

Genome annotation of *Aspergillus melleus* strain CBS 546.65

Jason E. Stajich,^{1,2} Sean M. Garvey,³ J  ssica Gil-Serna⁴

AUTHOR AFFILIATIONS See affiliation list on p. 2.

ABSTRACT The fungus *Aspergillus melleus* is an important biosynthesis host for varied commercial applications. Gene annotation of a previously published genome produced 12,841 protein-coding genes and identified 102 biosynthetic gene clusters.

KEYWORDS *Aspergillus*, gene prediction, genome annotation

Aspergillus melleus is an industrially important fungus that can synthesize the insecticide and nematocide compounds aspyrone (1) and mellamide (2), mycotoxins including ochratoxin A (3, 4), and antibiotic compounds (5). Secreted proteases obtained from *A. melleus* have also been developed as nutraceuticals to support immunity or digestive health (6–8). The neotype strain for the species was isolated from soil collected near Allahabad, India, in 1965 by B.S. Mehrotra and deposited by K.B. Raper and D.I. Fennell and deposited in the NRRL collection as NRRL 5103 (USDA-ARS Collection), American Type Culture Collection (ATCC) 16889, and CBS 546.65 in the Westerdijk Institute CBS collection (Fig. 1). The complete genome assembly was previously produced from a hybrid assembly of PacBio and Illumina DNA sequences but was not annotated (4).

Gene annotation was performed with Funannotate v1.8.8 (9) using a combination of RNA-seq and homologous protein sequences to inform gene prediction. The pipeline steps are archived in zenodo (10). First, the genome was masked for repetitive sequences and transposons by developing a *de novo* repeat library with RepeatModeler v2.0.1 (11) using default parameters and the “-LTRstruct” option. This library was used to lowercase mask repetitive genome regions by running RepeatMasker v4-1-1 (12) with Funannotate “mask.” Next, the Funannotate “train” step ran Trinity-GG v2.11.0 (13) and PASA v2.4.1 (14) to assemble transcripts from RNA-Seq alignments of *Aspergillus sesamicola*, a closely related species. Default Funannotate parameters were used to run Trimmomatic v0.39 (15), Hisat v2.2.1 (16), Kallisto v0.46.2 (17), Trinity-GG, and PASA to produce an assembly of 33,620 transcripts. The Funannotate “train” command trained gene prediction parameters for Augustus v3.3.3 (18), GlimmerHMM v3.0.4 (19), and SNAP v2013_11_29 (20) using the high-quality full-length PASA gene models. GeneMarkES-ET v4.62 (21) parameters were generated by its self-training protocol. Trinity-GG transcripts and Swissprot database protein sequences that matched the genome by diamond v2.0.13.151 BLASTX (22, 23) were aligned with the splice-site aware tool exonate v2.4.0 (24) and provided as exon evidence to *ab initio* gene predictors. CodingQuarry v2.0 (25) was run with the RNASeq alignment file from the “train” step. The gene predictions from all tools were combined to produce a composite set of 12,974 protein-coding genes by Evidence Modeler v1.1.1 (14). The predicted genes were filtered to remove proteins less than 50 amino acids in length or those with matches to transposases by diamond BLASTP of the Funannotate repeat library to produce a final set of 12,841 genes. The tool tRNAscan v1.3.1 (26) was run and predicted 213 non-overlapping tRNA gene models. The protein coding gene set was refined further with the Funannotate “update” step to add 5′- and 3′-untranslated region (UTR) exons to the gene models and predict alternatively spliced transcripts based on the Trinity transcripts and two rounds of PASA. The update

Editor Leighton Pritchard, University of Strathclyde, Glasgow, United Kingdom

Address correspondence to Jason E. Stajich, jason.stajich@ucr.edu.

The authors declare no conflict of interest.

See the funding table on p. 3.

Received 19 May 2025

Accepted 29 July 2025

Published 25 August 2025

Copyright   2025 Stajich et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).



FIG 1 *Aspergillus melleus* CBS 546.65 colony after incubation for 7 days on potato dextrose agar.

step added 19 new gene models and 2,444 UTRs to produce a final data set of 13,061 genes and 13,395 transcripts (including the tRNA genes). Few genes had more than one isoform predicted (256), and there were 64 genes with at least three and 11 with four isoforms. To assign putative protein function, sequence similarity to UniProt v2021_02 (22), InterProScan v5.51-85.0 (27), EggNog v1.0.3 (28, 29), dbCAN v9.0 (30), and MEROPS v12.0 (37) databases was computed and associated with the protein sequences and CDS records. Prediction of biosynthetic gene clusters by antiSMASH v5.2.0 (31) identified 102 clusters, including 30 Type I polyketide synthases (PKS), 1 Type III PKS, 21 non-ribosomal peptide synthases (NRPS), 26 NRPS-like, 15 terpene, 6 indole, and 1 beta-lactone.

ACKNOWLEDGMENTS

J.E.S. is a CIFAR Fellow in the program Fungal Kingdom: Threats and Opportunities and was supported by US Department of Agriculture-National Institute of Food and Agriculture Hatch project CA-R-PPA-211-5062-H, National Science Foundation (NSF) award EF-2125066, and NIH award 2R01AI130128. Data analyses were performed on the High-Performance Computing Cluster at the University of California-Riverside in the Institute of Integrative Genome Biology supported by NSF DBI-1429826 and NIH S10-OD016290. Funding provided by BIO-CAT, Inc. to J.E.S. supported annotation and defrayed publication costs.

AUTHOR AFFILIATIONS

¹Department of Microbiology and Plant Pathology, University of California, Riverside, California, USA

²Institute for Integrative Genome Biology, University of California, Riverside, California, USA

³Department of Research and Development, BIO-CAT, Inc., Troy, Virginia, USA

⁴Department of Genetics, Physiology and Microbiology, Faculty of Biological Sciences, Complutense University of Madrid, Madrid, Spain

PRESENT ADDRESS

Sean M. Garvey, SAPIOME LLC, Jericho, Vermont, USA

AUTHOR ORCID*s*

Jason E. Stajich  <http://orcid.org/0000-0002-7591-0020>

Sean M. Garvey  <http://orcid.org/0000-0003-1013-7871>

Jéssica Gil-Serna  <http://orcid.org/0000-0002-2113-1830>

FUNDING

Funder	Grant(s)	Author(s)
Canadian Institute for Advanced Research		Jason E. Stajich
National Institute of Food and Agriculture	CA-R-PPA-211-5062-H	Jason E. Stajich
National Science Foundation	EF-2125066,DBI-1429826	Jason E. Stajich
National Institute of Allergy and Infectious Diseases	2R01AI130128,S10-OD016290	Jason E. Stajich

AUTHOR CONTRIBUTIONS

Jason E. Stajich, Data curation, Formal analysis, Project administration, Software, Writing – original draft, Writing – review and editing | Sean M. Garvey, Conceptualization, Funding acquisition, Writing – review and editing | Jéssica Gil-Serna, Data curation, Methodology, Visualization, Writing – review and editing

DATA AVAILABILITY

This Whole Genome Shotgun project had been published and deposited previously at DDBJ/ENA/GenBank as accession [JADPPX000000000](https://www.ncbi.nlm.nih.gov/assembly/JADPPX000000000/). The record was updated from its initial deposit to include the gene predictions with permission of original authors. The RNA-Seq reads from *A. sesamicola* are associated with BioProject [PRJNA585261](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA585261) and SRA accessions [SRR12142522](https://www.ncbi.nlm.nih.gov/sra/SRR12142522) and [SRR12142523](https://www.ncbi.nlm.nih.gov/sra/SRR12142523). Annotation pipeline, antiSMASH results and logfiles are available in github https://github.com/stajichlab/Annotation_Aspergillus_melleus and archived at zenodo ([10.5281/zenodo.15460055](https://doi.org/10.5281/zenodo.15460055)).

REFERENCES

- Kimura Y, Nakahara S, Fujioka S. 1996. Aspyrone, a nematocidal compound isolated from the fungus, *Aspergillus melleus*. Biosci Biotechnol Biochem 60:1375–1376. <https://doi.org/10.1271/bbb.60.1375>
- Ondeyka JG, Dombrowski AW, Polishook JP, Felcetto T, Shoop WL, Guan Z, Singh SB. 2003. Isolation and insecticidal activity of mellamide from *Aspergillus melleus*. J Ind Microbiol Biotechnol 30:220–224. <https://doi.org/10.1007/s10295-003-0048-y>
- Lai M, Semenik G, Hesselstine CW. 1970. Conditions for production of ochratoxin A by *Aspergillus* species in a synthetic medium. Appl Microbiol 19:542–544. <https://doi.org/10.1128/am.19.3.542-544.1970>
- Gil-Serna J, Vázquez C, Patiño B. 2020. The genomic regions that contain ochratoxin A biosynthetic genes widely differ in *Aspergillus* section *Circumdati* species. Toxins (Basel) 12:754. <https://doi.org/10.3390/toxins12120754>
- Burton HS. 1950. Antibiotics from *Aspergillus melleus*. Nature 165:274. <https://doi.org/10.1038/165274a0>
- Moretti M, Bertoli E, Bulgarelli S, Testoni C, Guffanti EE, Marchioni CF, Braga PC. 1993. Effects of seaprose on sputum biochemical components in chronic bronchitic patients: a double-blind study vs placebo. Int J Clin Pharmacol Res 13:275–280.
- Paulussen KJM, Askow AT, Deutz MT, McKenna CF, Garvey SM, Guice JL, Kesler RM, Barnes TM, Tinker KM, Paluska SA, Ulanov AV, Bauer LL, Dilger RN, Burd NA. 2024. Acute microbial protease supplementation increases net postprandial plasma amino acid concentrations after pea protein ingestion in healthy adults: a randomized, double-blind, placebo-controlled trial. J Nutr 154:1549–1560. <https://doi.org/10.1016/j.tjnut.2024.03.009>
- Zhou Z, Liu Y, Ishigaki Y, Yamaguchi S, Chen J, Liu X. 2025. Microbial protease supplementation improves gastric emptying and protein digestive fate of beef for the elderly under dynamic *in vitro* digestion. Food Res Int 202:115721. <https://doi.org/10.1016/j.foodres.2025.115721>
- Palmer JM, Stajich JE. 2023. Funannotate: eukaryotic genome annotation (v1.8.14). Zenodo. <https://dx.doi.org/10.5281/zenodo.4054262>
- Stajich JE. 2025. Pipeline for annotation of *Aspergillus melleus*. Zenodo. <https://zenodo.org/doi/10.5281/zenodo.15460055>
- Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, Smit AF. 2020. RepeatModeler2 for automated genomic discovery of transposable element families. Proc Natl Acad Sci USA 117:9451–9457. <https://doi.org/10.1073/pnas.1921046117>

12. Smit AFA, Hubley R, Green P. 2013. RepeatMasker Open-4.0 (4-1-1). <http://www.repeatmasker.org>.
13. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, MacManes MD, Ott M, Orvis J, Pochet N, Strozzi F, Weeks N, Westerman R, William T, Dewey CN, Henschel R, LeDuc RD, Friedman N, Regev A. 2013. *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc* 8:1494–1512. <https://doi.org/10.1038/nprot.2013.084>
14. Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, White O, Buell CR, Wortman JR. 2008. Automated eukaryotic gene structure annotation using EVIDENCEModeler and the program to assemble spliced alignments. *Genome Biol* 9:R7. <https://doi.org/10.1186/gb-2008-9-1-r7>
15. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>
16. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. 2019. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol* 37:907–915. <https://doi.org/10.1038/s41587-019-0201-4>
17. Bray NL, Pimentel H, Melsted P, Pachter L. 2016. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol* 34:525–527. <https://doi.org/10.1038/nbt.3519>
18. Stanke M, Keller O, Gunduz I, Hayes A, Waack S, Morgenstern B. 2006. AUGUSTUS: *ab initio* prediction of alternative transcripts. *Nucleic Acids Res* 34:W435–9. <https://doi.org/10.1093/nar/gkl200>
19. Majoros WH, Pertea M, Salzberg SL. 2004. TigrScan and GlimmerHMM: two open source *ab initio* eukaryotic gene-finders. *Bioinformatics* 20:2878–2879. <https://doi.org/10.1093/bioinformatics/bth315>
20. Korf I. 2004. Gene finding in novel genomes. *BMC Bioinformatics* 5:59. <https://doi.org/10.1186/1471-2105-5-59>
21. 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>
22. Bateman A, Martin M-J, Orchard S, Magrane M, Ahmad S, Alpi E, Bowler-Barnett EH, Britto R, Bye-A-Jee H, Cukura A, et al. 2023. UniProt: the universal protein knowledgebase in 2023. *Nucleic Acids Res* 51:D523–D531. <https://doi.org/10.1093/nar/gkac1052>
23. Buchfink B, Reuter K, Drost H-G. 2021. Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat Methods* 18:366–368. <https://doi.org/10.1038/s41592-021-01101-x>
24. Slater GSC, Birney E. 2005. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* 6:31. <https://doi.org/10.1186/1471-2105-6-31>
25. Testa AC, Hane JK, Ellwood SR, Oliver RP. 2015. CodingQuarry: highly accurate hidden Markov model gene prediction in fungal genomes using RNA-seq transcripts. *BMC Genomics* 16:170. <https://doi.org/10.1186/s12864-015-1344-4>
26. Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* 25:955–964. <https://doi.org/10.1093/nar/25.5.955>
27. Blum M, Chang H-Y, Chuguransky S, Grego T, Kandasamy S, Mitchell A, Nuka G, Paysan-Lafosse T, Qureshi M, Raj S, et al. 2021. The InterPro protein families and domains database: 20 years on. *Nucleic Acids Res* 49:D344–D354. <https://doi.org/10.1093/nar/gkaa977>
28. Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J. 2021. eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Mol Biol Evol* 38:5825–5829. <https://doi.org/10.1093/molbev/msab293>
29. Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H, Mende DR, Letunic I, Rattei T, Jensen LJ, von Mering C, Bork P. 2019. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res* 47:D309–D314. <https://doi.org/10.1093/nar/gky1085>
30. Huang L, Zhang H, Wu P, Entwistle S, Li X, Yohe T, Yi H, Yang Z, Yin Y. 2018. dbCAN-seq: a database of carbohydrate-active enzyme (CAZyme) sequence and annotation. *Nucleic Acids Res* 46:D516–D521. <https://doi.org/10.1093/nar/gkx894>
31. Blin K, Shaw S, Steinke K, Villebro R, Ziemert N, Lee SY, Medema MH, Weber T. 2019. antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Res* 47:W81–W87. <https://doi.org/10.1093/nar/gkz310>