



Long-read Sequencing and de novo Genome Assembly of Three *Aspergillus fumigatus* Genomes

Received: 24 March 2023 / Accepted: 21 April 2023 / Published online: 25 May 2023
© The Author(s) 2023

Abstract *Aspergillus fumigatus* is a genetically diverse fungal species, which is near ubiquitous in its global distribution and is the major cause of the life-threatening disease invasive aspergillosis. We present 3 de novo genome assemblies that were selected to be representative of the genetic diversity of clinical and environmental *A. fumigatus*. Sequencing using long-read Oxford Nanopore and subsequent assembly of the genomes yielded 10–23 contigs with an N50 of 4.05 Mbp to 4.93 Mbp.

Keywords *Aspergillus fumigatus* · Oxford nanopore sequencing · de novo genome assembly · Fungal genome · Azole resistance

Introduction

Aspergillus fumigatus is a globally ubiquitous environmental mould that was recently highlighted in the World Health Organization (WHO) fungal priority pathogens list as a species of critical concern [1]. *A. fumigatus* can cause invasive and chronic forms of the disease aspergillosis which results in more than 300,000 deaths per year [2]. Unfortunately, resistance of *A. fumigatus* to triazole antifungals (the first-line therapy for aspergillosis) is emerging worldwide [3].

Previous phylogenomic analysis has shown that the population of *A. fumigatus* is genetically diverse and

clusters into two clades (A and B) [4, 5]. This extensive genetic diversity provides ample opportunity for new drug resistance polymorphisms to arise. However, the current reference genomes, Af293 [6] and A1163 [7] do not span the existing known diversity. To assist in investigating why most environmental triazole resistance occurs in clade A, we have resequenced three isolates from our laboratories in-house *A. fumigatus* collection that are representative of the main diversity of *A. fumigatus* [4] (Fig. 1). Sequencing was achieved using deep nanopore sequencing to generate de novo assemblies of two clade A isolates (one of which contains the predominant resistance allele TR₃₄/L98H) and a single clade B isolate. Although at time of original submission, there are currently 321 *A. fumigatus* isolates available on NCBI [8] the 3 de novo assemblies we present here are assembled into fewer contigs than > 99%. Moreover, these genomes were sequenced cheaply and inhouse using long-read sequencing and we provide a freely available, downloadable bioinformatic pipeline for research groups who also wish to produce de novo genome assemblies of fungal species with small genomes from long-read sequencing data (<https://github.com/SJHemmings/afasont>).

Method

The isolates selected for sequencing were C6 (a clinical wildtype isolate from clade A, U.K.), C87 (a clinical isolate from clade A with resistant TR₃₄/L98H

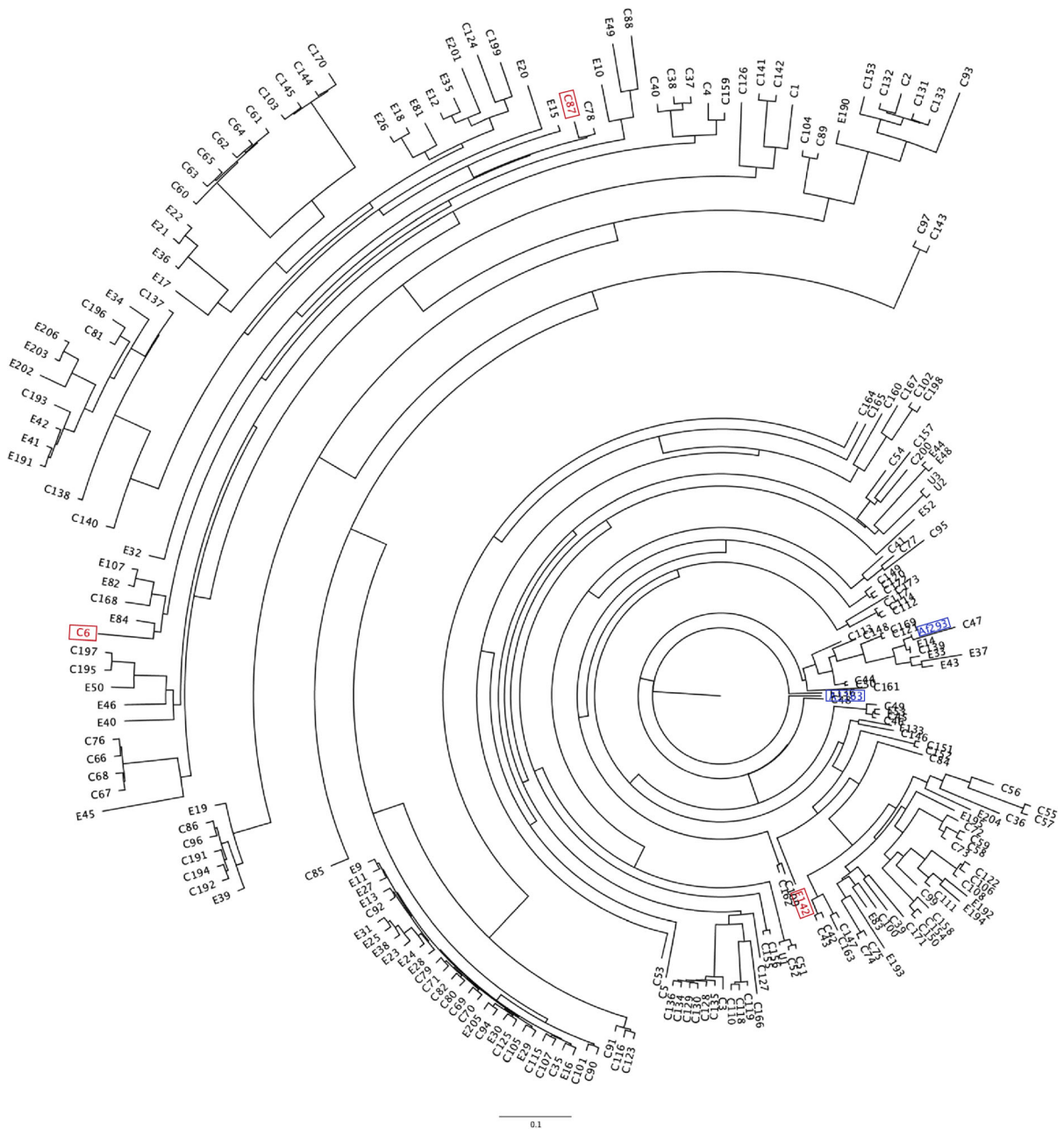


Fig. 1 Unrooted maximum likelihood phylogenetic tree (constructed in RAxML using genome-wide SNPs) of *Aspergillus fumigatus* isolates in the UK and Republic of Ireland. The reference genomes Af293 and A1163 are highlighted in blue and C6, C87 and E142 are highlighted in red.

allele, U.K.) and E142 (an environmental wildtype isolate from clade B, U.S.).

A. fumigatus isolates were inoculated in vented 25 cm³ tissue culture flasks with Sabouraud Dextrose agar

Branch lengths refer to the mean number of substitutions per site. Average nucleotide identity (ANI) between Af293 and A1163 was calculating to be 99.69% while ANI for the two Clade A isolates (C6, C87) compared to the Clade B isolate (E142) were calculated to be 99.64% and 99.66% respectively

(Oxoid, Hampshire, U.K.) and incubated for 48 h at 37 °C. Spores were harvested in PBS + 0.01% Tween-20 by filtration through glass wool (Thermo Fisher Scientific, Massachusetts, U.S.). Spores were

centrifuged (5000 rpm for 10 min) and resuspended in Yeast Cell Lysis Solution (Biosearch Technologies, Hoddesdon, U.K.) and vortexed at maximum speed for 10 min with 1.0 mm zirconia/silica beads (Thistle Scientific, Glasgow, U.K.). The suspension was then centrifuged (14,000 rpm for 2 min) and supernatant was removed and treated with RNase Cocktail™ Enzyme Mix (Thermo Fisher Scientific) according to the manufacturer's instruction. DNA was then isolated on spin columns using AW1 and AW2 wash buffers (Qiagen, Venlo, Netherlands) and eluted in nuclease free water. To achieve the required relative absorption ratios for Oxford Nanopore sequencing and to concentrate DNA, additional washing steps were carried out using 0.6X AMPure Reagent (Beckman Coulter, California, U.S.) and 70% ethanol. An SRE XS kit (Pacific Biosciences, California, U.S.) was used to deplete any remaining reads below 10 kb. For quality control, DNA was visualised using Genomic DNA Screentape on a TapeStation (Agilent Technologies, California, U.S.) to ensure the average DNA length was above 20 kbp. 1 µg of DNA was prepared for sequencing using an SQK-LSK110 ligation sequencing kit (Oxford Nanopore Technologies, Oxford, U.K.) and NEBNext Companion Module (New England Biolabs, Massachusetts, U.S.) following the manufacturer's instructions. Isolates were sequenced on a minION using an R10.4 flow cell (Oxford Nanopore Technologies) for a total of 18 h.

Live base calling was performed using Guppy v6.3.9. Porechop v0.2.4 [9] and NanoLyse v1.2.1 [10] were used to remove adapters and sequences and CS DNA (Oxford Nanopore Technologies) from raw fastq files. Reads were filtered using NanoFilt v2.8.0 [10] to remove reads with a quality score below Q10 or less than 1 kbp in length. Reads passing quality control were used for de novo assembly using Canu v2.2 [11] with a specified genome length of 29 Mbp. The assembled genomes were polished with Pilon v1.24 [12] using Illumina paired-end reads (150 bp) sequenced on a NovaSeq 6000 SP v1.5 (185X coverage for C6, 29X coverage for C87 and 51X coverage for E142) at the Earlham Institute (UK). Illumina paired-end reads can be accessed from the European Nucleotide Archive at EMBL-EBI under accession code PRJEB27135. The number of tRNA and protein coding genes within the assemblies were then estimated using tRNAscan-SE v2.0.9 [13] and AUGUSTUS v3.8.0 [14]. Genome completeness was then predicted

using BUSCO coupled with the ascomycota_odb10 lineage dataset [15]. The full pipeline ('afasont') used to generate these assemblies is available from <https://github.com/SJHemmings/afasont>. Finally, BLAST v2.12.0 + [16] analysis was used to screen the individual contigs for contamination.

Genome Details

After passing through quality control, raw reads from the minION showed coverage of 82X for C6, 73X for C87 and 149X for E142. Isolates C6, C87 and E142 were then assembled into genomes of 29,266,253 bp, 28,591,451 bp and 28,644,426 bp in length.

C6 was assembled into 10 contigs with an N50 of 3.99 Mbp with a longest contig of 4.93 Mbp. AUGUSTUS [14] predicted 8,802 protein coding genes and tRNAscan-SE [13] detected 211 genes which encode for transfer RNA. Using the ascomycota_odb10 lineage dataset, BUSCO estimated genome completeness to be 98.5% [15].

C87 assembled into 23 contigs with an N50 of 2.55 Mbp, the longest contig reached 4.05 Mbp in length. AUGUSTUS [14] found 8830 protein coding genes and tRNAscan-SE [13] estimated there were 200 genes which encode for tRNA. Genome completeness was predicted to be 97.6% using BUSCO [15].

E142 was assembled into 15 contigs and has a N50 of 2.67 Mbp, the longest contig was 4.20 Mbp. 8,806 protein coding genes were found by AUGUSTUS [14] and tRNAscan-SE [13] detected 208 genes which encode for tRNA. 98.1% genome completeness was estimated using BUSCO [15].

Raw read files sequenced with Oxford Nanopore Technologies and de novo genome assemblies can be accessed from the European Nucleotide Archive at EMBL-EBI under the accession code PRJEB59410. The sequence accessions for the individual assemblies are: CASBLW01 (GCA_949125545) (C6); CASBLU01 (GCA_949125165) (E142); CASBLV01 (GCA_949125185) (C87).

Acknowledgements We thank all reviewers for taking the time to assess our manuscript. We also acknowledge funding from the Wellcome Trust and Natural Environmental Research Council (NERC). We also thank Dr S. R. Lockhart (CDC Mycotic Diseases Branch) for provision of the isolate E142.


Author contributions All authors contributed to the study conception, design and writing of the manuscript. Material preparation and data collection was performed by SJH. Analysis was performed by SJH & JLR.


Funding Supported by the Wellcome Trust Collaborative Awards in Science grant: “Understanding and mitigating the impact of emerging antifungal resistance” & NERC: “Understanding the eco-evolutionary drivers of emerging antifungal resistance”. MCF is a CIFAR Fellow in the ‘Fungal Kingdom’ programme.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Author name and affiliation

¹Samuel J. Hemmings 

²Johanna L. Rhodes 

¹Matthew C. Fisher 

¹Department of Infectious Disease Epidemiology, Imperial College London, London, UK
e-mail: s.hemmings@imperial.ac.uk

²Department of Medical Microbiology, Radboud University Medical Centre, Nijmegen, Netherlands

References

1. WHO fungal priority pathogens list to guide research, Development and Public Health Action (2022) World Health Organization. World Health Organization. 2022. <https://www.who.int/publications/i/item/9789240060241> Accessed 17 Jan 2023.
2. Bongomin F, Gago S, Oladele RO, Denning DW. Global and multi-national prevalence of fungal diseases—estimate precision. *J Fungi*. 2017;3(4):57.
3. Fisher MC, Hawkins NJ, Sanglard D, Gurr SJ. Worldwide emergence of resistance to antifungal drugs challenges human health and food security. *Science*. 2018;360(6390):739–42.
4. Rhodes J, Abdolrasouli A, Dunne K, Sewell TR, Zhang Y, Ballard E, Brackin AP, van Rhijn N, Chown H, Tsit-sopoulou A, Posso RB. Population genomics confirms acquisition of drug-resistant *Aspergillus fumigatus* infection by humans from the environment. *Nat Microbiol*. 2022;7(5):663–74.
5. Sewell TR, Zhu J, Rhodes J, Hagen F, Meis JF, Fisher MC, Jombart T. Nonrandom distribution of azole resistance across the global population of *Aspergillus fumigatus*. *MBio*. 2019;10(3):e00392-e419.
6. Nierman WC, Pain A, Anderson MJ, Wortman JR, Kim HS, Arroyo J, Berriman M, Abe K, Archer DB, Bermejo C, Bennett J. Genomic sequence of the pathogenic and allergenic filamentous fungus *Aspergillus fumigatus*. *Nature*. 2005;438(7071):1151–6.
7. Fedorova ND, Khaldi N, Joardar VS, Maiti R, Amedeo P, Anderson MJ, Crabtree J, Silva JC, Badger JH, Albarraq A, Angiuoli S. Genomic islands in the pathogenic filamentous fungus *Aspergillus fumigatus*. *PLoS Genet*. 2008;4(4):e1000046.
8. *Aspergillus fumigatus* genome list - genome - NCBI. National Center for Biotechnology Information <https://www.ncbi.nlm.nih.gov/genome/browse/#!/eukaryotes/18/>. Accessed 2 Feb 2023.
9. Wick RR, Judd LM, Gorrie CL, Holt KE. Completing bacterial genome assemblies with multiplex MinION sequencing. *Microb Genomics*. 2017;3(10):e000132.
10. De Coster W, D’hert S, Schultz DT, Cruts M, Van Broeckhoven C. NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics*. 2018;34(15):2666–9.
11. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res*. 2017;27(5):722–36.
12. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE*. 2014;9(11):e112963.
13. Chan PP, Lin BY, Mak AJ, Lowe TM. tRNAscan-SE 2.0: improved detection and functional classification of transfer RNA genes. *Nucleic Acids Res*. 2021;49(16):9077–96.
14. Stanke M, Diekhans M, Baertsch R, Haussler D. Using native and syntenically mapped cDNA alignments to improve *de novo* gene finding. *Bioinformatics*. 2008;24(5):637–44.
15. Seppey M, Manni M, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness. *Methods Mol Biol*. 2019;1962:227–45. https://doi.org/10.1007/978-1-4939-9173-0_14.
16. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. BLAST+: architecture and applications. *BMC Bioinform*. 2009;10:1–9.

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.