Draft Genome Sequence of the Oomycete *Globisporangium splendens* Strain rgcb-1

Rani Jalajakumari Reghu,[a] Biju Vadakkemukadiyil Chellappan,[a] Shenu Hudson Beena,[a] Anu Sasi,[a] Soniya Eppurathu Vasudevan,[b] Achuthsankar Sukumaran Nair[a]

[a]Department of Computational Biology and Bioinformatics, University of Kerala, Kariavattom Campus, Trivandrum, Kerala, India
[b]Rajiv Gandhi Centre for Biotechnology, Trivandrum, Kerala, India

ABSTRACT *Globisporangium splendens* (formerly *Pythium splendens*) is an oomycete pathogen of many economically important vegetable crops. Here, we present the first draft genome of *P. splendens*, which comprises 197 scaffolds with a total length of 53.3 Mb and 17,350 predicted protein-coding genes.

Pythium is a genus of oomycetes, comprising over 355 species with a worldwide distribution (1). Many *Pythium* species are destructive pathogens, especially of monocotyledonous plants, whereas some species can cause diseases in fish, red algae, and mammals, including humans. *Globisporangium splendens* (formerly *Pythium splendens*) is an important plant pathogen that causes rots of stems and roots and pre- and postemergence seedling damping off in many plants, including (but not limited to) black pepper, maize, muskmelon, ginger, and oil palm (2–6). Although the whole genomes of many *Pythium* species are available, the whole genome of *Pythium splendens* is completely absent, which limits the understanding of its virulence mechanism. Here, we present the first draft genome sequence of *P. splendens*, which infects black pepper from Kerala, India. We hope that this genome sequence will be a valuable resource for gaining a better understanding of the mechanism underlying the pathogenicity of this destructive pathogen.

P. splendens culture was obtained from Kerala Agricultural University (Thiruvananthapuram, India). The isolate was cultured on potato-dextrose agar plates, and DNA was extracted from mycelia using the DNeasy plant minikit (Qiagen). Illumina paired-end (69 million reads) and PacBio single-molecule real-time (SMRT) (0.7 million reads) libraries were sequenced using the Illumina HiSeq 2500 and PacBio Sequel platforms, respectively, according to the instruction manual for each platform. Cutadapt (v2.3) was used with default parameters to remove adapter sequences, primers, poly(A) tails, and low-quality sequences (7). In total, 6.9 Gb and 10.2 Gb of Illumina and PacBio data, respectively, were obtained. MaSuRCA (v2.2.1) was used to build a hybrid assembly using both Illumina and PacBio data (8). The total length of the final assembly was 53.3 Mb, distributed in 198 scaffolds with a maximum scaffold length of 1.7 Mb. QUAST was used to assess the assembly scaffold quality (N_{50}, 342.1 kb; N_{75}, 237.8 kb; L_{50}, 44; L_{75}, 77; longest scaffold, 1.7 Mb; number of scaffolds of >50 kb, 180; number of scaffolds of >10 kb, 195) (9). The total GC content of the assembly was 52.6%. The Benchmarking Universal Single-Copy Orthologs (BUSCO) (v3.0.1) program identified 201 of 215 conserved protist genes from protists_ensembl (http://busco.ezlab.org), reflecting the fact that the *P. splendens* draft genome is 93.5% complete. A combination of *de novo* and homology-based approaches was used to mask 40% of the repetitive sequences in the genome. The protein-coding genes were predicted with the AUGUSTUS program implemented in Blast2Go (v5.1.13) software, using the previously sequenced transcriptome data of *P. splendens* as hints, with the following parameters: map

threshold, 30; minimum read alignment, 11; minimum intron length, 32; minimum exon length, 300; depth coverage, 20; this resulted in the prediction of 17,350 protein-coding genes (10, 11). Functional annotation using Blast2Go (v5.1.13) software revealed that 10,145 proteins had homologs in the NCBI databases. SECRETOOL analysis predicted 713 secretory proteins, of which 88 were found to possess signatures corresponding to RXLR effectors (12, 13) and 1,760 had Pfam domains (14).

Data availability. Both Illumina and PacBio raw reads have been submitted to the NCBI SRA under BioProject accession number PRJNA548776. The whole-genome shotgun project and annotation file have been deposited in DDBJ/ENA/GenBank under accession number VFIW00000000 (BioSample accession number SAMN12049464).

REFERENCES

1. Ho HH. 2009. The genus *Pythium* in Taiwan, China (1): a synoptic review. Front Biol China 4:15–28. https://doi.org/10.1007/s11515-009-0009-6.
2. Matsuda A, Moya Franco JD, González JL, Watanabe T. 1998. Root rots of black pepper caused by *Pythium splendens* in the Dominican Republic. Ann Phytopathol Soc Jpn 64:303–306. https://doi.org/10.3186/jjphytopath.64.303.
3. Al-Sa'di AM, Deadman ML, Al-Said FA, Khan I, Al-Azri M, Drenth A, Aitken E. 2008. First report of *Pythium splendens* associated with severe wilt of muskmelon (*Cucumis melo*) in Oman. Plant Dis 92:313. https://doi.org/10.1094/PDIS-92-2-0313C.
4. Dorrance AE, Berry SA, Bowen P, Lipps PE. 2004. Characterization of *Pythium* spp. from three Ohio fields for pathogenicity on corn and soybean and metalaxyl sensitivity. Plant Health Prog 5:1. https://doi.org/10.1094/PHP-2004-0202-01-RS.
5. Le DP, Smith MK, Aitken E. 2016. An assessment of *Pythium* spp. associated with soft rot disease of ginger (*Zingiber officinale*) in Queensland, Australia. Australas Plant Pathol 45:377–387. https://doi.org/10.1007/s13313-016-0424-5.
6. Aderungboye FO, Esuruoso OF. 1976. Ecological studies on *Pythium splendens* braun in oil palm (*Elaeis guineensis* Jacq.) plantation soils. Plant Soil 44:397–406. https://doi.org/10.1007/BF00015891.
7. Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J 17:10. https://doi.org/10.14806/ej.17.1.200.
8. 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.
9. Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics 29:1072–1075. https://doi.org/10.1093/bioinformatics/btt086.
10. 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.
11. Zhu Y, Zhou P, Hu J, Zhang R, Ren L, Li M, Ning F, Chen W, Yu L. 2013. Characterization of *Pythium* transcriptome and gene expression analysis at different stages of fermentation. PLoS One 8:e65552. https://doi.org/10.1371/journal.pone.0065552.
12. Cortázar AR, Aransay AM, Alfaro M, Oguiza JA, Lavín JL. 2014. SECRETOOL: integrated secretome analysis tool for fungi. Amino Acids 46:471–473. https://doi.org/10.1007/s00726-013-1649-z.
13. McGowan J, Fitzpatrick DA. 2017. Genomic, network, and phylogenetic analysis of the oomycete effector arsenal. mSphere 2:e00408-17. https://doi.org/10.1128/mSphere.00408-17.
14. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J, Sonnhammer EL, Tate J, Punta M. 2014. Pfam: the protein families database. Nucleic Acids Res 42:D222–D230. https://doi.org/10.1093/nar/gkt1223.