1 program, we also predicted 397 noncoding RNA (ncRNA), including 295 transfer RNA, 62 ribosomal RNA, and another 40 ncRNA on the genome (Lowe and Eddy 1997; Nawrocki and Eddy 2013). The data provided a better understanding of the *F. andiyazi* pathogenesis.

The biosynthetic gene clusters (BGCs) on a genome indicate the secondary metabolite production potential. BGCs play an important role in fungus–host interactions (Rokas et al. 2020). The BGCs in the YN28 genome were predicted on the basis of Antibiotics & Secondary Metabolite Analysis Shell (antiSMASH) (fungal version 4.0.0), including 13 polyketide synthases (PKS), 14 nonribosomal peptide synthetases (NRPS), 3 PKS-NRPS, 3 dimethylallyl tryptophan synthases, and 7 terpene cyclases.

Collectively, the combination of next- and third-generation sequencing technology was used for genome sequencing of *F. andiyazi* YN28. This is the first report of the high-quality genome of *F. andiyazi*. The whole-genome data will provide insight into *F. andiyazi* pathogenesis. This study highlights the significance of conducting an omics study to reveal the scientific apprehension of sugarcane PBD and to divulge the potential approaches to breed new resistant varieties. This Whole-Genome Shotgun project has been deposited at DNA Data Bank of Japan/European Nucleotide Archive/GenBank under the accession JADOXN000000000 (BioProject: PRJNA674737; BioSample: SAMN16674204).

Author-Recommended Internet Resource

antiSMASH, fungal version 4.0.0: <https://fungismash.secondarymetabolites.org>

Literature Cited

- Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., and Sherlock, G. 2000. Gene ontology: Tool for the unification of biology. *Nat. Genet.* 25:25-29.
- Bao, Y. X., Huang, Z., Li, T. M., Yao, Z. T., Powell, C. A., and Zhang, M. Q. 2020. First report of *Fusarium andiyazi* causing sugarcane pokkah boeng disease in China. *Plant Dis.* 104:286.
- Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. 2009. The Carbohydrate-Active EnZymes database (CAZy): An expert resource for Glycogenomics. *Nucleic Acids Res.* 37:D233-D238.
- Chen, N. 2004. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinf.* 5:4.10.1-4.10.14.
- Conesa, A., Götz, S., García-Gómez, J. M., Terol, J., Talón, M., and Robles, M. 2005. Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21:3674-3676.
- Cuomo, C. A., Güldener, 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.
- Deng, Y. Y., Li, J. Q., Wu, S. F., Zhu, Y. P., Chen, Y. W., and He, F. C. 2006. Integrated nr database in protein annotation system and its localization. *Comput. Eng.* 32:71-72.
- Haas, B. J., Salzberg, S. L., Zhu, W., Pertea, M., Allen, J. E., Orvis, J., White, O., Buell, C. R., and Wortman, J. R. 2008. Automated eukaryotic gene structure annotation using EVIDENCEModeler and the Program to Assemble Spliced Alignments. *Genome Biol.* 9:R7.
- Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y., and Hattori, M. 2004. The KEGG resource for deciphering the genome. *Nucleic Acids Res.* 32:D277-D280.
- Keilwagen, J., Wenk, M., Erickson, J. L., Schattat, M. H., Grau, J., and Hartung, F. 2016. Using intron position conservation for homology-based gene prediction. *Nucleic Acids Res.* 44:e89.
- Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., and Phillippy, A. M. 2017. Canu: Scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* 27:722-736.
- Korf, I. 2004. Gene finding in novel genomes. *BMC Bioinf.* 5:59.
- Lin, Z., Xu, S., Que, Y., Wang, J., Comstock, J. C., Wei, J., McCord, P. H., Chen, B., Chen, R., and Zhang, M. 2014. Species-specific detection and identification of *Fusarium* species complex, the causal agent of sugarcane pokkah boeng in China. *PLoS One* 9:e104195.
- Lowe, T. M., and Eddy, S. R. 1997. tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 25:955-964.

- 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.
- Majoros, W. H., Pertea, M., and Salzberg, S. L. 2004. TigrScan and GlimmerHMM: Two open source ab initio eukaryotic gene-finders. *Bioinformatics* 20:2878-2879.
- Mentlak, T. A., Kombrink, A., Shinya, T., Ryder, L. S., Otomo, I., Saitoh, H., Terauchi, R., Nishizawa, Y., Shibuya, N., Thomma, B. P. H. J., and Talbot, N. J. 2012. Effector-mediated suppression of chitin-triggered immunity by *Magnaporthe oryzae* is necessary for rice blast disease. *Plant Cell* 24:322-335.
- Mohamed Sidiq, S. N., Mohd Zainuddin, N. A. I., Abdul Razak, A., and Salleh, B. 2008. Distribution, morphological characterization and pathogenicity of *Fusarium sacchari* associated with pokkah boeng disease of sugarcane in peninsular Malaysia. *IEEE J. Quantum Electron.* 25:965-975.
- Nawrocki, E. P., and Eddy, S. R. 2013. Infernal 1.1: 100-Fold faster RNA homology searches. *Bioinformatics* 29:2933-2935.
- Rokas, A., Mead, M. E., Steenwyk, J. L., Raja, H. A., and Oberlies, N. H. 2020. Biosynthetic gene clusters and the evolution of fungal chemodiversity. *Nat. Prod. Rep.* 37:868-878.
- Ruan, J., and Li, H. 2020. Fast and accurate long-read assembly with wtdbg2. *Nat. Methods* 17:155-158.
- Saier, M. H., Jr., Tran, C. V., and Barabote, R. D. 2006. TCDB: The Transporter Classification Database for membrane transport protein analyses and information. *Nucleic Acids Res.* 34:D181-D186.
- Sidiq, M., and Nordahliawate, S. 2007. Pathogenicity and aethiology of *Fusarium* species associated with pokkah boeng disease on sugarcane. Masters thesis, Universiti Sains Malaysia.
- 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.
- Sperschneider, J., Gardiner, D. M., Dodds, P. N., Tini, F., Covarelli, L., Singh, K. B., Manners, J. M., and Taylor, J. M. 2016. EffectorP: Predicting fungal effector proteins from secretomes using machine learning. *New Phytol.* 210:743-761.
- Stanke, M., and Waack, S. 2003. Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics* 19:ii215-ii225.
- Tamazian, G., Dobrynin, P., Krashenninnikova, K., Komissarov, A., Koepfli, K. P., and O'Brien, S. J. 2016. Chromosomer: A reference-based genome arrangement tool for producing draft chromosome sequences. *Gigascience* 5:s13742-016-0141-6.
- Vishwakarma, S. K., Kumar, P., Nigam, A., Singh, A., and Kumar, A. 2013. Pokkah boeng: An emerging disease of sugarcane. *J. Plant Pathol. Microbiol.* 4:1000170.
- Voigt, C. A., Schäfer, W., and Salomon, S. 2005. A secreted lipase of *Fusarium graminearum* is a virulence factor required for infection of cereals. *Plant J.* 42:364-375.
- Walker, B. J., Abee, 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.
- Wang, Z. P., Duan, W. X., Li, Y. J., Liang, Q., Zhou, Z. G., Zhang, M. Q., and Lin, S. H. 2016. Establishment of an evaluation system for field resistance against sugarcane pokkah boeng. *J. South China Agric. Univ.* 37:67-72.
- Wiemann, P., Sieber, C. M. K., von Barga, K. W., Studt, L., Niehaus, E. M., Espino, J. J., Huß, K., Michielse, C. B., Albermann, S., Wagner, D., Bergner, S. V., Connolly, L. R., Fischer, A., Reuter, G., Kleigrewe, K., Bald, T., Wingfield, B. D., Ophir, R., Freeman, S., Hippler, M., Smith, K. M., Brown, D. W., Proctor, R. H., Münsterkötter, M., Freitag, M., Humpf, H. U., Güldener, U., and Tudzynski, B. 2013. Deciphering the cryptic genome: Genome-wide analyses of the rice pathogen *Fusarium fujikuroi* reveal complex regulation of secondary metabolism and novel metabolites. *PLoS Pathog.* 9:e1003475.
- Wingfield, B. D., Steenkamp, E. T., Santana, Q. C., Coetzee, M. P. A., Bam, S., Barnes, I., Beukes, C. W., Chan, W. Y., de Vos, L., Fourie, G., Friend, M., Gordon, T. R., Herron, D. A., Holt, C., Korf, I., Kvas, M., Martin, S. H., Mlonyeni, X. O., Naidoo, K., Phasha, M. M., Postma, A., Reva, O., Roos, H., Simpson, M., Slinski, S., Slippers, B., Sutherland, R., van der Merwe, N. A., van der Nest, M. A., Venter, S. N., Wilken, P. M., Yandell, M., Zipfel, R., and Wingfield, M. J. 2012. First fungal genome sequence from Africa: A preliminary analysis. *S. Afr. J. Sci.* 108:104-112.
- Winnenburg, R., Baldwin, T. K., Urban, M., Rawlings, C., Köhler, J., and Hammond-Kosack, K. E. 2006. PHI-base: A new database for pathogen host interactions. *Nucleic Acids Res.* 34:D459-D464.