Whole Genome Sequences of 20 *Aspergillus Flavus* Isolates from Corn Kernels and Cornfield Soils in Louisiana

Solomon T. Gebru[a], Mark K. Mammel[a] Jayanthi Gangiredla[a], Carmen Tartera[a], Jeffrey W. Cary[b], Geromy G. Moore[b], Rebecca R. Sweany[b]

[a]Division of Molecular Biology, Office of Applied Research and Safety Assessment, Center for Food Safety and Applied Nutrition, U.S. Food and Drug Administration, Laurel, MD, 20708, USA

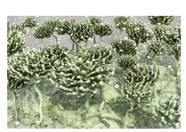
[b]Food and Feed Safety Research Unit, Southern Regional Research Center (SRRC), U.S. Department of Agriculture (USDA), New Orleans, LA, 70124, USA

Abstract

Aspergillus flavus is an opportunistic pathogen that infects plants, animals and humans, and produces the carcinogenic mycotoxin, aflatoxin. Aflatoxin consumed in grains or milk is acutely toxic, stunts children's growth and induces liver cancer. The fungus may also cause allergic reactions in humans and aspergillosis diseases, notably in lungs of immunocompromised people. Therefore, the occurrence of aflatoxin in the food chain is a public health concern and it is crucial to develop a method to identify this fungal contaminant in foods. The aim of this study was to develop a rapid means of identifying this fungal contaminant using whole genome sequencing. *Aspergillus flavus* populations are very diverse and may consist of many different vegetative compatibility groups (VCGs). Twenty isolates encompassing all 16 VCGs were selected for genome sequencing. After overnight growth of these isolates, mycelial pellets were collected, flash-frozen in liquid nitrogen, freeze-dried, and genomic DNA was extracted. Paired-end sequencing libraries were prepared using Nextera XT and sequenced on an Illumina Nextseq. The sequences were assembled using the SPAdes assembler v3.12.0, and the assembly quality assessment utilized QUAST. The sequences were identified to the species level, using an MLST and custom kmer database and were submitted in NCBI under Bio Project PRJNA482816. There are only a few publicly available fungal genomes sequences from food isolates. The availability of high-quality genome sequences for fungi will allow detection and identification of these potential pathogens and creation of an accurate catalog of mycotoxins.

Introduction

Aspergillus flavus is an opportunistic pathogen that infects plants, animals and humans, and produces aflatoxin, a natural carcinogen. Aflatoxin consumed in grains or milk is acutely toxic, stunts children's growth and induces liver cancer (1). The fungus may also cause allergic reactions in humans and aspergillosis diseases, notably in lungs of immunocompromised people (2). Aflatoxin contamination in corn threatens both consumer food safety and grower economic stability, causing an estimated $52 million to over $1 billion of economic loss per year in the U.S. (3). In Louisiana, a severe corn aflatoxin outbreak in 1998 resulted in an almost total crop loss, and periodic aflatoxin outbreaks still occur (4). The aim of this study was to develop a rapid means of identifying this fungal contaminant using whole genome sequencing.



Materials and Methods

Strains and DNA extraction

Aspergillus flavus populations are very diverse and may consist of many different vegetative compatibility groups (VCGs). Twenty isolates encompassing all 16 VCGs were selected for genome sequencing. 75 ml of potato dextrose broth conidial cultures were shaken overnight at 125 rpm and 30 °C. Mycelial pellets were collected, flash-frozen with liquid nitrogen, freeze-dried, and then DNA was extracted with a Zymo Quick-DNA Fungal/Bacterial DNA Miniprep kit (Zymo Research, Irvine, CA). Genomic DNA (gDNA) quality and purity were assessed using a Nanodrop spectrophotometer (Thermo Scientific, Wilmington, DE), and quality control was performed using a Qubit 2.0 fluorometer (Life Technologies, Burlington, Canada).

Whole genome sequencing

Whole genome sequencing was performed using a Nextera XT DNA Library Prep Kit (Illumina, Inc., San Diego, CA) with 2 x150 bp paired-end sequencing on an Illumina NextSeq Sequencer with a NextSeq 500/550 Mid output reagent cartridge V2 (n = 8).

Strain Name/Genome ID	Organism	Tax ID	Isolation Source
MOD1-573	*Aspergillus flavus*	5059	Corn
MOD1-575	*Aspergillus flavus*	5059	Corn
MOD1-576	*Aspergillus flavus*	5059	Corn
MOD1-578	*Aspergillus flavus*	5059	Corn
MOD1-580	*Aspergillus flavus*	5059	Corn
MOD1-581	*Aspergillus flavus*	5059	Corn
MOD1-584	*Aspergillus flavus*	5059	Corn field soil
MOD1-586	*Aspergillus flavus*	5059	Corn field soil
MOD1-587	*Aspergillus flavus*	5059	Corn field soil
MOD1-590	*Aspergillus flavus*	5059	Corn field soil
MOD1-591	*Aspergillus flavus*	5059	Corn field soil
MOD1-595	*Aspergillus flavus*	5059	Corn field soil
MOD1-599	*Aspergillus flavus*	5059	Corn field soil
MOD1-601	*Aspergillus flavus*	5059	Corn field soil
MOD1-605	*Aspergillus flavus*	5059	Corn field soil
MOD1-607	*Aspergillus flavus*	5059	Corn field soil
MOD1-618	*Aspergillus flavus*	5059	Corn field soil
MOD1-619	*Aspergillus flavus*	5059	Corn
MOD1-620	*Aspergillus flavus*	5059	Corn field soil
MOD1-621	*Aspergillus flavus*	5059	Corn

Table 1. Strains used in this study and source

Results and Discussion

Fungal sequence assembly and mycotoxin gene identification

Fast QC (Q >30) was used to check the raw sequence data. Low-quality reads were trimmed to a quality threshold of Q>30 using Trimmomatic (5) with NexteraPE adapter file. Trimmed reads were subjected to de novo assembly using the SPAdes assembler v3.12.0 (6), and assembly quality assessment utilized QUAST (7). Default settings were used for all software programs unless otherwise noted. The analysis of the sequence data is shown in Table 2. The sequences were identified to the species level, using an MLST and custom kmer database, as *Aspergillus flavus* species. The raw sequence data and some of the draft genomes were submitted in NCBI under a new project specific for fungi, BioProject PRJNA482816.

Genome ID	Strain/VCG	Aflatoxigenicity	Aflatoxin B_1 production (parts per billion) ±SE[b]	Contigs	Reads	Genome size (bp)	N_{50}	Coverage	G+C (%)	SRA accession no.	GenBank accession no.
MOD1-573	RRS4	Aflatoxigenic	11,900 ± 836	670	34,952,640	37,023,989	132,464	42X	48.3	SRR11596619	JABVYR000000000
MOD1-575	RRS7	Aflatoxigenic	20 ± 6	624	31,401,164	36,757,764	142,633	39X	48.3	SRR11596618	JABVYS000000000
MOD1-576	RRS5	Aflatoxigenic	24,006 ± 3,918	688	34,833,608	37,068,150	126,389	42X	48.2	SRR11596607	JABVYT000000000
MOD1-578	RRS9	Aflatoxigenic	3,872 ± 1,026	652	40,610,742	37,193,948	152,925	48X	48.3	SRR11596606	JABVYU000000000
MOD1-580	RRS1	Aflatoxigenic	Less than 0	683	35,257,442	36,938,434	121,985	42X	48.3	SRR11596605	JABVYV000000000
MOD1-581	RRS10	Non-aflatoxigenic	0 ± 0	723	39,417,907	36,954,782	112,933	39X	48.3	SRR11596604	JABVYW000000000
MOD1-584	RRS11	Aflatoxigenic	1,714 ± 120	595	29,738,152	36,865,287	144,290	41X	48.3	SRR11596603	JABVYX000000000
MOD1-586	RRS3	Aflatoxigenic	1,6162 ± 150	756	35,670,908	37,08,9482	143,643	42X	48.3	SRR11596602	JABVYY000000000
MOD1-587	RRS12	Aflatoxigenic	4,2548 ± 9,686	701	30,328,260	36,985,390	136,341	40X	48.3	SRR11596601	JABVYZ000000000
MOD1-590	RRS2	Aflatoxigenic	14,032 ± 4,858	635	32,785,962	36,946,864	138,751	45X	48.3	SRR11596600	JABVZA000000000
MOD1-591	RRS13	Non-aflatoxigenic	0 ± 0	626	35,283,328	36,626,824	141,548	49X	48.3	SRR11596617	JABVZB000000000
MOD1-595	RRS14	Non-aflatoxigenic	0 ± 0	623	32,152,852	36,915,067	152,851	45X	48.2	SRR11596616	JABVZC000000000
MOD1-599	RRS8	Aflatoxigenic	27,998 ± 9,260	723	38,755,062	37,006,081	152,745	42X	48.2	SRR11596615	JABVZD000000000
MOD1-601	RRS5	Aflatoxigenic	19,114 ± 332	867	40,051,400	37,973,590	170,945	57X	48.2	SRR11596614	JABVZE000000000
MOD1-605	RRS6	Non-aflatoxigenic	0 ± 0	627	36,745,078	36,775,823	155,768	52X	48.2	SRR11596613	JABVZF000000000
MOD1-607	RRS15	Aflatoxigenic	3,968 ± 432	1,167	21,467,712	38,231,283	85,770	30X	48.2	SRR11596612	JABVZG000000000
MOD1-618	RRSSOLO[a]	Aflatoxigenic	12,460 ± 2,650	546	36,203,968	36,963,370	169,107	51X	48.3	SRR11596611	JABVZH000000000
MOD1-619	RRS1	Aflatoxigenic	35	908	39,434,866	36,935,936	129,357	110X	48.3	SRR11596610	JABVZI000000000
MOD1-620	RRS1	Aflatoxigenic	3	1,226	42,922,302	37,016,679	143,589	108X	48.2	SRR11596610	JABVZJ000000000
MOD1-621	RRS1	Aflatoxigenic	3	1,660	37,110,996	36,895,533	138,343	105X	48.3	SRR11596608	JABVZK000000000

Table 2. *Aspergillus flavus strains*, aflatoxin production, and genomic statistics data

a. RRSSOLOa refers to an isolate that did not complement with any other isolates and therefore was in a singleton VCG (vegetative compatibility group).

b. Aflatoxin B1 from glucose-salts medium of 4-day old cultures was quantified with Ultra Performance Liquid Chromatography. AFB2 was detected in lower quantities for each extract with AFB1.

Conclusion

In this study, genomes of 20 fungal isolates from corn and corn field soil were sequenced. We show that whole genome sequences can be used to identify isolates at the species level and note the potential to identify genes from mycotoxin synthesis pathways to determine whether they produce mycotoxins. Therefore, the availability of high-quality genome sequences for fungi will allow detection and identification of these potential pathogens and creation of an accurate catalog of mycotoxins.

References

1. Hedayati MT, Pasqualotto AC, Warn PA, Bowyer P, Denning DW. 2007. *Aspergillus flavus*: human pathogen, allergen and mycotoxin producer. Microbiology 153:1677–1692. https://doi.org/10.1099/mic.0.2007/007641-0.

2. Kowalska A, Walkiewicz K, Kozieł P, Muc-Wierzgoń M. 2017. Aflatoxins: characteristics and impact on human health. Postepy Hig Med Dosw. 71:315-327. https://doi.org/10.5604/01.3001.0010.3816.

3. Mitchell NJ, Bowers E, Hurburgh C, Wu F. 2016. Potential economic losses to the USA corn industry from aflatoxin contamination. Food Additives & Contaminants Part A, Chemistry, Analysis, Control, Exposure & Risk Assessment 33: 540-550. https://doi.org/10.1080/19440049.2016.1138545.

4. Lanclos DY. 2002. Aflatoxin levels low in Louisiana corn. Delta Farm Press: Aug 16.

5. 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.

6. 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.

7. 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.

Contact information: Dr. Solomon Gebru, E-mail: solomon.gebru@fda.hhs.gov, FDA/CFSAN/OARSA/DMB, 8301 Muirkirk Rd, Laurel MD, 20708