AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

GENOME SEQUENCES

Draft Genome Sequence of an *Aspergillus* Strain Isolated from a Honey Bee Pupa

E. Anne Hatmaker,[a,b] Delaney L. Miller,[c] Irene Newton,[c] Antonis Rokas[a,b]

[a]Department of Biological Sciences, Vanderbilt University, Nashville, Tennessee, USA
[b]Vanderbilt Evolutionary Studies Initiative, Vanderbilt University, Nashville, Tennessee, USA
[c]Department of Biology, Indiana University, Bloomington, Indiana, USA

ABSTRACT Insect-associated fungi play an important role in wild and agricultural communities. We present a draft genome sequence of an entomopathogenic strain from the fungal genus *Aspergillus*, isolated from a honey bee pupa.

Fungi are a leading cause of insect disease (1) and negatively impact the productivity of an economically and ecologically important pollinator, the honey bee (2–4). The risk of fungal disease is greatest for honey bee brood (larvae and pupae). Several species of *Aspergillus* are opportunistic brood pathogens (5). *Aspergillus* species also infect plants (6) and humans (7).

We isolated a fungal strain (DMIN) from an infected pupa in Bloomington, IN, in June 2018. Inoculation of *in vitro*-reared bee brood with 10^3 fungal spores confirmed the pathogenic potential of the strain (8). Spores were collected with a sterile swab and struck to isolation on potato dextrose agar (PDA) at 34°C, ambient humidity, in incubation chambers under conditions favorable for microaerophile growth. Total DNA was extracted with the DNeasy blood and tissue kit (Qiagen, Hilden, Germany). The isolate ITS region was amplified using ITS4/ITS5 (9) and compared to the NCBI nonredundant/nucleotide (nr/nt) database using BLASTN (10); the region was most similar (>90%) to multiple *Aspergillus* species, with 100% query coverage.

Libraries were prepared using the TruSeq DNA PCR-free kit (Illumina, Inc.), and genomic DNA (gDNA) was sequenced at the Vanderbilt Technologies for Advanced Genomics facility on the Illumina NovaSeq 6000 platform, resulting in 150-bp paired-end reads. The reads were filtered for quality and adaptors removed using Trimmomatic v0.39 (11). The trimmed reads were assembled using SPAdes v3.15.3 (12). The mitochondrial genome was assembled using GetOrganelle v1.6.4 (13) and annotated using GeSeq v2.03 (14, 15). The nuclear genome was annotated using Liftoff v1.2.0 (16) with *Aspergillus flavus* NRRL 3357 as the reference genome (GenBank accession number GCA_000006275.3) (17). We evaluated the genome completeness using the BUSCO v4.0.4 Eurotiales database (18). The reads were mapped to assemblies using Bowtie2 v2.3.4.1 (19). The average nucleotide identity (ANI) was calculated using FastANI v0.1.3 (20). ANI comparisons were made between DMIN and four *Flavi* species: *Aspergillus arachidicola* CBS 117612 (GCA_009193545.1), *Aspergillus flavus* NRRL 3357 (CP044616 to CP044623), *Aspergillus parasiticus* SU-1 (GCA_000956085.1), and *Aspergillus minisclerotigenes* CBS 117635 (GCA_009176455.1). For all analyses, default parameters were used except where otherwise noted.

Sequencing resulted in 53,363,706 paired reads (42,289,484 after trimming). The assembly revealed a coculture, with both bacterial and fungal scaffolds. The fungal scaffolds were longer, had higher coverage (over 200×), and had a lower GC content (48%) than the bacterial scaffolds (coverage, 40×; GC content, 60%). Scaffolds under 500 bp were removed from the assembly, and scaffolds 500 bp and longer were compared to the NCBI BLASTN nonredundant/nucleotide (nr/nt) database. Scaffolds with

Editor Jason E. Stajich, University of California, Riverside

Address correspondence to Antonis Rokas, antonis.rokas@vanderbilt.edu.

The authors declare a conflict of interest.

Received 3 August 2022
Accepted 22 September 2022
Published 4 October 2022

TABLE 1 Genome sequencing and assembly statistics for the bacterial and fungal genomes from isolate DMIN

Species	Genome	No. of scaffolds	N_{50} (bp)	Coverage (×)	Size (bp)	No. of mapped reads (%)	No. of predicted proteins	GenBank accession no.
Aspergillus sp. DMIN	Nuclear	244	14	290	39,974,781	96.22	13,169	JALMFT000000000
	Mitochondrial	1	1	3,100	29,141	1.52	15	CM042804
Bombella sp. DMIN-2	Bacterial	21	4	45	2,057,252	0.77	1,883	JAMWFD000000000

high similarity (≥90%) to *Aspergillus* species and with ≥80% query coverage were retained to ensure that the assembly represented only fungal scaffolds. Of the 6,156 scaffolds removed, 5,281 were under 1,000 bp; the mean length of the removed scaffolds was 1,111.78 bp, and the median length was 641 bp.

ANI comparisons with *Aspergillus* section *Flavi* species showed that DMIN was most closely related to *Aspergillus parasiticus* (98% similarity). DMIN shared 96.5%, 94.2%, and 94% identity with *A. arachidicola*, *A. minisclerotigenes*, and *A. flavus*, respectively. The *Aspergillus* sp. strain DMIN nuclear genome contained 244 scaffolds; the mitochondrial genome was a single, circular chromosome (Table 1). Predicted proteins encompassed 93% of the BUSCO single-copy orthologs (3,911), with 5% (223) missing.

Reads that did not map to the fungal assembly were assembled independently to capture the genome of the bacterial contaminant and annotated using PGAP v6.0 (21). BLASTN comparisons of the bacterial scaffolds indicated high similarity (99%) to *Bombella* sp. strain ESL0368, which is related to bacteria that can provide protection against fungal infection in honey bee brood (8). The *Bombella* DMIN-2 assembly contained 21 scaffolds (Table 1).

Understanding *Aspergillus* pathogenicity is crucial to protecting insect health. The *Aspergillus* sp. DMIN genome provides valuable information about the genetic content of an entomopathogenic strain belonging to a widespread fungus genus.

Data availability. The metadata and complete assemblies are available at GenBank under BioProject accession number PRJNA815363. The raw sequence reads have been deposited in the Sequence Read Archive (SRA) under accession number SRR18306612. The *Aspergillus* sp. DMIN whole-genome shotgun sequencing project has been deposited at DDBJ/ENA/GenBank under the accession number JALMFT000000000. The version described in this paper is version JALMFT010000000. The *Bombella* sp. DMIN-2 whole-genome shotgun sequencing project has been deposited at DDBJ/ENA/GenBank under the accession number JAMWFD000000000. The version described in this paper is version JAMWFD000000000.

ACKNOWLEDGMENTS

Research in A.R.'s lab is supported by grants from the National Institutes of Health/National Institute of Allergy and Infectious Diseases (R56 AI146096 and R01 AI153356), the National Science Foundation (DEB-2110404), and the Burroughs Wellcome Fund. E.A.H. is supported by the National Institutes of Health/National Eye Institute (F31 EY033235).

A.R. is a scientific consultant for LifeMine Therapeutics, Inc.

REFERENCES

1. St Leger RJ, Wang C. 2010. Genetic engineering of fungal biocontrol agents to achieve greater efficacy against insect pests. Appl Microbiol Biotechnol 85:901–907. https://doi.org/10.1007/s00253-009-2306-z.
2. Klein A-M, Boreux V, Fornoff F, Mupepele A-C, Pufal G. 2018. Relevance of wild and managed bees for human well-being. Curr Opin Insect Sci 26:82–88. https://doi.org/10.1016/j.cois.2018.02.011.
3. Southwick EE, Southwick L, Jr. 1992. Estimating the economic value of honey bees (Hymenoptera: Apidae) as agricultural pollinators in the United States. J Econ Entomol 85:621–633. https://doi.org/10.1093/jee/85.3.621.
4. Brodschneider R, Gray A, Adjlane N, Ballis A, Brusbardis V, Charrière J-D, Chlebo R, Coffey MF, Dahle B, de Graaf DC, Maja Dražić M, Evans G, Fedoriak M, Forsythe I, Gregorc A, Grzęda U, Hetzroni A, Kauko L, Kristiansen P, Martikkala M, Martín-Hernández R, Aurelio Medina-Flores C, Mutinelli F, Raudmets A, A Ryzhikov V, Simon-Delso N, Stevanovic J, Uzunov A, Vejsnæs F, Wöhl S, Zammit-Mangion M, Danihlík J. 2018. Multi-country loss rates of honey bee colonies during winter 2016/2017 from the COLOSS survey. J Apic Res 57:452–457. https://doi.org/10.1080/00218839.2018.1460911.
5. Evison SEF, Jensen AB. 2018. The biology and prevalence of fungal diseases in managed and wild bees. Curr Opin Insect Sci 26:105–113. https://doi.org/10.1016/j.cois.2018.02.010.
6. St Leger RJ, Screen SE, Shams-Pirzadeh B. 2000. Lack of host specialization in *Aspergillus flavus*. Appl Environ Microbiol 66:320–324. https://doi.org/10.1128/AEM.66.1.320-324.2000.
7. Hedayati MT, Pasqualotto AC, Warn PA, Bowyer P, Denning DW. 2007. *Aspergillus flavus*: human pathogen, allergen and mycotoxin producer. Microbiology (Reading) 153:1677–1692. https://doi.org/10.1099/mic.0.2007/007641-0.

8. Miller DL, Smith EA, Newton ILG. 2021. A bacterial symbiont protects honey bees from fungal disease. mBio 12:e00503-21. https://doi.org/10.1128/mBio.00503-21.
9. White TJ, Bruns T, Lee S, Taylor J. 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics, p 315–322. *In* Innis MA, Gelfand DH, Sninsky JJ, White TJ (ed), PCR protocols: a guide to methods and applications. Academic Press, London, UK.
10. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. BMC Bioinformatics 10:421. https://doi.org/10.1186/1471-2105-10-421.
11. 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.
12. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477. https://doi.org/10.1089/cmb.2012.0021.
13. Jin J-J, Yu W-B, Yang J-B, Song Y, DePamphilis CW, Yi T-S, Li D-Z. 2020. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. Genome Biol 21:241. https://doi.org/10.1186/s13059-020-02154-5.
14. Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, Greiner S. 2017. GeSeq—versatile and accurate annotation of organelle genomes. Nucleic Acids Res 45:W6–W11. https://doi.org/10.1093/nar/gkx391.
15. Hugaboom M, Beck ML, Carrubba KH, Chennupati DV, Gupta A, Liu Q, Reddy MK, Roque F, Hatmaker EA. 2021. Complete mitochondrial genome sequences of nine *Aspergillus flavus* strains. Microbiol Resour Announc 10:e00971-21. https://doi.org/10.1128/MRA.00971-21.
16. Shumate A, Salzberg SL. 2021. Liftoff: accurate mapping of gene annotations. Bioinformatics 37:1639–1643. https://doi.org/10.1093/bioinformatics/btaa1016.
17. Hatmaker EA, Zhou X, Mead ME, Moon H, Yu J-H, Rokas A. 2020. Revised transcriptome-based gene annotation for *Aspergillus flavus* strain NRRL 3357. Microbiol Resour Announc 9:e01155-20. https://doi.org/10.1128/MRA.01155-20.
18. 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.
19. Langmead B, Salzberg SL. 2012. Bowtie2. Nat Methods 9:357–359. https://doi.org/10.1038/nmeth.1923.
20. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. 2018. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat Commun 9:5114. https://doi.org/10.1038/s41467-018-07641-9.
21. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J. 2016. NCBI Prokaryotic Genome Annotation Pipeline. Nucleic Acids Res 44:6614–6624. https://doi.org/10.1093/nar/gkw569.