

# Whole-genome sequence of an *Aspergillus parasiticus* strain isolated from Kenyan soil

Alexandra Schamann,<sup>1</sup> Rolf Geisen,<sup>1</sup> Markus Schmidt-Heydt<sup>1</sup>

**AUTHOR AFFILIATION** See affiliation list on p. 3.

**ABSTRACT** *Aspergillus parasiticus* is an important aflatoxigenic fungus, frequently found in soil samples. Here, we report the sequencing of *A. parasiticus* strain MRI410 using Illumina MiSeq and Oxford Nanopore platforms. This strain was isolated from soil of a Kenyan maize field.

**KEYWORDS** aflatoxin, *Aspergillus parasiticus*, high-quality genome sequence

The filamentous fungus *Aspergillus parasiticus* is, in addition to *A. flavus*, the most serious aflatoxigenic *Aspergillus* section *Flavi* species. It is described to prefer ground crop hosts to plant hosts and was already repeatedly isolated from peanuts (1, 2). To our knowledge, genomes of *A. parasiticus* have only been sequenced by short-read sequencing technologies until now. To provide a genome assembly of an *A. parasiticus* strain sequenced also with a long-read sequencing technology, strain MRI410 was sequenced using Oxford Nanopore in addition to Illumina MiSeq technology.

For the isolation of this strain, soil samples from the rhizosphere of a maize field in Makueni (1.382 S, 37.322 E), Kenya, were diluted using Tween-80/NaCl-mixture (NaCl 9 g/L, Tween-80 1 g/L, and agar 1 g/L), and serial dilutions were incubated on selective nutrient media. Then, individual colonies were separated. The strain was identified by partial sequencing of  $\beta$ -tubulin, calmodulin, and nitrate reductase genes (Fig. 1), and the identification was verified by the Westerdijk Fungal Biodiversity Institute (Utrecht, Netherlands). After the incubation of *A. parasiticus* MRI410 on malt extract agar (malt extract 17 g/L, glucose 5 g/L, and agar 16 g/L) at 25°C for 4 d and its homogenization in liquid nitrogen, DNA was isolated using the NucleoSpin Plant II kit (Macherey-Nagel, Düren, Germany) following the official protocol. DNA quality and concentration were verified using a NanoDrop 1000 spectrophotometer and a Qubit 3.0 fluorometer (both Thermo Fisher Scientific GmbH, Bremen, Germany). The same DNA sample was used for the following sequencing with Oxford Nanopore and Illumina MiSeq technologies.

For long-read sequencing, a library was prepared using the Rapid Sequencing kit (Oxford Nanopore Technologies, Oxford, United Kingdom). Sequencing was performed on a MinION Mk1C instrument (Oxford Nanopore Technologies, Oxford, United Kingdom) resulting in 573,070 raw reads (1.56 Gb, N<sub>50</sub> 6.35 kb). Subsequent base calling of Nanopore data was done using Guppy, and quality was checked using pycoQC (13). For short-read sequencing, a library was generated using the DNA Prep kit (Illumina, San Diego, CA, USA), and its quality was controlled with the Experion DNA 1k analysis kit (Bio-Rad, Feldkirchen, Germany). The 2 × 300 bp sequencing was performed on a MiSeq platform (Illumina, San Diego, CA, USA). A total of 26,943,603 raw read pairs were trimmed (Trimmomatic-0.39) based on the quality control (FastQC 0.11.3) (14).

Default software parameters were used except where otherwise noted. A *de novo* assembly of the Nanopore data was generated using Flye 2.8.2 (average coverage: 22.9×) (15, 16). Subsequently, the MiSeq data were aligned (average coverage: 235.2×) to the resulting contigs using BWA 0.7.17, SAMtools 1.10, and QualiMap 2.2.1 (17–19), and the

**Editor** Antonis Rokas, Vanderbilt University, Nashville, Tennessee, USA

Address correspondence to Markus Schmidt-Heydt, markus.schmidt-heydt@mri.bund.de.

The authors declare no conflict of interest.

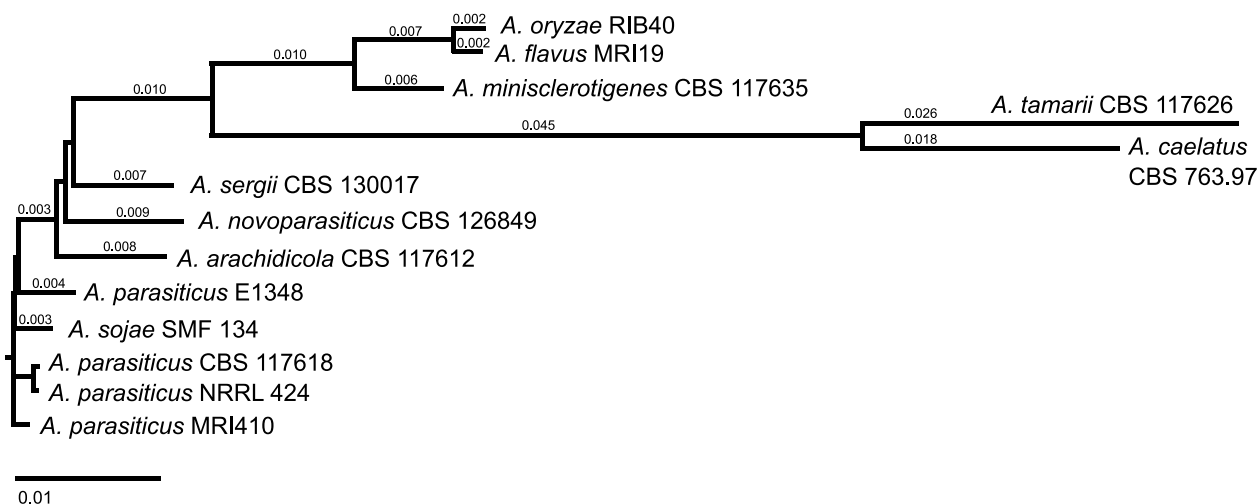
See the funding table on p. 3.

**Received** 14 March 2023

**Accepted** 14 June 2023

**Published** 18 July 2023

Copyright © 2023 Schamann et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license.



**FIG 1** Phylogenetic tree of different *Aspergillus* species. *A. parasiticus* MRI410 was initially identified by sequencing PCR products of partial  $\beta$ -tubulin [Bt2a/2b (3)], calmodulin [cmd5/6 (4), cmd2F/2R (5)], and nitrate reductase [niaDF/AR (5), niaDBF/BR (5, 6), niaDCF/CR (7)] genes. The recommended mastermix composition of the peqGOLD Taq DNA polymerase all-inclusive kit (VWR International GmbH, Darmstadt, Germany) with 2.5  $\mu$ L of each primer (5 pmol/ $\mu$ L) and 5  $\mu$ L DNA was used for the PCR. The following cycling program was chosen: 95°C for 3 min; 40 cycles of 95°C for 30 s, 52°C (cmd2F/2R)/55°C (niaDF/AR)/57°C (niaDBF/BR, niaDCF/CR)/60°C (Bt2a/2b, cmd5/6) for 40 s, and 72°C for 90 s; and 72°C for 3 min. PCR products were Sanger sequenced by Eurofins Genomics (Cologne, Germany). Forward and reverse sequences were assembled with SeqMan Pro (LaserGene v17, DNASTAR, Inc, Madison, WI, USA). For calmodulin and nitrate reductase, the two, respectively, three overlapping consensus sequences were concatenated using MegAlign Pro and SeqBuilder Pro (LaserGene v17, DNASTAR, Inc, Madison, WI, USA). Resulting sequences were compared to the database of the NCBI using BLASTN. To generate a phylogenetic tree using the neighbor-joining algorithm, the three partial genes were concatenated and aligned (Mauve, MegAlign Pro, LaserGene v17, DNASTAR, Inc, Madison, WI, USA) with the same sequences of other *Aspergillus* species [*A. arachidicola* CBS 117612: [ML737115.1](#), [ML737234.1](#), [ML737155.1](#) (8); *A. caelatus* CBS 763.97: [NW\\_022475357.1](#), [NW\\_022475408.1](#), [NW\\_022475603.1](#) (8); *A. flavus* MRI19: [JAGYXF010000057.1](#), [JAGYXF010000047.1](#), [JAGYXF010000013.1](#) (9); *A. minisclerotigenes* CBS 117635: [ML732812.1](#), [ML732765.1](#), [ML732764.1](#) (8); *A. novoparasiticus* CBS 126849: [ML733430.1](#), [ML733443.1](#), [ML733467.1](#) (8); *A. oryzae* RIB40: [NC\\_036440.1](#), [NC\\_036436.1](#), [NC\\_036438.1](#) (10); *A. parasiticus* NRRL 424: [MK119739.1](#), [MK119705.1](#), [MK119671.1](#) (6); *A. parasiticus* CBS 117618: [ML734942.1](#), [ML734938.1](#), [ML734939.1](#) (8); *A. parasiticus* E1348: [SJFE01000042.1](#), [SJFE01000024.1](#), [SJFE01000049.1](#) (11); *A. sergii* CBS 130017: [ML741807.1](#), [ML741799.1](#), [ML741762.1](#) (8); *A. sojae* SMF 134: [CP035525.1](#), [CP035529.1](#), [CP035527.1](#) (12), *A. tamarii* CBS 117626: [ML738700.1](#), [ML738591.1](#), [ML738590.1](#) (8)].

alignment was polished with Pilon 1.23 and SAMtools 1.10 (without activating the IUPAC nucleotide codes) (16, 20). The resulting assembly consisted of 60 contigs and a total of 38.68 Mb ( $N_{50}$ : 2,119,486 bp, GC content: 47.68%) of gDNA. A further contig contained the complete mitochondrial DNA (29,136 bp). To assess the completeness of the genome assembly, a BUSCO (v5.4.6) analysis using the lineage database *ascomycota\_odb10* was performed (21). A total of 1,638 (96.0%) complete single-copy and 15 (0.9%) complete duplicated orthologs, 2 (0.1%) fragmented orthologs, and 51 (3.0%) missing orthologs were retrieved.

In total, 55 gene clusters coding secondary metabolites were predicted using antiSMASH 6.1.0 with the cluster finder algorithm for BGC border prediction (22, 23). The aflatoxin gene cluster will be further analyzed and compared to that of other *Aspergillus* species.

## ACKNOWLEDGMENTS

The work described here is part of the project AflaZ, which is being funded by the Federal Office for Agriculture and Food (BLE) under the reference AflaZ 2816PROC11.

We thank Charles Nkonge for providing the soil samples, Christian Roder for the isolation of the strain, Karla Hell for her excellent technical support, and Dr. Andreas Dötsch for the bioinformatic support.

## AUTHOR AFFILIATION

<sup>1</sup>Department of Safety and Quality of Fruit and Vegetables, Max Rubner-Institut (MRI), Federal Research Institute of Nutrition and Food, Karlsruhe, Germany

## AUTHOR ORCID*s*

Alexandra Schamann  <http://orcid.org/0000-0003-2844-9873>

Markus Schmidt-Heydt  <http://orcid.org/0000-0003-2844-9873>

## FUNDING

Funder	Grant(s)	Author(s)
Bundesanstalt für Landwirtschaft und Ernährung (BLE)	AflaZ 2816PROC11	Alexandra Schamann

## AUTHOR CONTRIBUTIONS

Alexandra Schamann, Data curation, Formal analysis, Investigation, Methodology, Validation, Visualization, Writing – original draft, Writing – review and editing | Rolf Geisen, Writing – review and editing | Markus Schmidt-Heydt, Investigation, Writing – review and editing, Conceptualization, Funding acquisition, Project administration, Resources, Software, Supervision

## DATA AVAILABILITY

This Whole Genome Shotgun project has been deposited at DDBJ/ENA/GenBank under the accession no. [JAMRJN000000000](https://doi.org/10.1093/bioinformatics/btad000) and BioProject accession no. [PRJNA835319](https://doi.org/10.1093/bioinformatics/btad000). The version described in this paper is version [JAMRJN010000000](https://doi.org/10.1093/bioinformatics/btad000). The raw sequence reads have been deposited in the Sequence Read Archive (SRA) under accession no. [SRX17682209](https://doi.org/10.1093/bioinformatics/btad000) and [SRX17682210](https://doi.org/10.1093/bioinformatics/btad000). The partial gene sequences have been deposited at the GenBank under accession no. [OQ909814](https://doi.org/10.1093/bioinformatics/btad000), [OQ909817](https://doi.org/10.1093/bioinformatics/btad000), and [OQ909820](https://doi.org/10.1093/bioinformatics/btad000).

## REFERENCES

- Barros GG, Torres AM, Rodriguez MI, Chulze SN. 2006. Genetic diversity within *Aspergillus flavus* strains isolated from peanut-cropped soils in Argentina. *Soil Biol Biochem* 38:145–152. <https://doi.org/10.1016/j.soilbio.2005.04.028>
- Amaike S, Keller NP. 2011. *Aspergillus flavus*. *Annu Rev Phytopathol* 49:107–133. <https://doi.org/10.1146/annurev-phyto-072910-095221>
- 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>
- Hong S-B, Cho H-S, Shin H-D, 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>
- Probst C, Callicott KA, Cotty PJ. 2012. Deadly strains of Kenyan *Aspergillus* are distinct from other aflatoxin producers. *Eur J Plant Pathol* 132:419–429. <https://doi.org/10.1007/s10658-011-9887-y>
- Singh P, Orbach MJ, Cotty PJ. 2018. *Aspergillus texensis*: a novel aflatoxin producer with S morphology from the United States. *Toxins (Basel)* 10:513. <https://doi.org/10.3390/toxins10120513>
- Singh P, Cotty PJ. 2019. Characterization of *Aspergilli* from dried red chilies (*Capsicum* spp.): insights into the etiology of aflatoxin contamination. *Int J Food Microbiol* 289:145–153. <https://doi.org/10.1016/j.ijfoodmicro.2018.08.025>
- Kjærboelling 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>
- Schamann A, Geisen R, Schmidt-Heydt M. 2022. Draft genome sequence of an aflatoxin-producing *Aspergillus flavus* strain isolated from food. *Microbiol Resour Announc* 11:e0089421. <https://doi.org/10.1128/mra.00894-21>
- Machida M, Asai K, Sano M, Tanaka T, Kumagai T, Terai G, Kusumoto K-I, 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 J, 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>
- Arias RS, Mohammed A, Orner VA, Faustini PC, Lamb MC, Sobolev VS. 2020. Sixteen draft genome sequences representing the genetic diversity of *Aspergillus flavus* and *Aspergillus parasiticus* colonizing peanut seeds in Ethiopia. *Microbiol Resour Announc* 9:e00591-20. <https://doi.org/10.1128/MRA.00591-20>
- Kim KU, Kim KM, Choi Y-H, Hurh B-S, Lee I. 2019. Whole genome analysis of *Aspergillus sojae* SMF 134 supports its merits as a starter for soybean fermentation. *J Microbiol* 57:874–883. <https://doi.org/10.1007/s12275-019-9152-1>
- Leger A, Leonardi T. 2019. PycoQC, interactive quality control for Oxford Nanopore Sequencing. *JOSS* 4:1236. <https://doi.org/10.21105/joss.01236>

14. 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>
15. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol* 37:540–546. <https://doi.org/10.1038/s41587-019-0072-8>
16. 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>
17. Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>
18. 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>
19. Okonechnikov K, Conesa A, García-Alcalde F. 2016. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics* 32:292–294. <https://doi.org/10.1093/bioinformatics/btv566>
20. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9:e112963. <https://doi.org/10.1371/journal.pone.0112963>
21. Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. 2021. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol Biol Evol* 38:4647–4654. <https://doi.org/10.1093/molbev/msab199>
22. 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>
23. 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>