AMERICAN SOCIETY FOR MICROBIOLOGY

Microbiology
Resource Announcements

First Draft Genome Sequence of Wheat Spot Blotch Pathogen *Bipolaris sorokiniana* BS_112 from India, Obtained Using Hybrid Assembly

Rashmi Aggarwal,[a] Sapna Sharma,[a] Kartar Singh,[a*] Malkhan Singh Gurjar,[a] Mahender Singh Saharan,[a] Sangeeta Gupta,[a] Bishnu Maya Bashyal,[a] Kishor Gaikwad[b]

[a]Fungal Molecular Biology Laboratory, Division of Plant Pathology, ICAR-Indian Agricultural Research Institute, New Delhi, India
[b]ICAR-National Research Centre on Plant Biotechnology, New Delhi, India

ABSTRACT *Bipolaris sorokiniana* is a devastating fungal pathogen causing spot blotch of wheat. We report here the first draft genome of *Bipolaris sorokiniana* strain BS_112 from India using sequence reads from the Ion Torrent, Illumina HiSeq, and Nanopore platforms. The genome size was estimated at 35.64 Mb with an average G+C content of 50.20%.

Spot blotch caused by *Bipolaris sorokiniana* (phylum Ascomycota, class Dothideomycetes) is one of the most devastating diseases of wheat in the warmer areas of the world (1). The disease is distributed widely across the wheat-growing areas of the world, including Africa, South America, Australia, Canada, Asia, and particularly the Indian subcontinent under the prevalent warm and humid conditions. Yield losses of between 20 and 80% have been reported and may lead to complete devastation under the most severe conditions (2). The pathogen is hemibiotrophic, initially biotrophic and turning to a necrotrophic phase at later stages (3). PCR-based DNA fingerprinting grouped *B. sorokiniana* isolates according to their geographic locations in India (1). A sequence-characterized amplified region (SCAR) marker has been developed for detection of *B. sorokiniana* (4), but little is known about the molecular basis of pathogenesis in *B. sorokiniana* due to a paucity of resources for molecular genetics. Therefore, a draft genome sequence of *B. sorokiniana* was generated, assembled, and analyzed.

High-quality genomic DNA was extracted from monoconidial culture of a highly virulent strain of *B. sorokiniana*, BS_112 (GenBank accession number KU201275), grown at 26°C with constant shaking (150 rpm) in potato dextrose broth (HiMedia, India) for 3 to 4 days using a Qiagen DNeasy minikit. Then, the quality and quantity were assessed using a spectrophotometer, Qubit fluorometer, and Agarose gel electrophoresis. This pathogen was sequenced using three platforms, *viz.*, Illumina HiSeq, Ion Torrent, and Nanopore, following the manufacturers' instructions, including library preparation and sequencing. The TruSeq DNA PCR-free kit was used for Illumina library preparation with an insert size of 350 bp and a read size of 150 bp from each end, and a total of 36,915,848 paired-end reads were obtained. In the Ion Torrent library preparation, the fungal DNA was enzymatically sheared into suitably sized fragments using the Ion Xpress Plus fragment library kit. The barcode was applied, and the adapter-ligated library was size selected for a target peak of 330 bp using E-Gel Size Select 2% agarose gel (Invitrogen Corporation, USA). Finally, the adapter-ligated and size-selected library was amplified using the Ion Plus fragment library kit (Life Technologies [LT], USA), and the run was performed on an Ion PGM system version 1. A total of 845,589 reads were obtained using the Ion Torrent platform. The library preparation for the Oxford Nanopore platform was performed using a ligation sequencing kit

(SQK-LSK108) from Oxford Nanopore Technologies according to the instructions provided with the kit, and a total of 560,150 reads were generated. The Illumina reads were preprocessed using Trim Galore version 0.4.0 (5); Nanopore Fast5 data were base called using Albacore version 2.3.1 (6), and for the Ion Torrent reads, the bases with a quality (Q) score of $\geq$20 were considered for assembly. All three types of sequencing reads were used for the hybrid genome assembly using the MaSuRCA tool version 3.2.2 (7). Further, the contigs were processed to scaffolds using pyScaf version 3 (https://pypi.org/search/?q=PyScaf+version+3). The generated whole-genome size is 35.64 Mb with a G+C content of 50.2%, 235 scaffolds, and an N_{50} value of 1,654,800 bp. BUSCO provided quantitative measures for the assessment of the genome assembly using ortholog groups, and the evaluation was performed using fungal data sets. The BUSCO evaluation of completeness of the *B. sorokiniana* genome sequence assembly predicted that it was complete to about 97.6%. The gene prediction was performed using the AUGUSTUS tool version 2.2.5 (8) with the reference *Bipolaris sorokiniana* ND90Pr (GenBank accession number AEIN00000000). A total of 10,460 genes were predicted with an average gene density of 250 to 300 genes/Mb, which covers around 98% of predicted genes. The average gene length was 435 to 545 bp, the maximum gene length was 8,506 bp, and the minimum gene length was 50 bp. Gene ontology (GO) annotations of the genes were determined using Blast2GO version 2 (9). Out of 10,460 annotated genes, 1,024 (3,491), 493 (4,165), and 1,274 (7,784) genes (hits) were for biological processes, cellular components, and molecular functions, respectively. The *B. sorokiniana* genome was interrogated with the Pathogen Host Interactions (PHI) database version 4.3 (10), and in total, 3,627 genes were found to be homologous to proteins in the PHI database with an E value of $\leq$1e-05. Single-nucleotide polymorphism (SNP) prediction was processed using the SnpEff tool version 3.3h (11) with the Illumina HiSeq data and identified 93,122 variants containing 88,672 SNPs and 4,450 indels. Further, simple sequence repeat (SSR) identification using the MISA tool version 1 (https://webblast.ipk-gatersleben.de/misa/) showed 5,996 SSRs, and 146 of the 235 SSR-containing sequences were examined. The genome sequence of *Bipolaris sorokiniana* ND90Pr was used as reference for the analysis of SNPs and SSRs.

Data availability. This whole-genome sequencing project has been deposited in DDBJ/ENA/GenBank under the accession number RCTM00000000, BioProject number PRJNA489551, and BioSample identification (ID) SAMN09901300. The version described in this paper is the first version (GenBank accession number RCTM01000000). The internal transcribed spacer (ITS) sequence was deposited at NCBI with the accession number KU201275. The raw sequencing data of all the three platforms were deposited in the DDBJ SRA database under the accession numbers SRX6673825 (Ion Torrent), SRX6673824 (Illumina HiSeq), and SRX6100476 (Oxford Nanopore).

REFERENCES

1. Aggarwal R, Singh VB, Shukla R, Gurjar MS, Gupta S, Sharma TR. 2010. URP-based DNA fingerprinting of *Bipolaris sorokiniana* isolates causing spot blotch of wheat. J Phytopathol 158:210–216. https://doi.org/10.1111/j.1439-0434.2009.01603.x.

2. Duveiller E, Altamirano IG. 2000. Pathogenicity of *Bipolaris sorokiniana* isolates from wheat roots, leaves and grains in Mexico. Plant Pathol 49:235–242. https://doi.org/10.1046/j.1365-3059.2000.00443.x.

3. Aggarwal R, Das S, Jahani M, Singh DV. 2008. Histopathology of spot

blotch disease of wheat caused by *Bipolaris sorokiniana* (teleomorph: *Cochliobolus sativus*). Acta Phytopathol 43:23–30. https://doi.org/10.1556/APhyt.43.2008.1.3.

4. Aggarwal R, Gupta S, Banerjee S, Singh VB. 2011. Development of a SCAR marker for detection of *Bipolaris sorokiniana* causing spot blotch of wheat. Can J Microbiol 57:934–942. https://doi.org/10.1139/w11-089.

5. Krueger F. 2015. Trim Galore. A wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files. https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/.

6. Sahoo N. 2017. Sequence base-calling through Albacore software: a part of the Oxford Nanopore technology. Doctoral dissertation, Orissa University of Agriculture and Technology, Bhubaneswar, Orrisa, India.

7. Zimin AV, Marçais G, Puiu D, Roberts M, Salzberg SL, Yorke JA. 2013. The MaSuRCA genome assembler. Bioinformatics 29:2669–2677. https://doi.org/10.1093/bioinformatics/btt476.

8. Stanke M, Morgenstern B. 2005. AUGUSTUS: a Web server for gene prediction in eukaryotes that allows user-defined constraints. Nucleic Acids Res 33:W465–W467. https://doi.org/10.1093/nar/gki458.

9. Gotz S, Gomez JMG, Terol J, Williams TD, Nagaraj SH, Nueda MJ, Robles M, Talon M, Dopazo J, Conesa A. 2008. High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Res 36:3420–3435. https://doi.org/10.1093/nar/gkn176.

10. Urban M, Pant R, Raghunath A, Irvine AG, Pedro H, Hammond-Kosack KE. 2015. The Pathogen-Host Interactions database (PHI-base): additions and future developments. Nucleic Acids Res 43:D645–D655. https://doi.org/10.1093/nar/gku1165.

11. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118. iso-2; iso-3. Fly (Austin) 6:80–92. https://doi.org/10.4161/fly.19695.