Revised Transcriptome-Based Gene Annotation for *Aspergillus flavus* Strain NRRL 3357

E. Anne Hatmaker, Xiaofan Zhou, Matthew E. Mead, Heungyun Moon, Jae-Hyuk Yu, Antonis Rokas

**ABSTRACT**  *Aspergillus flavus* is an agriculturally and medically important filamentous fungus that produces mycotoxins, including aflatoxins, which are potent carcinogens. Here, we generated short- and long-read transcript sequence data from the growth of *A. flavus* strain NRRL 3357 under both typical and stress conditions to produce a new annotation of its genome.
Totals of 27,714,710 (library from mix of six samples), 29,581,787 (mycelia grown at 30°C), and 24,719,957 (mycelia grown at 37°C) raw reads were generated. An additional Iso-Seq library was prepared from the mix of the six samples, size selected for fragments of \(4\) kb using the BluePippin size selection system (Sage Science), and analyzed on the RS II instrument (PacBio Biosciences), resulting in 5,173,949 reads with a mean length of 2,785 bp. Short and long reads were used together to generate a de novo transcriptome assembly using Trinity v.2.10.0 (4). As part of the Trinity assembly pipeline, low-quality bases and adaptor contamination were trimmed from the Illumina reads using Trimmomatic v.0.39.0 (5). Short reads were also used for genome-guided assembly using Trinity.

Initial gene models were generated using default parameters for MAKER v.2.31.9 (6), which included three \textit{ab initio} gene predictors, AUGUSTUS v.3.3.3 (7), GeneMark-ES v.4.39 (8), and SNAP v.2013-11-29 (9). Exonerate v.2.4.0 (10) was provided for alignment to the MAKER pipeline. Isoforms were added to PASA v.2.3.3 (11) from both the de novo and genome-guided transcriptome assemblies as “EST evidence.” The new annotation predicts 14,313 genes and 14,699 proteins. We compared our annotation to the \textit{Aspergillus flavus} NRRL 3357 genome annotations from AspGD and NCBI (accession number GCA_000006275.2) using BUSCO v.3.0.2 (12), which estimated a copy number of 4,046 preselected genes considered single copy within the OrthoDB (13) database (Table 1). The new annotation contains 96.3% of these 4,046 genes, substantially higher than the percentages contained by the previous two annotations (AspGD, 90.5%; NCBI, 88.0%; Table 1). The transcriptome-based annotation contains 12,569 genes which overlap one or more gene models in the previous NCBI annotation, as well as 1,744 predicted genes without any overlap with the previous annotation, suggesting that our transcriptome-based annotation provides new information for studying \textit{A. flavus}.

\textbf{Data availability.} \textit{A. flavus} NRRL 3357 RNA sequences from this study have been deposited under Sequence Read Archive (SRA) study SRP213830, with individual samples deposited in the SRA under accession numbers SRX6419472, SRX6419473, SRX6419474, and SRX6419475 and BioSample accession numbers SAMN12235769 through SAMN12235771. This whole-genome sequencing project has been deposited at DDBJ/ENA/GenBank under the accession number AAIH00000000. The annotation has been updated for \textit{A. flavus} NRRL 3557 assembly accession number GCA_000006275.3 (GenBank), and the annotation described in this paper is the third version, AAIH03000000.

\textbf{ACKNOWLEDGMENT}

We thank William C. Nierman for giving us permission to update the gene annotation of \textit{Aspergillus flavus} NRRL 3357.

\textbf{REFERENCES}

1. Klich MA. 2007. \textit{Aspergillus flavus}: the major producer of aflatoxin. Mol Plant Pathol 8:713–722. https://doi.org/10.1111/j.1364-3703.2007.00436.x.

2. Cunha C, Aversa F, Romani L, Carvalho A. 2013. Human genetic susceptibility to invasive aspergillosis. PLoS Pathog 9:e1003434. https://doi.org/10.1371/journal.ppat.1003434.

3. Amaike S, Keller NP. 2011. \textit{Aspergillus flavus}. Annu Rev Phytopathol 49:107–133. https://doi.org/10.1146/annurev-phyto-072910-095221.
4. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A. 2011. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. Nat Biotechnol 29:644.

5. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120.

6. Holt C, Yandell M. 2011. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. BMC Bioinformatics 12:491.

7. 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.

8. 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.

9. Korf I. 2004. Gene finding in novel genomes. BMC Bioinformatics 5:59.

10. Slater GSC, Birney E. 2005. Automated generation of heuristics for biological sequence comparison. BMC Bioinformatics 6:31.

11. Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, White O, Robin CR, Wortman JR. 2008. Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. Genome Biol 9:R7.

12. 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.

13. Zdobnov EM, Tegenfeldt F, Kuznetsov D, Waterhouse RM, Simao FA, Ioannidis P, Seppey M, Loetscher A, Kriventseva EV. 2017. OrthoDB v9.1: cataloging evolutionary and functional annotations for animal, fungal, plant, archaeal, bacterial and viral orthologs. Nucleic Acids Res 45:D744–D749.

Microbiology Resource Announcement

December 2020 Volume 9 Issue 49 e01155-20 mra.asm.org