



Improved Draft Genome Sequence of a Monoteliosporic Culture of the Karnal Bunt (*Tilletia indica*) Pathogen of Wheat

 Anil Kumar,^a  Pallavi Mishra,^{b,c}  Ranjeet Maurya,^a A. K. Mishra,^b Vijai K. Gupta,^d Pramod W. Ramteke,^e Soma S. Marla^f

^aDepartment of Molecular Biology and Genetic Engineering, College of Basic Sciences and Humanities, G. B. Pant University of Agriculture and Technology, Pantnagar, India

^bICAR-Agricultural Knowledge Management Unit, Indian Agricultural Research Institute, Pusa Campus, New Delhi, India

^cDepartment of Computational Biology and Bioinformatics, Sam Higginbottom University of Agriculture, Technology and Sciences, Allahabad, India

^dDepartment of Chemistry and Biotechnology, ERA Chair of Green Chemistry, Tallinn University of Technology, Tallinn, Estonia

^eDepartment of Biological Sciences, Sam Higginbottom University of Agriculture, Technology and Sciences, Allahabad, India

^fDivision of Genomic Resources, ICAR-National Bureau of Plant Genetic Resources, New Delhi, India

ABSTRACT Karnal bunt of wheat is an internationally quarantined fungal pathogen disease caused by *Tilletia indica* and affects the international commercial seed trade of wheat. We announce here the first improved draft genome assembly of a monoteliosporic culture of the *Tilletia indica* fungus, consisting of 787 scaffolds with an approximate total genome size of 31.83 Mbp, which is more accurate and near to complete than the previous version.

Karnal bunt (KB) of wheat crops was first discovered in Karnal, India, in 1931 (1) and is caused by the smut fungus *Tilletia indica*, a basidiomycete belonging to the subdivision Ustilaginomycotina. KB has become a major disease which hampers international wheat trade due to quarantine regulations imposed by several countries (2). In order to understand the offensive and defensive mechanism(s) in the wheat-*T. indica* interaction at the molecular level, complete genomic information of fungus is of paramount importance. To date, attempts have been made by several research groups to decipher the genome sequence using next-generation sequencing platforms (3–5). The redundancy in the genetic profiling of the fungus *T. indica* necessitates the refinement of available genome sequences of monoteliosporic cultures of *T. indica* for better appreciation of fungal biology and disease management (6).

The Karnal bunt fungus *T. indica* assemblies for isolates DAOM 236416 (4) and RAKB_UP_1 (5) were retrieved from the NCBI database (<https://www.ncbi.nlm.nih.gov>) and were preprocessed accordingly for the reconciliation algorithms.

The Illumina and PacBio sequence reads of the TiK (*Tilletia indica* Karnal) isolate (3) were quality checked using FastQC version 0.11.5 (7) and fastQValidator version 0.1.1 (8), and the reads were repaired using BBmap version 37.66 (9). Adapter sequences along with low-quality bases were removed using PRINSEQ version 0.20.4 (10). The high-quality Illumina and PacBio reads were *de novo* assembled using hybridSPAdes version 3.11.0 (11). The improved draft version of the assembly was generated by using Metassembler version 1.5 (12) by merging the draft monoteliosporic sequence-based assemblies from the DAOM 236416 and RAKB_UP_1 isolates with the improved and reassembled hybrid assembly. Genome assembly gap filling and polishing on the merged assembly were done by GapFiller version 1.10 (13, 14) and Pilon version 1.22 (15), respectively. The improved merged draft assembly consists of 787 scaffolds with

Received 7 January 2018 Accepted 20 March 2018 Published 17 May 2018

Citation Kumar A, Mishra P, Maurya R, Mishra AK, Gupta VK, Ramteke PW, Marla SS. 2018. Improved draft genome sequence of a monoteliosporic culture of the Karnal bunt (*Tilletia indica*) pathogen of wheat. Genome Announc 6:e00015-18. <https://doi.org/10.1128/genomeA.00015-18>.

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

Address correspondence to Anil Kumar, anilkumar.mbg@gmail.com, or Soma S. Marla, ssmarla@yahoo.com.

A.K., P.M., R.M., and S.S.M. contributed equally to this work.

a total genome size of 31,836,179 bp (N_{50} , 80,772 bp), with a GC content of 54.79% and an average coverage depth of at least $107\times$, which covers higher genome statistics (high N_{50} value, maximum contig length, and number of minimum contigs in scaffolding with respect to the high depth of coverage, which leads to an improved genome assembly) than the other published draft genome sequences of KB pathogen.

For comprehensive genome annotation, the improved draft scaffolds were first repeat masked using RepeatMasker version 4.0.7 (16), followed by genomic annotation using MAKER version 2.31.9 (17), AUGUSTUS (18–20), and SNAP (21). A total of 9,209 protein-coding gene models and 0.21% repeat elements were observed for the improved TiK isolate. The improved draft version was also screened for the presence of simple sequence repeat (SSR) loci using MISA (22), and 5,734 SSR loci were identified. The most abundant SSR type in the genome was trinucleotides, with 2,449 (42.71% of all SSRs). This is the first improved genome sequence of a monoteliosporic culture of the KB pathogen of wheat from the order Georgerfisheriales (Exobasidiomycetes).

The availability of a near-complete, more accurate, and nonredundant genome sequence serves as baseline data to provide ample opportunities to understand the pathogenic mechanisms as the model for the identification of the fungal pathogenic determinants involved in disease development, which will be used for devising effective crop protection strategies as part of the development of resistant wheat cultivars showing immunity against KB (23).

Accession number(s). This improved whole-genome shotgun project has been deposited at DDBJ/ENA/GenBank under the accession number [PKQB00000000](https://www.ncbi.nlm.nih.gov/nuccore/PKQB00000000). The version described in this report is version PKQB01000000 for TiK.

ACKNOWLEDGMENTS

We are grateful to the Bioinformatics Centre (Sub DIC) at G. B. Pant University of Agriculture and Technology, Pantnagar, India, for encouragement and providing the research facilities; we also acknowledge the Biotechnology Information System Network (BTISNet), Department of Biotechnology (DBT), Government of India, New Delhi, and Division of Genomic Resources (DGR), National Bureau of Plant Genetic Resources (NBPGR) for providing the necessary facilities.

This research was financially supported by the grant of Department of Biotechnology (DBT), Government of India, New Delhi, India, under sanction BT/PR14987/AGR/21/914/2015.

REFERENCES

- Mitra M. 1931. A new bunt on wheat in India. *Ann Appl Biol* 18:178–179. <https://doi.org/10.1111/j.1744-7348.1931.tb02294.x>.
- Kumar A, Singh US, Kumar J, Garg GK. 2008. Application of molecular and immuno-diagnostic tools for detection, surveillance and quarantine regulation of Karnal bunt (*Tilletia indica*) of wheat. *Food Agric Immunol* 19:293–311. <https://doi.org/10.1080/09540100802478194>.
- Kumar A, Pandey V, Singh M, Pandey D, Saharan MS, Marla SS. 2017. Draft genome sequence of Karnal bunt pathogen (*Tilletia indica*) of wheat provides insights into the pathogenic mechanisms of quarantined fungus. *PLoS One* 12:e0171323. <https://doi.org/10.1371/journal.pone.0171323>.
- Nguyen HD, Samba Siva P, Cullis J, Levesque CA, Hambleton S. 2016. Draft genome sequence of *Tilletia indica* and *Tilletia walkeri*. EMBL/GenBank/DDBJ databases (accession no. GCA_001645015).
- Aggarwal R, Gurjar MS, Jogawat A, Gogoi R, Basyal BM, Kulshreshtha D, Sharma S, Rao AR, Rai A. 2017. Data from GenBank (accession no. GCA_002220835).
- Kumar A, Singh US, Singh A, Malik VS, Garg GK. 2000. Molecular signaling in pathogenicity and host recognition in smut fungi taking Karnal bunt as a model system. *Indian J Exp Biol* 38:525–539.
- Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>.
- University of Michigan. 2009. fastQValidator (v0.1.1). <https://github.com/statgen/fastQValidator>.
- Bushnell B. 2015. BMap v35.66. <https://sourceforge.net/projects/bbmap>.
- Schmieder R. 2010. PRINSEQ v0.20.4. <https://sourceforge.net/projects/prinseq>.
- 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>.
- Wence AH, Schatz MC. 2015. Metassembler: merging and optimizing *de novo* genome assemblies. *Genome Biol* 16:207. <https://doi.org/10.1186/s13059-015-0764-4>.
- Nadalin F, Vezzi F, Policriti A. 2012. GapFiller: a *de novo* assembly approach to fill the gap within paired reads. *BMC Bioinformatics* 13(Suppl 14):S8. <https://doi.org/10.1186/1471-2105-13-S14-S8>.
- Boetzer M, Pirovano W. 2012. Toward almost closed genomes with GapFiller. *Genome Biol* 13:R56. <https://doi.org/10.1186/gb-2012-13-6-r56>.
- Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9:e112963. <https://doi.org/10.1371/journal.pone.0112963>.
- Smit AFA, Hubley R, Green P. 2017. RepeatMasker open 4.0. <http://www.repeatmasker.org>.
- Cantarel BL, Korf I, Robb SMC, Parra G, Ross E, Moore B, Holt C, Sánchez Alvarado A, Yandell M. 2008. MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res* 18:188–196. <https://doi.org/10.1101/gr.6743907>.
- Stanke M, Schöffmann O, Morgenstern B, Waack S. 2006. Gene predic-

- tion in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC Bioinformatics* 7:62–73. <https://doi.org/10.1186/1471-2105-7-62>.
19. 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>.
 20. Stanke M, Steinkamp R, Waack S, Morgenstern B. 2004. AUGUSTUS: a Web server for gene finding in eukaryotes. *Nucleic Acids Res* 32:W309–W312. <https://doi.org/10.1093/nar/gkh379>.
 21. Korf I. 2004. Gene finding in novel genomes. *BMC Bioinformatics* 5:59. <https://doi.org/10.1186/1471-2105-5-59>.
 22. Thiel W, Michalek R, Varshney RK, Graner A. 2003. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theor Appl Genet* 106:411–422. <https://doi.org/10.1007/s00122-002-1031-0>.
 23. Kumar A, Singh A, Garg GK. 1998. Development of seed immunoblot binding assay for detection of Karnal bunt (*Tilletia indica*) of wheat. *J Plant Biochem Biotechnol* 7:119–120. <https://doi.org/10.1007/BF03263048>.