Chromosome-Level *Aspergillus flavus* Strain CA14 Genome Assembly

Mark A. Weaver,a Brian M. Mack,b Matthew K. Gilbertb

aBiological Control of Pests Research Unit, National Biological Control Laboratory, USDA ARS, Stoneville, Mississippi, USA
bFood and Feed Safety Unit, USDA ARS, New Orleans, Louisiana, USA

**ABSTRACT**  We report here a chromosome-level genome assembly of the *flatoxic* fungus *Aspergillus flavus* strain CA14. This strain is the basis for numerous studies in fungal physiology and secondary metabolism. This full-length assembly will aid in subsequent genomics research.

The filamentous fungus *Aspergillus flavus* is an opportunistic pathogen of several crops, including corn, cotton, and peanut, and colonizes many stored grains intended for food and feed. It may be best known as a producer of aflatoxins, the most potent carcinogens found in nature (1). Isolates of *A. flavus* within a field population vary greatly with regard to aflatoxin production and microsatellite-based haplotype (2, 3). Additionally, numerous species-specific genes and genomic rearrangements have been documented within the *A. flavus* clade of *Aspergillus* section Flavi (4). Until recently, the reference genome for *A. flavus* was the Sanger-sequenced assembly of strain 3357, with 5× coverage and 331 scaffolds (5) (Table 1). Presently, >100 sequenced genomes have been made publicly available for *A. flavus*, created using various sequencing technologies or combinations of available systems (https://www.ncbi.nlm.nih.gov/genome/browse#%21/overview/aspergillus%20flavus). A pseudomolecule-level assembly of strain 3357 and AF13 (6) and full chromosome-level assembly of strain 3357 (7) were also reported in 2020 (Table 1). We report here the sequence of strain CA14, assembled at the chromosome level. Strain CA14 is a wild-type, large-sclerotia-producing, *flatoxic* strain isolated from pistachio at the Wolfskill Grant Experimental Farm of University of California, Davis (8), that has been used in numerous knockout and functional studies of *A. flavus* isolates (9, 10).

Genomic DNA (gDNA) was extracted from ~10⁷ conidia of *A. flavus* strain CA14 by grinding in liquid nitrogen and extracting in hot cetyltrimethylammonium bromide buffer (11). After gDNA cleanup, creation of a PacBio Express library, and size selection,

**TABLE 1** Assembly statistics of some representative *Aspergillus flavus* genomes

<table>
<thead>
<tr>
<th>Aspergillus flavus strain</th>
<th>GenBank assembly accession no.</th>
<th>Genome size (Mb)</th>
<th>Fold coverage (x)</th>
<th>No. of contigs/scaffolds</th>
<th>N₅₀ (Mb)</th>
<th>No. of genes predicted</th>
</tr>
</thead>
<tbody>
<tr>
<td>3357a</td>
<td>GCA_000006275.2</td>
<td>36.89</td>
<td>5</td>
<td>331</td>
<td>2.39</td>
<td>13,485</td>
</tr>
<tr>
<td>3357b</td>
<td>GCA_009017415.1</td>
<td>37.75</td>
<td>600</td>
<td>8</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>3357c</td>
<td>GCA_014117465.1</td>
<td>37.0</td>
<td>70</td>
<td>8</td>
<td>2.40</td>
<td></td>
</tr>
<tr>
<td>AF13b</td>
<td>GCA_014117485.1</td>
<td>37.44</td>
<td>60</td>
<td>8</td>
<td>2.39</td>
<td></td>
</tr>
<tr>
<td>KuPG1d</td>
<td>GCA_003709025.1</td>
<td>37.7</td>
<td>121</td>
<td>199</td>
<td>1.3</td>
<td>12,846</td>
</tr>
<tr>
<td>CA14</td>
<td>GCA_014784225.1</td>
<td>37.81</td>
<td>140</td>
<td>8</td>
<td>6.27</td>
<td></td>
</tr>
</tbody>
</table>

a From reference 5.
b From reference 7.
c From reference 6.
d Chang et al. (12), derived from isolate CA14.
e Pseudochromosomes inferred from alignment to *Aspergillus oryzae* strain RIB40.

Editor Christina A. Cuomo, Broad Institute
This is a work of the U.S. Government and is not subject to copyright protection in the United States. Foreign copyrights may apply.
Address correspondence to Mark A. Weaver, mark.weaver@usda.gov.
Received 1 October 2020
Accepted 1 December 2020
Published 7 January 2021
6.75 Gb of sequence (140× coverage) was generated from a single 10-hour single-molecule real-time (SMRT) cell using PacBio Sequel. De novo genome assembly was done using Flye (version 2.5, settings at “genome size 37m\textsuperscript{-t}14”) (https://github.com/fenderglass/Flye) followed by three rounds of polishing with Arrow (version 2.3, default settings) (https://github.com/PacificBiosciences/GenomicConsensus) yielding 8 full chromosome-length scaffolds with an N\textsubscript{50} value of 6.27 Mb. Nine to 11 telomere repeat sequences were present on six of the contig ends.

Data availability. The GenBank accession number is GCA_014784225.1, and the SRA accession number is SRR12683076.

ACKNOWLEDGMENTS

We thank Perng-Kuang Chang for providing strain CA14 and sharing data from strain KuPG#1.

We thank the University of Minnesota Genomic Center for helpful discussion and technical services.

REFERENCES


