AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

GENOME SEQUENCES

Complete Mitochondrial Genome Sequences of Nine *Aspergillus flavus* Strains

Miya Hugaboom,[a] Max L. Beck,[a] Katherine H. Carrubba,[a] D. Vignesh Chennupati,[a] Aryan Gupta,[a] Qi Liu,[a] Maya K. Reddy,[a] Fatima Roque,[a] E. Anne Hatmaker[a,b]

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

ABSTRACT Nuclear genome sequences incompletely characterize the genomic content and thus the genetic diversity of fungal species. Here, we present the complete mitochondrial genome sequences of nine *Aspergillus flavus* strains, providing useful information for inter- and intraspecific analyses.

The genus *Aspergillus* contains over 300 fungal species of varied industrial, agricultural, and medical relevance (1). *Aspergillus flavus* is a potent producer of aflatoxin B, a carcinogenic mycotoxin and a plant and opportunistic human pathogen (2). Despite the widespread availability of *A. flavus* genomic sequences, few assembled and annotated mitochondrial genome (mitogenome) sequences are available (3–5). Mitochondrial genes are linked to processes including metabolism, cell differentiation, and drug resistance (6, 7). The assembly and annotation of nine *A. flavus* strain mitogenome sequences from publicly available sequencing data provides valuable insight into the full genetic profile, evolution, and population genetics of *A. flavus*.

Genomic DNA was previously isolated and sequenced using an Illumina HiSeq 2500 paired-end 2 × 250-bp platform as described by Drott et al. (8). Paired-end reads from whole-genome sequencing of *A. flavus* were downloaded from NCBI's Sequence Read Archive (8), extracted, and split into forward and reverse FASTQ files using SRA Toolkit v2.9.6-1 (9). The reads were trimmed using Trimmomatic v.0.39 (10). Mitogenome sequences were assembled using the specialized genome assembler GetOrganelle v1.7.4.1 (11), with SPAdes v.3.12.0 (12) as the internal assembler. We used the GetOrganelle fungal database (-F fungus_mt) to identify, filter, and assemble target-associated reads with default parameters unless otherwise noted. The complete mitogenome sequence for *Aspergillus fumigatus* SGAir0713 (GenBank accession number CM016889.1) was used as a reference for the seed database (-s) for each assembly.

A single contig was generated from the GetOrganelle assembly for each strain. Circularization was accomplished via the identification of overlapping nucleotide sequences within the contig FASTA file and the subsequent manual trimming of redundant nucleotides within a text editor. The percentage of mitochondrial reads used for assembly ranged from 1.2 to 2.8% of the total trimmed reads. Read mapping to correct errors was carried out using Bowtie2 v2.3.4.1 (13) and SAMtools v1.6 (14). Bowtie2 was used to align the raw paired-end reads from *A. flavus* strains against the corresponding circularized mitogenome, and SAMtools was used to identify variants. The read mapping was also visualized and the variants identified using the Integrative Genomics Viewer (IGV) v2.9.4 (15). The mitogenomes had high coverage (830 to 1,300×) when the raw reads were mapped back to the circularized assemblies (Table 1). GeSeq v2.03 (16), a rapid organellar genome annotator, was used to annotate the mitogenomes, with *A. flavus* (GenBank accession number NC_026920.1), *Aspergillus oryzae* (NC_008282.1), *A. oryzae* 3.042 (NC_018100.1), *Aspergillus parasiticus* (NC_041445.1), and *A. fumigatus* (NC_017016.1) serving as the references. The gene names

Editor Jason E. Stajich, University of California, Riverside

Address correspondence to E. Anne Hatmaker, e.anne.hatmaker@vanderbilt.edu.

Received 28 September 2021
Accepted 29 October 2021
Published 11 November 2021

TABLE 1 Summary of assembly statistics and genomic content for the mitogenomes of nine *Aspergillus flavus* strains

Strain	No. of raw reads	Total no. of trimmed reads	No. of mitochondrial reads	Length (bp)	Coverage (×)	GC content (%)	No. of tRNA genes	No. of protein-coding genes	N_{50} (bp)
FL-B-1-1-1	8,491,957	8,136,443	97,689	29,198	830	26.2	27	16	1
PA-C-1-1-1	4,067,886	3,923,475	110,110	29,323	930	26.1	27	16	1
NC-E-3-2	5,854,147	5,718,609	109,793	29,208	940	26.1	27	16	1
TX-A-2-1-1	8,039,120	7,852,211	104,180	29,204	890	26.2	27	16	1
TX-A-13-1-1	7,750,291	7,599,530	93,233	29,207	890	26.1	27	16	1
TX-A-20-1-1	6,229,075	6,119,690	122,629	29,208	1,050	26.2	27	16	1
TX-A-1-1-1	7,180,878	7,218,213	152,774	29,220	1,300	26.1	27	16	1
TX-B-1-1-1	9,890,328	9,180,887	130,715	29,198	1,120	26.2	27	16	1
TX-B-2-1-1	7,108,842	6,927,904	131,197	29,210	1,120	26.2	27	16	1

were manually revised, and genes with low coverage and sequence similarity (<50%) were discarded. The annotation was finalized following inspection using Geneious Prime v2021.1 (17) to ensure appropriate reading frames and start/stop codons, and the genome sequences were rotated using Geneious to orient the genome start upstream of *cox1*.

All mitogenomes were circular DNA molecules ranging from 29,198 to 29,323 bp with GC contents of 26.1 to 26.2% (Table 1). All mitogenomes contained 16 protein-coding genes, including 14 highly conserved fungal mitochondrial core genes: cytochrome oxidase subunits 1, 2, and 3; NADH dehydrogenase subunits 1, 2, 3, 4, 4L, 5, and 6; ATP synthase subunits 6, 8, and 9; and cytochrome b (3). All mitogenomes also contained ribosomal protein s5 and an intron-encoded LAGLIDADG endonuclease. Two mitochondrial rRNA genes encoding small and large ribosomal subunits and 27 tRNAs were identified in each mitogenome, as expected based on previously assembled *Aspergillus* mitogenome sequences (3–5).

Data availability. The *Aspergillus flavus* mitogenome sequences from this study are available in GenBank under accession numbers MZ714575.1 (FL-B-1-1-1), MZ714576.1 (PA-C-1-1-1), MZ714577.1 (NC-E-3-2), MZ714578.1 (TX-A-2-1-1), MZ714579.1 (TX-A-13-1-1), MZ714580.1 (TX-A-20-1-1), MZ714581.1 (TX-A-1-1-1), MZ714582.1 (TX-B-1-1-1), and MZ714583.1 (TX-B-2-1-1). The SRA accession numbers for the whole-genome sequencing data used for mitogenome assembly are as follows: SRR12001149 (FL-B-1-1-1), SRR12001150 (PA-C-1-1-1), SRR12001141 (NC-E-3-2), SRR12001147 (TX-A-2-1-1), SRR12001145 (TX-A-13-1-1), SRR12001144 (TX-A-20-1-1), SRR12001148 (TX-A-1-1-1), SRR12001143 (TX-B-1-1-1), and SRR12001142 (TX-B-2-1-1).

ACKNOWLEDGMENTS

This work was performed using resources within the Advanced Computing Center for Research and Education at Vanderbilt University in Nashville, TN. Funding was provided by the Vanderbilt University Department of Biological Sciences and the Vanderbilt Evolutionary Studies Initiative.

Special thanks to Antonis Rokas and Katherine Friedman for facilitating the research.

REFERENCES

1. Samson RA, Visagie CM, Houbraken J, Hong S-B, Hubka V, Klaassen CHW, Perrone G, Seifert KA, Susca A, Tanney JB, Varga J, Kocsubé S, Szigeti G, Yaguchi T, Frisvad JC. 2014. Phylogeny, identification and nomenclature of the genus Aspergillus. Stud Mycol 78:141–173. https://doi.org/10.1016/j.simyco.2014.07.004.
2. Hoffmeister D, Keller NP. 2007. Natural products of filamentous fungi: enzymes, genes, and their regulation. Nat Prod Rep 24:393–416. https://doi.org/10.1039/b603084j.
3. Joardar V, Abrams NF, Hostetler J, Paukstelis PJ, Pakala S, Pakala SB, Zafar N, Abolude OO, Payne G, Andrianopoulos A, Denning DW, Nierman WC. 2012. Sequencing of mitochondrial genomes of nine Aspergillus and Penicillium species identifies mobile introns and accessory genes as main sources of genome size variability. BMC Genomics 13:698. https://doi.org/10.1186/1471-2164-13-698.
4. Yan Z, Chen D, Shen Y, Ye B. 2016. The complete mitochondrial genome sequence of Aspergillus flavus. Mitochondrial DNA A DNA Mapp Seq Anal 27:2671–2672. https://doi.org/10.3109/19401736.2015.1022752.
5. Park J, Lee M-K, Yu J-H, Kim J-H, Han K-H. 2020. Complete mitochondrial genome sequence of Afla-Guard®, commercially available non-toxigenic Aspergillus flavus. Mitochondrial DNA B Resour 5:3572–3574. https://doi.org/10.1080/23802359.2020.1825129.
6. Sanglard D, Ischer F, Bille J. 2001. Role of ATP-binding-cassette transporter genes in high-frequency acquisition of resistance to azole antifungals in Candida glabrata. Antimicrob Agents Chemother 45:1174–1183. https://doi.org/10.1128/AAC.45.4.1174-1183.2001.
7. Martins VP, Dinamarco TM, Soriani FM, Tudella VG, Oliveira SC, Goldman GH, Curti C, Uyemura SA. 2011. Involvement of an alternative oxidase in oxidative stress and mycelium-to-yeast differentiation in Paracoccidioides brasiliensis. Eukaryot Cell 10:237–248. https://doi.org/10.1128/EC.00194-10.
8. Drott MT, Satterlee TR, Skerker JM, Pfannenstiel BT, Glass NL, Keller NP, Milgroom MG. 2020. The frequency of sex: population genomics reveals differences in recombination and population structure of the aflatoxin-

producing fungus Aspergillus flavus. mBio 11:e00963-20. https://doi.org/10.1128/mBio.00963-20.
9. Kodama Y, Shumway M, Leinonen R, Collaboration INSD, International Nucleotide Sequence Database Collaboration. 2012. The Sequence Read Archive: explosive growth of sequencing data. Nucleic Acids Res 40:D54–D56. https://doi.org/10.1093/nar/gkr854.
10. 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.
11. 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.
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. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359. https://doi.org/10.1038/nmeth.1923.
14. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079. https://doi.org/10.1093/bioinformatics/btp352.
15. Robinson JT, Thorvaldsdóttir H, Wenger AM, Zehir A, Mesirov JP. 2017. Variant review with the Integrative Genomics Viewer. Cancer Res 77:e31–e34. https://doi.org/10.1158/0008-5472.CAN-17-0337.
16. 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.
17. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28:1647–1649. https://doi.org/10.1093/bioinformatics/bts199.