AMERICAN SOCIETY FOR MICROBIOLOGY | genomeAnnouncements™

EUKARYOTES

Draft Genome Sequence of Sorghum Grain Mold Fungus *Epicoccum sorghinum*, a Producer of Tenuazonic Acid

Rodrigo C. Oliveira,[a] Karen W. Davenport,[b] Blake Hovde,[b] Danielle Silva,[a] Patrick S. G. Chain,[b] Benedito Correa,[a] Debora F. Rodrigues[c]

Department of Microbiology, University of São Paulo, São Paulo, Brazil[a]; Biosciences Division, Los Alamos National Laboratory, Los Alamos, New Mexico, USA[b]; Civil and Environmental Engineering, University of Houston, Houston, Texas, USA[c]

ABSTRACT The facultative plant pathogen *Epicoccum sorghinum* is associated with grain mold of sorghum and produces the mycotoxin tenuazonic acid. This fungus can have serious economic impact on sorghum production. Here, we report the draft genome sequence of *E. sorghinum* (USPMTOX48).

Received 7 November 2016 **Accepted** 22 November 2016 **Published** 26 January 2017

Citation Oliveira RC, Davenport KW, Hovde B, Silva D, Chain PSG, Correa B, Rodrigues DF. 2017. Draft genome sequence of sorghum grain mold fungus *Epicoccum sorghinum*, a producer of tenuazonic acid. Genome Announc 5:e01495-16. https://doi.org/10.1128/genomeA.01495-16.

Address correspondence to Debora F. Rodrigues, dfrigirodrigues@uh.edu.

Epicoccum sorghinum (Sacc.) (also known as *Phoma sorghina*) (1) is one of the most important fungi in the grain-mold complex in sorghum (2). The presence of this pathogen in sorghum results in significant economic losses due to reduced crop yields, seed viability, and kernel weight (3). This fungus produces tenuazonic acid (TA), a mycotoxin that produces acute toxicity to organisms and therefore prevents the consumption of sorghum grains as food and feed (4, 5). The draft genome of this fungus has genes involved in the TA pathway.

To begin to access the genetic mechanism of tenuazonic acid production in *E. sorghinum*, we report the draft genome sequence of *Epicoccum sorghinum* strain USPMTOX48, which was recovered from contaminated sorghum grains (*Sorghum bicolor* [L.] Moench cv. DKB 550) cultured in Votuporanga, Brazil, in 2013. A polyphasic approach consisting of molecular and morphological characterization was performed for species identification (1). For sequencing analysis, genomic DNA was extracted using the Easy-DNA kit (Invitrogen, USA) and used to generate a short-insert paired-end library on an Illumina HiSeq 2000 instrument.

The library generated 58,194,228 reads (read lengths, 101 bp) totaling 5,878 Mbp (176× genome coverage). Raw data underwent quality control using FaQCs, which trimmed and filtered the reads (6, 7). The resulting data were assembled with IDBA_UD (8) and Velvet (9). Consensus sequences of both assemblies were computationally shredded and merged with Phrap (10, 11). The genome assembly consisted of 391 contigs (>1 kb); the estimated genome size is 33.4 Mbp, and the G+C content is 52%.

Gene annotation was carried out using the MAKER2 training and annotation pipeline (12). Briefly, repeated genomic regions were masked using RepeatMasker (http://www.repeatmasker.org). Genes were then modeled by combining several gene annotations methods as inputs into MAKER2, namely: (i) BLASTx alignment of proteins of a related species, *Phoma tracheiphila*; (ii) Augustus (13) *ab initio* gene models trained on the gene structures of the fungal Benchmarking Universal Single-Copy Orthologs (BUSCO) (14); (iii) SNAP (15) *ab initio* models trained on Hidden Markov Models of the CEGMA core eukaryotic genes (16); and (iv) Genemark-ES *ab initio* gene models (17). BUSCO quality analysis of the output gene annotations resulted in a high-quality gene annotation. The output resulted in MAKER calling a total of 9,495 genes. The average gene length, the mean exon length, and the mean intron length were determined to

be 1,658 bp, 574 bp, and 84 bp, respectively. Functional annotation of the 9,495 genes was performed by InterProScan 5 (18) and BLASTp searches against the UniProt (UniProt Consortium) protein blast database.

In agreement with the capability of *E. sorghinum* to produce TA, we found an identical domain of the TA biosynthetic gene described from *Magnaporthe oryzae* genome (19). TAS1 is a nonribosomal peptide synthetase (NRPS)-polyketide synthase (PKS) hybrid enzyme with a C-A-PCP-KS domain organization. The TAS1 identified in the *E. sorghinum* genome was highly conserved. This genomic information will contribute to a better understanding of the TA biosynthetic pathways and its regulatory mechanisms.

Accession number(s). The draft genome sequence of *Epicoccum sorghinum* (USPMTOX48) has been deposited at DDBJ/EMBL/GenBank under the accession no. MIEO00000000. The version described in this paper is version MIEO01000000.

ACKNOWLEDGMENTS

This work was supported by the Research Support Foundation of the State of São Paulo (FAPESP) and CAPES (project 006869/2015-07).

REFERENCES

1. Aveskamp MM, de Gruyter J, Woudenberg JHC, Verkley GJM, Crous PW. 2010. Highlights of the *Didymellaceae*: a polyphasic approach to characterise *Phoma* and related pleosporalean genera. Stud Mycol 65:1–60. https://doi.org/10.3114/sim.2010.65.01.
2. Navi SS, Bandyopadhyay R, Reddy RK, Thakur RP, Yang XB. 2005. Effects of wetness duration and grain development stages on sorghum grain mold infection. Plant Dis 89:872–878. https://doi.org/10.1094/PD-89-0872.
3. Forbes GA, Bandyopadhyay R, Garcia G. 1992, A review of sorghum grain mold. ICRISAT, India.
4. Shephard GS, Thiel PG, Sydenham EW, Vleggaar R, Marasas WF. 1991. Reversed-phase high-performance liquid chromatography of tenuazonic acid and related tetramic acids. J Chromatogr 566:195–205. https://doi.org/10.1016/0378-4347(91)80124-U.
5. Codex Committee on Contaminants in Foods. 2012. Discussion paper on fungi and mycotoxins in sorghum. Food and Agriculture Organization of the United Nations, Maastricht, The Netherlands.
6. Bennett S. 2004. Solexa Ltd. Pharmacogenomics 5:433–438. https://doi.org/10.1517/14622416.5.4.433.
7. Lo CC, Chain PSG. 2014. Rapid evaluation and quality control of next generation sequencing data with FaQCs. BMC Bioinformatics 15:366. https://doi.org/10.1186/s12859-014-0366-2.
8. Peng Y, Leung HCM, Yiu SM, Chin FYL. 2012. IDBA-UD: a *de novo* assembler for single-cell and metagenomic sequencing data with highly uneven depth. Bioinformatics 28:1420–1428. https://doi.org/10.1093/bioinformatics/bts174.
9. Zerbino DR, Birney E. 2008. Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. Genome Res 18:821–829. https://doi.org/10.1101/gr.074492.107.
10. Ewing B, Green P. 1998. Base-calling of automated sequencer traces using Phred. II. Error probabilities. Genome Res 8:186–194. https://doi.org/10.1101/gr.8.3.186.
11. Ewing B, Hillier L, Wendl MC, Green P. 1998. Base-calling of automated sequencer traces using Phred. I. Accuracy assessment. Genome Res 8:175–185.
12. Holt C, Yandell M. 2011. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. BMC Bioinformatics 12:491. https://doi.org/10.1186/1471-2105-12-491.
13. Stanke M, Schöffmann O, Morgenstern B, Waack S. 2006. Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. BMC Bioinformatics 7:62. https://doi.org/10.1186/1471-2105-7-62.
14. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31:3210–3212. https://doi.org/10.1093/bioinformatics/btv351.
15. Korf I. 2004. Gene finding in novel genomes. BMC Bioinformatics 5:59. https://doi.org/10.1186/1471-2105-5-59.
16. Parra G, Bradnam K, Korf I. 2007. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. Bioinformatics 23:1061–1067. https://doi.org/10.1093/bioinformatics/btm071.
17. Ter-Hovhannisyan V, Lomsadze A, Chernoff YO, Borodovsky M. 2008. Gene prediction in novel fungal genomes using an *ab initio* algorithm with unsupervised training. Genome Res 18:1979–1990. https://doi.org/10.1101/gr.081612.108.
18. Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong SY, Lopez R, Hunter S. 2014. InterProScan 5: genome-scale protein function classification. Bioinformatics 30:1236–1240. https://doi.org/10.1093/bioinformatics/btu031.
19. Yun CS, Motoyama T, Osada H. 2015. Biosynthesis of the mycotoxin tenuazonic acid by a fungal NRPS-PKS hybrid enzyme. Nat Commun 6:8758. https://doi.org/10.1038/ncomms9758.