Draft Genome Sequences of Four *Aspergillus* Section *Fumigati* Clinical Strains

Renato Augusto Corrêa dos Santos,a,b Olga Rivero-Menéndez,c Jacob L. Steenwyk,b Matthew E. Mead,b Gustavo Henrique Goldman,a Ana Alastruey-Izquierdo,c Antonis Rokasb

aDepartmento de Ciências Farmacêuticas, Faculdade de Ciências Farmacêuticas de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, São Paulo, Brazil bDepartment of Biological Sciences, Vanderbilt University, Nashville, Tennessee, USA cMedical Mycology Reference Laboratory, National Center for Microbiology, Instituto de Salud Carlos III, Madrid, Spain

**ABSTRACT** *Aspergillus* fungi in section *Fumigati* include important human pathogens. Here, we sequenced the genomes of two strains of *Aspergillus hiratsukae* and two strains of *Aspergillus felis*. The average genome sizes are 29.5 Mb for *A. hiratsukae* and 31.8 Mb for *A. felis*.

Aspergillus is a highly diverse genus of industrially and medically important fungi (1, 2). The genus is taxonomically divided into 27 sections (3). Section *Fumigati* contains the major human pathogen *Aspergillus fumigatus* (4) and several so-called cryptic species, such as *Aspergillus hiratsukae* and *Aspergillus felis* (5–7), which are morphologically similar but genetically distinct from *A. fumigatus*. Cryptic species account for over 10% of cases of *Aspergillus* infection (8). Here, we sequenced the genomes of two clinical strains of *A. hiratsukae*, CNM-CM5793 and CNM-CM6106, from nail and ear infections, respectively, both from Spain. We also sequenced two clinical strains of *A. felis*, strain CNM-CM7691 from an ear infection in Spain and strain CNM-CM5623 from Portugal. All four isolates were recovered from clinical samples following standard procedures and sent to the Medical Mycology Reference Laboratory (at the National Center for Microbiology, Instituto de Salud Carlos III, Madrid, Spain) for identification and susceptibility testing. Except for infection type, no clinical data were recorded. Therefore, the fungal isolates were judged to be exempt from informed consent of the patients and institutional review board approval.

Species assignment was based on a maximum-likelihood phylogenetic analysis (Fig. 1). For genome sequencing, we grew all strains in glucose-yeast extract-peptone (GYEP) liquid medium (0.3% yeast extract and 1% peptone; Difco, Soria Melguizo) with 2% glucose (Sigma-Aldrich, Spain) for 24 to 48 h at 30°C. The mycelium was mechanically disrupted by vortex mixing with glass beads and used to extract genomic DNA using the phenol-chloroform method (9). DNA was quantified using the Quantifluor double-stranded DNA (dsDNA) system and the Quantifluor ST fluorometer (Promega, Madison, WI, USA). DNA quality was checked with the Agilent 2100 bioanalyzer (Agilent Technologies, Inc., Santa Clara, CA, USA). DNA libraries were prepared using the Nextera DNA library prep kit (Illumina, Inc., San Diego, CA, USA) according to the manufacturer’s guidelines. Paired-end sequencing (2 × 150 bp) was performed using the NextSeq 500 platform following the manufacturer’s protocols (Illumina, Inc.).

For all software, default parameters were used except where otherwise noted. The numbers of sequencing read pairs generated for strains CNM-CM5793, CNM-CM6106, CNM-CM7691, and CNM-CM5623 were 7,733,508, 5,237,901, 9,555,248, and 6,768,577, respectively. Quality control of the sequence reads was performed with FastQC v0.11.7 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Raw reads were cleaned with Trimmomatic v0.38 (10) with the following parameters: NexteraPE-PE.fa:2:30:10:2:


Editor Vincent Bruno, University of Maryland School of Medicine

Copyright © 2020 dos Santos et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license.

Address correspondence to Gustavo Henrique Goldman, ggoldman@usp.br, or Antonis Rokas, antonis.rokas@vanderbilt.edu.

Received 25 August 2020
Accepted 9 September 2020
Published 1 October 2020
The genome sequences were assembled with SPAdes v3.14.0 employing multiple k-mers (31, 41, 51, 61, 71, 81, and 91) and the --careful parameter. The genomic reads were mapped to the assembly with Bowtie v2.3.4.1, followed by a single iteration of tree inference on IQ-TREE v2.0.3 with partitions with the option “MFP+MERGE,” which employs ModelFinder to find the best partition scheme. The final tree was edited in FigTree v1.4.4.

Support values are based on 1,000 bootstrap replicates. A. clavatus (section Clavati) was used to root the tree. Note that the species A. parafelis, A. pseudofelis, and A. felis were merged (synonymized) into a single species, A. felis (18); thus, we infer that the two sequenced strains belong to A. felis.

FIG 1 Maximum-likelihood phylogenetic tree of the four strains sequenced in this announcement (in bold) and related species in section Fumigati, based on the analysis of the four markers beta-tubulin gene (benA), calmodulin gene (CaM), actin gene (act), and RNA polymerase II second-largest subunit gene (RPB2), commonly used in Aspergillus taxonomy (18); sequences were obtained from reference 18 except for the sequences of the four newly sequenced strains, which were obtained by searching for markers of each strain in orthogroups generated by OrthoFinder v2.3.3 (19) using A. fumigatus AF293 (17) as the reference. Each marker was aligned with MAFFT v7.397, and a supermatrix was generated with FASconCAT v1.11. Tree inference was carried out on IQ-TREE v2.0.3 with partitions with the option “MFP+MERGE,” which employs ModelFinder to find the best partition scheme. The final tree was edited in FigTree v1.4.4. Support values are based on 1,000 bootstrap replicates. A. clavatus (section Clavati) was used to root the tree. Note that the species A. parafelis, A. pseudofelis, and A. felis were merged (synonymized) into a single species, A. felis (18); thus, we infer that the two sequenced strains belong to A. felis.
Overall genome assembly, completeness, and annotation statistics

<table>
<thead>
<tr>
<th>Strain</th>
<th>Assembly sizea (bp)</th>
<th>No. of contigs &gt;1,000 bp</th>
<th>Avg genome coverage (X)</th>
<th>GC content (%)</th>
<th>N50 (bp)</th>
<th>No. of genes</th>
<th>No. (%) of complete single-copy BUSCOs</th>
<th>No. (%) of fragmented BUSCOs</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. hiratsukae CNM-CM5793</td>
<td>29,562,918</td>
<td>745</td>
<td>63</td>
<td>50.38</td>
<td>100,935</td>
<td>9,685</td>
<td>3,487 (98.33)</td>
<td>31 (0.87)</td>
</tr>
<tr>
<td>A. hiratsukae CNM-CM6106</td>
<td>29,374,270</td>
<td>922</td>
<td>39</td>
<td>50.37</td>
<td>71,695</td>
<td>9,663</td>
<td>3,466 (97.74)</td>
<td>38 (1.07)</td>
</tr>
<tr>
<td>A. felis CNM-CM5623</td>
<td>31,643,783</td>
<td>663</td>
<td>47</td>
<td>49.93</td>
<td>112,776</td>
<td>10,161</td>
<td>3,494 (98.53)</td>
<td>26 (0.73)</td>
</tr>
<tr>
<td>A. felis CNM-CM7691</td>
<td>31,957,614</td>
<td>559</td>
<td>70</td>
<td>49.93</td>
<td>138,232</td>
<td>10,243</td>
<td>3,503 (98.78)</td>
<td>18 (0.51)</td>
</tr>
</tbody>
</table>

aBased on contigs with more than 1,000 bp.

ACKNOWLEDGMENTS

This research was funded by Brazilian São Paulo Research Foundation (FAPESP) grant 2016/07870-9 (to G.H.G.) and scholarships 2017/21983-3 and 2019/07526-4 (to R.A.C.D.S.) and by research projects from the Fondo de Investigación Sanitaria (PI13/02145 and PI16/CIII/00035) of the Instituto de Salud Carlos III. J.L.S. and A.R. were funded by the Howard Hughes Medical Institute through the James H. Gilliam Fellowships for Advanced Study program. A.R. also received funding from the National Institutes of Health/National Institute of Allergy and Infectious Diseases (1R56AI146096-01A1).

Computational infrastructure was provided by the Advanced Computing Center for Research and Education (ACCRE) at Vanderbilt University.

REFERENCES