AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

GENOME SEQUENCES

Whole-Genome Sequences of Two Kenyan *Aspergillus minisclerotigenes* Strains

Alexandra Schamann,[a] Rolf Geisen,[a] Markus Schmidt-Heydt[a]

[a]Max Rubner-Institut, Federal Research Institute of Nutrition and Food, Department of Safety and Quality of Fruit and Vegetables, Karlsruhe, Germany

ABSTRACT Here, we report the sequencing of the whole genome, including the mitochondrial DNA, of the two highly aflatoxigenic *Aspergillus minisclerotigenes* strains MRI390 and MRI400 using the MiSeq and PacBio platforms and the generated assemblies. The strains were isolated from Kenyan maize kernels.

Aspergillus minisclerotigenes is a highly aflatoxigenic *Aspergillus* section *Flavi* species repeatedly isolated from samples originating from Kenyan regions, where aflatoxin contamination poses an exceptional risk to food safety and several outbreaks of aflatoxicosis have been reported (1–3). Nevertheless, limited research has focused on *A. minisclerotigenes* until now. To increase our knowledge of this fungus, the genomes of the Kenyan *A. minisclerotigenes* strains MRI390 and MRI400 were sequenced using the MiSeq and PacBio platforms.

The strains were isolated from ground maize kernels from Katumani, Kenya, by mixing the ground kernels with a Tween 80-NaCl solution (9 g/L NaCl, 1 g/L Tween 80, 1 g/L agar), generating a dilution series, and cultivating the dilutions on selective nutrient medium. The isolates were identified by partial sequencing of the β-tubulin (Bt2a/2b [4]), calmodulin (cmd5/6 [5], cmd2F/2R [6]), and nitrate reductase (niaDF/AR [6], niaDBF/BR [6, 7], niaDCF/CR [8]) genes. For this, PCR was performed using the peqGOLD Taq DNA polymerase all-inclusive kit (VWR International GmbH, Darmstadt, Germany) with 2.5 μL of each primer (5 pmol/μL) and 5 μL DNA. Amplification was achieved with the following cycling program: 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. The PCR products were sequenced in both directions by Eurofins Genomics (Cologne, Germany) using Sanger technology, and the sequences were assembled (SeqMan Pro, LaserGene v17). The two respectively three overlapping consensus sequences of calmodulin respectively nitrate reductase were concatenated (MegAlign Pro, SeqBuilder Pro). The sequences of the three partial genes (β-tubulin, calmodulin, nitrate reductase) were compared to sequences in NCBI using BLASTN. Concatenating these three gene sequences, a phylogenetic tree was created using the neighbor-joining algorithm with the same partial genes of a variety of different *Aspergillus* strains: *Aspergillus aflatoxiformans* BN038-G (GenBank accession no. MK119747.1, MK119713.1, MK119679.1) (7), *Aspergillus arachidicola* CBS 117612 (ML737115.1, ML737234.1, ML737155.1) (9), *Aspergillus caelatus* CBS 763.97 (NW_022475357.1, NW_022475408.1, NW_022475603.1) (9), *Aspergillus flavus* MRI19 (JAGYXF010000057.1, JAGYXF010000047.1, JAGYXF010000013.1) (10), *A. minisclerotigenes* CBS 117635 (ML732812.1, ML732765.1, ML732764.1) (9), *A. minisclerotigenes* DTO 009-F5 (MT024508.1, MT024497.1, MT024519.1) (11), *A. minisclerotigenes* DTO 228-H1 (MT024515.1, MT024504.1, MT024526.1) (11), *A. minisclerotigenes* DTO 045-F6 (MT024512.1, MT024501.1, MT024523.1) (11), *A. minisclerotigenes* DTO 303-C6 (MT024516.1, MT024506.1, MT024528.1) (11), *Aspergillus novoparasiticus* CBS 126849 (ML733430.1, ML733443.1, ML733467.1) (9), *Aspergillus oryzae* RIB40 (NC_036440.1, NC_036436.1, NC_036438.1) (12), *Aspergillus parasiticus* CBS 117618 (ML734942.1,

Editor Antonis Rokas, Vanderbilt University

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

The authors declare no conflict of interest.

Received 17 March 2023
Accepted 14 June 2023

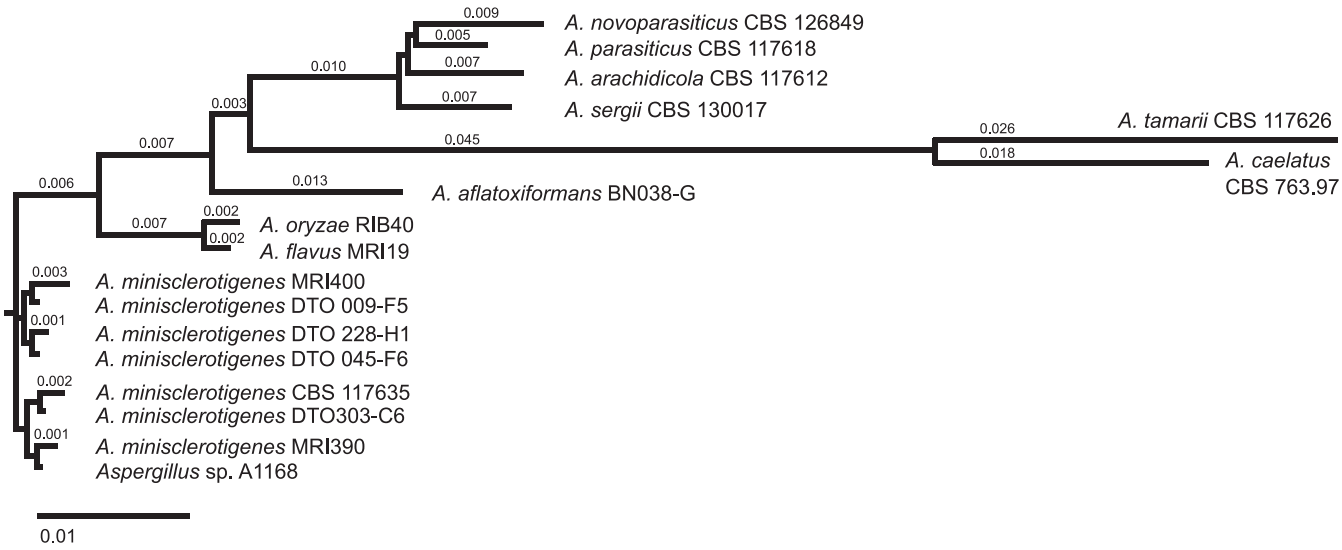

FIG 1 Phylogenetic tree of various *Aspergillus* strains, based on partial sequencing of the β-tubulin, calmodulin, and nitrate reductase genes.

ML734938.1, ML734939.1) (9), *Aspergillus sergii* CBS 130017 (ML741807.1, ML741799.1, ML741762.1) (9), *Aspergillus tamarii* CBS 117626 (ML738700.1, ML738591.1, ML738590.1) (9), and *Aspergillus* sp. strain A1168 (MK119750.1, MN987082.1, MK119682.1) (7, 13) (Fig. 1). The identification was confirmed by the experts at the Westerdijk Fungal Biodiversity Institute (Utrecht, Netherlands). For DNA extraction, the fungal strains were grown for 4 days on potato-dextrose agar at 25°C, and the mycelium was homogenized using liquid nitrogen and a mortar and pestle. For MiSeq sequencing, DNA was extracted using the NucleoSpin plant II kit (Macherey-Nagel, Düren, Germany) following the manufacturer's instructions. For PacBio sequencing, DNA of the homogenized mycelium was extracted using 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 having an average molecular weight of 10,000 [PVP-10]) for lysis, phenol-chloroform for extraction, and propan-2-ol and 7.5 M ammonium acetate for precipitation overnight. Two DNA extraction protocols were followed due to the different quality and quantity requirements of the DNA, which were checked using a NanoDrop 1000 spectrophotometer and a Qubit 3.0 fluorometer (both from Thermo Fisher Scientific GmbH, Bremen, Germany).

A library was prepared using the Illumina DNA prep kit (San Diego, USA) and sequenced (2×300 bp) on the MiSeq platform (Illumina). The raw reads were quality checked (FastQC v0.11.3), as well as quality trimmed and checked for remaining adapter sequences (Trimmomatic v0.39) (14). A PacBio sequencing library was prepared using the SMRTbell Express template prep kit 2.0 (Pacific BioSciences, Menlo Park, USA) following the manufacturer's instructions. DNA was sheared into fragments of 6 to 10 kb using g-TUBE devices (Covaris, Brighton, UK). BluePippin (Sage Science, Beverly, USA) was used for size selection, before sequencing on the Sequel platform by BGI (Hong Kong, China). The quality of the PacBio data was checked using LongQC v1.2.0 (15).

Default software parameters were used except where otherwise noted. Different assembly tools were tested for each strain, and the software fitting best to each was chosen for the final assembly. Combining the data from both technologies, *de novo* hybrid assembly was carried out using SPAdes v3.14.1 for strain MRI400 (16, 17). For strain MRI390, *de novo* assembly of the PacBio data was performed using Flye v2.8.2, and then the MiSeq data were aligned to the resulting contigs; the alignment was polished using Pilon v1.23 and SAMtools v1.10 (17–19). Short contigs (<400 bp) were excluded. In addition to the genomic DNA, the complete mitochondrial DNA was sequenced. The completeness of the genome assemblies was determined using BUSCO v5.4.6 with the lineage database ascomycota_odb10 (20). The sequencing data and assembly metrics are shown in Table 1. The genomes of *A.*

TABLE 1 Sequencing and assembly data

Parameter	Data for *Aspergillus minisclerotigenes* strain: MRI390	MRI400
Genome size (Mb)	38.03	37.88
Mitochondrial DNA (bp)	29,195	29,329
No. of contigs	28	52
GC content (%)	47.51	47.51
N_{50} value (bp)	3,823,310	1,345,410
Coverage (×)	74	102
Total no. of MiSeq paired-end raw reads	28,506,465	31,129,593
PacBio data		
Total no. of raw reads	288,480	161,135
N_{50} value (bp)	11,239	10,403
Avg read length (bp)	9,676	9,105
Total no. of BUSCO orthologs	1,706	1,706
Complete single-copy, complete multicopy, fragmented, and missing orthologs (%)	93.7, 0.9, 0.5, 4.9	93.4, 0.8, 0.6, 5.2

minisclerotigenes will be analyzed more deeply and compared to other aflatoxigenic *Aspergillus* strains, focusing especially on the aflatoxin gene cluster.

Data availability. This whole-genome shotgun project has been deposited at DDBJ/ENA/GenBank under accession no. JAHXGP000000000 and JAHYSD000000000 and BioProject accession no. PRJNA742918 and PRJNA741918. The versions described in this paper are versions JAHXGP010000000 and JAHYSD010000000. The raw sequence reads have been deposited in the Sequence Read Archive (SRA) under accession no. SRR15400241, SRR15400240, SRR15130309, and SRR15130308. The partial gene sequences have been deposited at GenBank under accession no. OQ909815, OQ909818, OQ909821, OQ909816, OQ909819, and OQ909822.

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 reference no. AflaZ 2816PROC11.

We thank Charles Nkonge for providing the maize samples and Andreas Dötsch for bioinformatic support.

REFERENCES

1. Azziz-Baumgartner E, Lindblade K, Gieseker K, Rogers HS, Kieszak S, Njapau H, Schleicher R, McCoy LF, Misore A, DeCock K, Rubin C, Slutsker L, Aflatoxin Investigation Group. 2005. Case-control study of an acute aflatoxicosis outbreak, Kenya, 2004. Environ Health Perspect 113:1779–1783. https://doi.org/10.1289/ehp.8384.
2. Okoth S, De Boevre M, Vidal A, Diana Di Mavungu J, Landschoot S, Kyallo M, Njuguna J, Harvey J, De Saeger S. 2018. Genetic and toxigenic variability within *Aspergillus flavus* population isolated from maize in two diverse environments in Kenya. Front Microbiol 9:57. https://doi.org/10.3389/fmicb.2018.00057.
3. Oloo RD, Okoth S, Wachira P, Mutiga S, Ochieng P, Kago L, Nganga F, Entfellner J-BD, Ghimire S. 2019. Genetic profiling of *Aspergillus* isolates with varying aflatoxin production potential from different maize-growing regions of Kenya. Toxins (Basel) 11:467. https://doi.org/10.3390/toxins11080467.
4. 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.
5. 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.
6. 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.
7. 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.
8. 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.
9. 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.
10. 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:e00894-21. https://doi.org/10.1128/mra.00894-21.
11. Houbraken J, Kocsubé S, Visagie CM, Yilmaz N, Wang XC, Meijer M, Kraak B, Hubka V, Bensch K, Samson RA, Frisvad JC. 2020. Classification of *Aspergillus*, *Penicillium*, *Talaromyces* and related genera (*Eurotiales*): an overview of families, genera, subgenera, sections, series and species. Stud Mycol 95:5–169. https://doi.org/10.1016/j.simyco.2020.05.002.
12. 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 J, Nagasaki H, Nakajima T, Oda K, Okada K, Paulsen I, Sakamoto K, Sawano T, Takahashi M, Takase K, Terabayashi Y, Wortman JR, et al. 2005. Genome sequencing and analysis of *Aspergillus oryzae*. Nature 438:1157–1161. https://doi.org/10.1038/nature04300.

13. Singh P, Callicott KA, Orbach MJ, Cotty PJ. 2020. Molecular analysis of S-morphology aflatoxin producers from the United States reveals previously unknown diversity and two new taxa. Front Microbiol 11:1236. https://doi.org/10.3389/fmicb.2020.01236.
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. 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.
16. 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.
17. 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.
18. 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.
19. 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.
20. 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.