Draft Genome Sequence of an Aflatoxin-Producing *Aspergillus flavus* Strain Isolated from Food

Alexandra Schamann,ª Rolf Geisen,ª ⓘ Markus Schmidt-Heydtª

ªMax Rubner-Institut, Federal Research Institute of Nutrition and Food, Department of Safety and Quality of Fruit and Vegetables, Karlsruhe, Germany

ABSTRACT *Aspergillus flavus* is the main producer of carcinogenic aflatoxins and thus is one of the most important fungal food contaminants. Here, we report that the genome of *A. flavus* strain MRI19 was sequenced using MiSeq and PacBio platforms and that a hybrid assembly was generated.

The filamentous fungus *Aspergillus flavus* contaminates economically important crops like maize and peanuts as a preharvest and postharvest contaminant, causing large economic losses (1, 2). Additionally, due to its production of hepatotoxic aflatoxins and neurotoxic cyclopiazonic acid, *A. flavus* is a threat to food safety in regions with warm and humid climates that are affected by high levels of contamination, such as Sub-Saharan Africa (3). To analyze gene clusters encoding secondary metabolites, the whole genome of *A. flavus* strain MRI19 was sequenced using MiSeq and PacBio platforms.

The strain was isolated from tiger nuts grown in the surroundings of Valencia, Spain, via dilution series and cultivation on selective nutrient media. Strain identification was based on microscopic morphological determination and partial Sanger sequencing by Eurofins Genomics (Cologne, Germany) of PCR products of the internal transcribed spacer (ITS) region and the β-tubulin and calmodulin genes (Fig. 1) (4–6). For isolation of genomic DNA (gDNA) for MiSeq sequencing, *A. flavus* was grown for 7 days at 25°C in YES broth (20 g/L yeast extract, 150 g/L saccharose); gDNA was extracted from the pure culture using the NucleoSpin plant II kit (Macherey-Nagel, Düren, Germany). For extraction of gDNA for PacBio sequencing, the strain was grown on potato dextrose agar (PDA) at 25°C for 4 days. The mycelium was homogenized in liquid nitrogen using a mortar and pestle, and lysis was achieved with cetyltrimethylammonium bromide (CTAB) buffer (0.1 M Tris [pH 8.0], 1.4 M NaCl, 20 mM EDTA, 2% [wt/vol] CTAB, 4% [wt/vol] polyvinylpyrrolidone with an average molecular weight of 10,000 [PVP-10]). Phenol-chloroform was used for extraction, and propan-2-ol and 7.5 M ammonium acetate were added overnight for precipitation. Different DNA extraction methods were applied for the two sequencing platforms due to different requirements regarding DNA quality and quantity, which were verified using a NanoDrop 1000 spectrophotometer and a Qubit 3.0 fluorometer (both from Thermo Fisher Scientific GmbH, Bremen, Germany), respectively.

Paired-end short-read genome sequencing was carried out on a MiSeq instrument (Illumina, San Diego, CA, USA) using a sequencing library generated with the Nextera DNA Flex library preparation kit (Illumina) and quality checked with the Experion DNA 1K analysis kit (Bio-Rad, Feldkirchen, Germany). Based on the quality control (FastQC v0.11.3) of the 300-bp-long raw reads (18,887,287 read pairs), reads were quality trimmed (Trimmomatic v0.39) (7). Long-read PacBio genome sequencing was performed on a Sequel platform by BGI (Hong Kong, China), resulting in 114,583 total reads (N_{50}, 9,775 bp). Quality was checked using LongQC v1.2.0 (8).

Default software parameters were used except where otherwise noted. *De novo* hybrid assembly was carried out with SPAdes v3.14.1, resulting in 68 contigs after the elimination of short contigs (<400 bp) and mitochondrial sequences (9, 10). The assembly size was
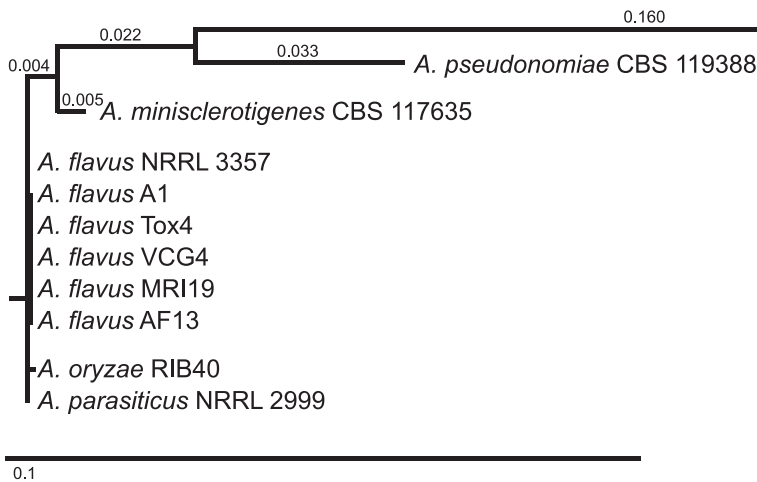
FIG 1 Phylogenetic analysis of *Aspergillus* strains. Identification of strain *A. flavus* MRI19 was based on sequencing of PCR products of the ITS region (ITS1/ITS4 primers) and the β-tubulin (Bt2a/Bt2b) and calmodulin (cmd5/cmd6) genes (4–6). PCR was performed using the peqGOLD *Taq* DNA polymerase all-inclusive kit (VWR International GmbH, Darmstadt, Germany) with the recommended mastermix composition, 2.5 μL of each primer (5 pmol/μL), and 5 μL DNA. The following cycling program was used: 95℃ for 3 min; 40 cycles of 95℃ for 30 s, 60℃ (β-tubulin and calmodulin) or 50℃ (ITS) for 40 s, and 72℃ for 90 s; and 72℃ for 3 min. Forward and reverse sequences were assembled using SeqMan Pro (LaserGene v16). Consensus sequences were compared at GenBank using BLASTN. For generation of the phylogenetic tree, the three sequences were concatenated and were compared to those of a variety of other *Aspergillus* strains as examples (*A. flavus* AF13, GenBank accession no. CP059863.1, CP059859.1, and CP059864.1 [13]; *A. flavus* A1, GenBank accession no. CP051064.1, CP051060.1, and CP051065.1 [14]; *A. flavus* NRRL 3357, GenBank accession no. CP044623.1, CP044622.1, and CP044617.1 [15]; *A. flavus* Tox4, GenBank accession no. CP051048.1, CP051044.1, and CP051049.1 [14]; *A. flavus* VCG4, GenBank accession no. CP051056.1, CP051052.1, and CP051057.1 [14]; *Aspergillus oryzae* RIB40, GenBank accession no. NC_036440.1, NC_036436.1, and NC_036441.1 [16]; *Aspergillus parasiticus* NRRL 2999, GenBank accession no. CP051032.1, CP051028.1, and CP051033.1 [14]; *Aspergillus minisclerotigenes* CBS 117635, GenBank accession no. ML732812.1, ML732765.1, and OL711675.1 [17]; *Aspergillus pseudonomiae* CBS 119388, GenBank accession no. NW_022476058.1, NW_022476011.1, and OL711683.1 [17]; *Aspergillus welwitschiae* CBS 139.54b, GenBank accession no. NW_020798696.1, NW_020798709.1, and OL711714.1 [18]), with the same fragments of the β-tubulin and calmodulin genes and the ITS region being concatenated. All sequences were aligned using MUSCLE (MegAlign Pro). A phylogenetic tree was generated using the neighbor-joining algorithm.

36.78 Mb, with a coverage of 75-fold. The average GC content was 48.36%, the N_{50} value was 1.46 Mb, and the L_{50} value was 10. Secondary metabolite biosynthetic gene clusters (BGCs) were predicted with antiSMASH v6.0.1, choosing the cluster finder algorithm for BGC border prediction (11, 12). In total, 46 BGCs were identified. In further analysis, the genome of *A. flavus* will be compared to that of other aflatoxin-producing *Aspergillus* species to find species-specific characteristics.

Data availability. This whole-genome shotgun project has been deposited in DDBJ/ENA/GenBank under accession no. JAGYXF000000000 and BioProject accession no. PRJNA727175. The version described in this paper is version no. JAGYXF010000000. The raw sequence reads have been deposited in the Sequence Read Archive (SRA) under accession no. SRX10945236 and SRX10945235. The partial gene sequences have been deposited in GenBank under accession no. OL513390, OL513391, and OL513392.

REFERENCES

1. Klich MA. 2007. *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. Taniwaki MH, Pitt JI, Magan N. 2018. *Aspergillus* species and mycotoxins: occurrence and importance in major food commodities. Curr Opin Food Sci 23:38–43. https://doi.org/10.1016/j.cofs.2018.05.008.
3. Benkerroum N. 2020. Aflatoxins: producing-molds, structure, health issues and incidence in Southeast Asian and Sub-Saharan African countries. Int J Environ Res Public Health 17:1215. https://doi.org/10.3390/ijerph17041215.
4. 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, Shinsky JJ, White TJ (ed), PCR protocols: a guide to methods applications. Academic Press, San Diego, CA.
5. Glass NL, Donaldson GC. 1995. Development of primer sets designed for use with the PCR to amplify conserved genes from filamentous ascomycetes. Appl Environ Microbiol 61:1323–1330. https://doi.org/10.1128/aem.61.4.1323-1330.1995.

6. Hong SB, Cho HS, Shin HD, Frisvad JC, Samson RA. 2006. Novel *Neosartorya* species isolated from soil in Korea. Int J Syst Evol Microbiol 56:477–486. https://doi.org/10.1099/ijs.0.63980-0.

7. 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.

8. Fukasawa Y, Ermini L, Wang H, Carty K, Cheung MS. 2020. LongQC: a quality control tool for third generation sequencing long read data. G3 (Bethesda) 10:1193–1196. https://doi.org/10.1534/g3.119.400864.

9. 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.

10. Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics 29:1072–1075. https://doi.org/10.1093/bioinformatics/btt086.

11. Medema MH, Blin K, Cimermancic P, De Jager V, Zakrzewski P, Fischbach MA, Weber T, Takano E, Breitling R. 2011. antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. Nucleic Acids Res 39:W339–W346. https://doi.org/10.1093/nar/gkr466.

12. Blin K, Shaw S, Kloosterman AM, Charlop-Powers Z, van Wezel GP, Medema MH, Weber T. 2021. antiSMASH 6.0: improving cluster detection and comparison capabilities. Nucleic Acids Res 49:W29–W35. https://doi.org/10.1093/nar/gkab335.

13. Fountain JC, Clevenger JP, Nadon B, Youngblood RC, Korani W, Chang PK, Starr D, Wang H, Isett B, Johnston HR, Wiggins R, Agarwal G, Chu Y, Kemerait RC, Pandey MK, Bhatnagar D, Ozias-Akins P, Varshney RK, Scheffler BE, Vaughn JN, Guo B. 2020. Two new *Aspergillus flavus* reference genomes reveal a large insertion potentially contributing to isolate stress tolerance and aflatoxin production. G3 (Bethesda) 10:3515–3531. https://doi.org/10.1534/g3.120.401405.

14. Fountain JC, Clevenger JP, Nadon B, Wang H, Abbas HK, Kemerait RC, Scully BT, Vaughn JN, Guo B. 2020. Draft genome sequences of one *Aspergillus parasiticus* isolate and nine *Aspergillus flavus* isolates with varying stress tolerance and aflatoxin production. Microbiol Resour Announc 9:e00478-20. https://doi.org/10.1128/MRA.00478-20.

15. Skerker JM, Pianalto KM, Mondo SJ, Yang K, Arkin AP, Keller NP, Grigoriev IV, Glass NL. 2021. Chromosome assembled and annotated genome sequence of *Aspergillus flavus* NRRL 3357. G3 (Bethesda) 11:jkab213. https://doi.org/10.1093/g3journal/jkab213.

16. Machida M, Asai K, Sano M, Tanaka T, Kumagai T, Terai G, Kusumoto KI, Arima T, Akita O, Kashiwagi Y, Abe K, Gomi K, Horiuchi H, Kitamoto K, Kobayashi T, Takeuchi M, Denning DW, Galagan JE, Nierman WC, Yu J, Archer DB, Bennett JW, Bhatnagar D, Cleveland TE, Fedorova ND, Gotoh O, Horikawa H, Hosoyama A, Ichinomiya M, Igarashi R, Iwashita K, Juvvadi PR, Kato M, Kato Y, Kin T, Kokubun A, Maeda H, Maeyama N, Maruyama JI, Nagasaki H, Nakajima T, Oda K, Okada K, Paulsen I, Sakamoto K, Sawano T, Takahashi M, Takase K, Terabayashi Y, Wortman JR, Yamada O, Yamagata Y, Anazawa H, Hata Y, Koide Y, Komori T, Koyama Y, Minetoki T, Suharnan S, Tanaka A, Isono K, Kuhara S, Ogasawara N, Kikuchi H. 2005. Genome sequencing and analysis of *Aspergillus oryzae*. Nature 438:1157–1161. https://doi.org/10.1038/nature04300.

17. Kjærbølling I, Vesth T, Frisvad JC, Nybo JL, Theobald S, Kildgaard S, Petersen TI, Kuo A, Sato A, Lyhne EK, Kogle ME, Wiebenga A, Kun RS, Lubbers RJM, Mäkelä MR, Barry K, Chovatia M, Clum A, Daum C, Haridas S, He G, LaButti K, Lipzen A, Mondo S, Pangilinan J, Riley R, Salamov A, Simmons BA, Magnuson JK, Henrissat B, Mortensen UH, Larsen TO, de Vries RP, Grigoriev IV, Machida M, Baker SE, Andersen MR. 2020. A comparative genomics study of 23 *Aspergillus* species from section *Flavi*. Nat Commun 11:1106. https://doi.org/10.1038/s41467-019-14051-y.

18. Vesth TC, Nybo JL, Theobald S, Frisvad JC, Larsen TO, Nielsen KF, Hoof JB, Brandl J, Salamov A, Riley R, Gladden JM, Phatale P, Nielsen MT, Lyhne EK, Kogle ME, Strasser K, McDonnell E, Barry K, Clum A, Chen C, LaButti K, Haridas S, Nolan M, Sandor L, Kuo A, Lipzen A, Hainaut M, Drula E, Tsang A, Magnuson JK, Henrissat B, Wiebenga A, Simmons BA, Mäkelä MR, de Vries RP, Grigoriev IV, Mortensen UH, Baker SE, Andersen MR. 2018. Investigation of inter- and intraspecies variation through genome sequencing of *Aspergillus* section *Nigri*. Nat Genet 50:1688–1695. https://doi.org/10.1038/s41588-018-0246-1.