

# Genome Sequence of *Madurella mycetomatis* mm55, Isolated from a Human Mycetoma Case in Sudan

Sandra Smit,<sup>b</sup> Martijn F. L. Derks,<sup>b</sup> Sander Bervoets,<sup>c</sup> Ahmed Fahal,<sup>d</sup> Willem van Leeuwen,<sup>e</sup> Alex van Belkum,<sup>f</sup> Wendy W. J. van de Sande<sup>a</sup>

Department of Medical Microbiology and Infectious Diseases, Erasmus MC, Rotterdam, The Netherlands<sup>a</sup>; Bioinformatics Group, Wageningen University, Wageningen, The Netherlands<sup>b</sup>; ServiceXS B.V., Leiden, The Netherlands<sup>c</sup>; Mycetoma Research Centre, University of Khartoum, Khartoum, Sudan<sup>d</sup>; Professorate Innovative Molecular Diagnostics, University of Applied Sciences Leiden, Leiden, The Netherlands<sup>e</sup>; Microbiology R&D Clinical Unit, bioMérieux, La Balme les Grottes, France<sup>f</sup>

**We present the first genome sequence for a strain of the main mycetoma causative agent, *Madurella mycetomatis*. This 36.7-Mb genome sequence will offer new insights into the pathogenesis of mycetoma, and it will contribute to the development of better therapies for this neglected tropical disease.**

Received 5 April 2016 Accepted 8 April 2016 Published 26 May 2016

**Citation** Smit S, Derks MFL, Bervoets S, Fahal A, van Leeuwen W, van Belkum A, van de Sande WWJ. 2016. Genome sequence of *Madurella mycetomatis* mm55, isolated from a human mycetoma case in Sudan. *Genome Announc* 4(3):e00418-16. doi:10.1128/genomeA.00418-16.

**Copyright** © 2016 Smit et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Wendy W. J. van de Sande, w.vandesande@erasmusmc.nl.

Mycetoma is a chronic granulomatous subcutaneous tropical infectious disease, caused by either fungi (eumycetoma) or bacteria (actinomycetoma). The most common causative agent of human mycetoma is the fungus *Madurella mycetomatis* (1). A characteristic of mycetoma is that the causative agents organize themselves in grains, which are black in the case of *M. mycetomatis*. The pathogenesis of mycetoma is barely understood. In recent years, genome sequences of the actinomycetoma causative agents, *Nocardia brasiliensis* (2, 3), *Streptomyces somaliensis* (4), and *Actinomadura madurae* (5) have been published, but no genome sequence for a fungal mycetoma agent was available until now. Here, we report the first genome sequence for a clinical isolate of *M. mycetomatis*.

*M. mycetomatis* mm55 was isolated on 25 November 1999 in the Mycetoma Research Centre (Khartoum, Sudan) from an extensive mycetoma case in the foot in a 22-year-old male patient. The patient from central Sudan had been suffering from mycetoma for more than 12 years. This strain was isolated by direct culture of the black grains obtained by a deep biopsy and identified by morphology, PCR-RFLP, and sequencing of the ITS region (6). Strain mm55 was kept on Sabouraud agar and ground in a mortar and pestle in liquid nitrogen, and the DNA was isolated with the Promega Wizard kit (Promega, Leiden, The Netherlands). The genome was reconstructed from Roche 454 (12.3× coverage), Illumina (28.2×), and PacBio (3.0×) data. Minimus2 (7) was used to merge one assembly made by the whole-genome sequencing assembler (8) and PBjelly (9), and another made by SPAdes (10) and SSPACE-LongRead (11). The resulting draft genome is 36.7 Mbp long, fragmented into 804 scaffolds ( $N_{50}$  of 81.8 kb; G+C content of 54.9%).

The genome was annotated using MAKER2 (12) and Evidence-Modeler (13) (predictions with long introns were removed). Gene predictors SNAP and Augustus were trained using *Chaetomium globosum*. Protein and expressed sequence tag sequences of eight fungi (six in the order of *Sordariales*) were used as homology evidence. Functional annotation was done using BLASTp (14) against the Swiss-Prot and TrEMBL databases (15). Protein domains, Pfam su-

perfamily IDs, and gene-ontology terms were assigned using InterProScan (16). The draft genome contains 10,707 protein-coding genes. CEGMA (17) and BUSCO (18) showed that 89% of core eukaryotic genes are completely present in the assembly, and OrthoFinder (19) indicated that 93% of the predicted genes have orthologs in other species.

The genomic data generated provide a potential leap forward in our comprehension of the biology and pathogenic mechanisms underlying human eumycetoma and will help our understanding of the formation of fungal grains. By comparing the genome to closely related fungi not causing mycetoma, proteins can be identified that could play a role in the development of the mycetoma grain. Furthermore, this sequence offers a rich resource for the identification of *M. mycetomatis*-specific proteins, which could be used to develop novel diagnostic tools or which could serve as novel drug targets. The latter is especially important, since the current treatment is often limited to azole treatment and amputation.

**Nucleotide sequence accession numbers.** This annotated genome (version 2.0) has been deposited at DDBJ/EMBL/GenBank under the accession number [LCTW000000000](https://www.ncbi.nlm.nih.gov/nuccore/LCTW000000000), BioProject PRJNA267680.

## ACKNOWLEDGMENTS

This research was financially supported by VENI grant 91611178 of the Netherlands Organization of Scientific Research (NWO). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## FUNDING INFORMATION

This work, including the efforts of Wendy W. J. van de Sande, was funded by Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO) (91611178).

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## REFERENCES

1. Van de Sande WWJ. 2013. Global burden of human mycetoma: a systematic review and meta-analysis. *PLoS Negl Trop Dis* 7:e2550. <http://dx.doi.org/10.1371/journal.pntd.0002550>.
2. Vera-Cabrera L, Ortiz-Lopez R, Elizondo-Gonzalez R, Ocampo-Candiani J. 2013. Complete genome sequence analysis of *Nocardia brasiliensis* HUJEG-1 reveals a saprobic lifestyle and the genes needed for human pathogenesis. *PLoS One* 8:e65425. <http://dx.doi.org/10.1371/journal.pone.0065425>.
3. Vera-Cabrera L, Ortiz-Lopez R, Elizondo-Gonzalez R, Perez-Maya AA, Ocampo-Candiani J. 2012. Complete genome sequence of *Nocardia brasiliensis* HUJEG-1. *J Bacteriol* 194:2761–2762. <http://dx.doi.org/10.1128/JB.00210-12>.
4. Kirby R, Sangal V, Tucker NP, Zakrzewska-Czerwinska J, Wierzbicka K, Herron PR, Chu C-J, Chandra G, Fahal AH, Goodfellow M, Hoskisson PA. 2012. Draft genome sequence of the human pathogen *Streptomyces somaliensis*, a significant cause of actinomycetoma. *J Bacteriol* 194:3544–3545. <http://dx.doi.org/10.1128/JB.00534-12>.
5. Vera-Cabrera L, Ortiz-Lopez R, Elizondo-González R, Campos-Rivera MP, Gallardo-Rocha A, Molina-Torres CA, Ocampo-Candiani J. 2014. Draft genome sequence of *Actinomadura madurae* LIID-AJ290, isolated from a human mycetoma case. *Genome Announc* 2(2):e00201-14. <http://dx.doi.org/10.1128/genomeA.00201-14>.
6. Ahmed AO, Mukhtar MM, Kools-Sijmons M, Fahal AH, de Hoog S, van den Ende BG, Zijlstra EE, Verbrugh H, Abugroun ES, Elhassan AM, van Belkum A. 1999. Development of a species-specific PCR-restriction fragment length polymorphism analysis procedure for identification of *Madurella mycetomatis*. *J Clin Microbiol* 37:3175–3178.
7. Sommer DD, Delcher AL, Salzberg SL, Pop M. 2007. Minimus: a fast, lightweight genome assembler. *BMC Bioinformatics* 8:64. <http://dx.doi.org/10.1186/1471-2105-8-64>.
8. Miller JR, Delcher AL, Koren S, Venter E, Walenz BP, Brownley A, Johnson J, Li K, Mobarry C, Sutton G. 2008. Aggressive assembly of pyrosequencing reads with mates. *Bioinformatics* 24:2818–2824. <http://dx.doi.org/10.1093/bioinformatics/btn548>.
9. English AC, Richards S, Han Y, Wang M, Vee V, Qu J, Qin X, Muzny DM, Reid JG, Worley KC, Gibbs RA. 2012. Mind the gap: upgrading genomes with Pacific Biosciences RS long-read sequencing technology. *PLoS One* 7:e47768. <http://dx.doi.org/10.1371/journal.pone.0047768>.
10. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Pribelski 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. <http://dx.doi.org/10.1089/cmb.2012.0021>.
11. Boetzer M, Pirovano W. 2014. SSPACE-LongRead: scaffolding bacterial draft genomes using long read sequence information. *BMC Bioinformatics* 15:211. <http://dx.doi.org/10.1186/1471-2105-15-211>.
12. Holt C, Yandell M. 2011. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics* 12:491. <http://dx.doi.org/10.1186/1471-2105-12-491>.
13. Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, White O, Buell CR, Wortman JR. 2008. Automated eukaryotic gene structure annotation using EVIDENCEModeler and the Program to Assemble Spliced Alignments. *Genome Biol* 9:R7. <http://dx.doi.org/10.1186/gb-2008-9-1-r7>.
14. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* 10:421. <http://dx.doi.org/10.1186/1471-2105-10-421>.
15. Apweiler R, Bairoch A, Wu CH, Barker WC, Boeckmann B, Ferro S, Gasteiger E, Huang H, Lopez R, Magrane M, Martin MJ, Natale DA, O'Donovan C, Redaschi N, Yeh L-SL. 2004. UniProt: the universal protein knowledgebase. *Nucleic Acids Res* 32(suppl 1):D115–D119. <http://dx.doi.org/10.1093/nar/gkh131>.
16. Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong S-Y, Lopez R, Hunter S. 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30:1236–1240. <http://dx.doi.org/10.1093/bioinformatics/btu031>.
17. Parra G, Bradnam K, Korf I. 2007. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* 23:1061–1067. <http://dx.doi.org/10.1093/bioinformatics/btm071>.
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. <http://dx.doi.org/10.1093/bioinformatics/btv351>.
19. Emms DM, Kelly S. 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol* 16:157. <http://dx.doi.org/10.1186/s13059-015-0721-2>.