2017

Study of the genetic diversity of the aflatoxin biosynthesis cluster in *Aspergillus* section *Flavi* using insertion/deletion markers in peanut seeds from Georgia, USA

Paola C. Faustinelli
National Peanut Research Laboratory, USDA-ARS, Paola.Faustinelli@ars.usda.gov

Edwin R. Palencia
National Peanut Research Laboratory, USDA-ARS

Victor S. Sobolev
USDA-ARS National Peanut Research Laboratory, Victor.Sobolev@usda.gov

Bruce W. Horn
National Peanut Research Laboratory, USDA-ARS, bruce.horn@ars.usda.gov

Hank T. Sheppard
National Peanut Research Laboratory, USDA-ARS, hank.sheppard@ars.usda.gov

See next page for additional authors

MYCOLOGIA
2017, VOL. 109, NO. 2, 200–209
https://doi.org/10.1080/00275514.2017.1307095

Taylor & Francis
Taylor & Francis Group

Study of the genetic diversity of the aflatoxin biosynthesis cluster in *Aspergillus* section *Flavi* using insertion/deletion markers in peanut seeds from Georgia, USA

Paola C. Faustinelli[a], Edwin R. Palencia[a], Victor S. Sobolev[a], Bruce W. Horn[a], Hank T. Sheppard[a], Marshall C. Lamb[a], Xinye M. Wang[a], Brian E. Scheffler[b], Jaime Martinez Castillo[c], and Renée S. Arias[a]

[a]National Peanut Research Laboratory, Agricultural Research Service, United States Department of Agriculture, 1011 Forrester Drive S.E., Dawson, Georgia 39842-0509; [b]Genomics and Bioinformatics Research Unit, Agricultural Research Service, United States Department of Agriculture, 141 Experiment Station Road, Stoneville, Mississippi 38776; [c]Centro de Investigación Científica de Yucatán, A.C. Calle 43 No. 130, Colonia Chuburná de Hidalgo, Mérida, Yucatán 97200, México

ABSTRACT

Aflatoxins are among the most powerful carcinogens in nature. The major aflatoxin-producing fungi are *Aspergillus flavus* and *A. parasiticus*. Numerous crops, including peanut, are susceptible to aflatoxin contamination by these fungi. There has been an increased use of RNA interference (RNAi) technology to control phytopathogenic fungi in recent years. In order to develop molecular tools targeting specific genes of these fungi for the control of aflatoxins, it is necessary to obtain their genome sequences. Although high-throughput sequencing is readily available, it is still impractical to sequence the genome of every isolate. Thus, in this work, the authors proposed a workflow that allowed prescreening of 238 *Aspergillus* section *Flavi* isolates from peanut seeds from Georgia, USA. The aflatoxin biosynthesis cluster (ABC) of the isolates was fingerprinted at 25 InDel (insertion/deletion) loci using capillary electrophoresis. All isolates were tested for aflatoxins using ultra-high-performance liquid chromatography. The neighbor-joining, three-dimension (3D) principal coordinate, and Structure analyses revealed that the *Aspergillus* isolates sampled consisted of three main groups determined by their capability to produce aflatoxins. Group I comprised 10 non-aflatoxigenic *A. flavus*; Group II included *A. parasiticus*; and Group III included mostly aflatoxigenic *A. flavus* and the three non-aflatoxigenic *A. caelatus*. Whole genomes of 10 representative isolates from different groups were sequenced. Although InDels in *Aspergillus* have been used by other research groups, this is the first time that the cluster analysis resulting from fingerprinting was followed by whole-genome sequencing of representative isolates. In our study, cluster analysis of ABC sequences validated the results obtained with fingerprinting. This shows that InDels used here can predict similarities at the genome level. Our results also revealed a relationship between groups and their capability to produce aflatoxins. The database generated of *Aspergillus* spp. can be used to select target genes and assess the effectiveness of RNAi technology to reduce aflatoxin contamination in peanut.

ARTICLE HISTORY
Received 18 January 2017
Accepted 30 January 2017

KEYWORDS
Arachis hypogaea; capillary electrophoresis; cluster analysis; genome sequencing; insertion/deletion; mycotoxin

INTRODUCTION

The genus *Arachis* has a considerable number of species, including *A. hypogaea* (peanut), the most important species for human consumption. The value of its seeds lies mainly in the high oil (42–52%) and protein (25–32%) contents (Roca and Mroginski 1991). Approximately 30 million tons of peanuts are produced annually, with the United States positioned as the third-ranked producer after China and India. In the United States, Georgia is the main producer, with over 49% of the national total (http://nationalpeanutboard.org).

Numerous agricultural products, including pistachio, almond, maize, fig, hazelnut, and peanut, have been reported to contain aflatoxins (Spanjer et al. 2008; Liao et al. 2013). These carcinogenic substances are secondary metabolites produced by multiple species from section *Flavi*, including *Aspergillus flavus* and *A. parasiticus* (Horn 2007). Aflatoxins are mainly associated with acute toxicity (Azziz-Baumgartner et al. 2005), liver cancer (Murugavel et al. 2007), and child growth impairment (Gong et al. 2004). The high susceptibility of peanut to *Aspergillus* spp. makes the crop a potential health risk because of aflatoxin contamination (Wu et al. 2014). Since approximately 5000 million people worldwide are at risk of exposure to aflatoxins (Strosnider et al. 2006), it is critical to control aflatoxins in the food supply (Wang et al. 2001).

Different control methods have been developed to prevent aflatoxin contamination in peanut and other crops. Several strategies, including biocontrol, and

CONTACT Paola Faustinelli Paola.Faustinelli@ars.usda.gov

Color versions of one or more of the figures in this article can be found online at www.tandfonline.com/umyc.
Supplemental data for this article can be accessed on the publisher's Web site.

Published online 16 May 2017

cultural and storage practices (Khlangwiset and Wu 2010), help minimize contamination, but they are costly and not entirely effective (Yu 2012). In the United States alone, control of aflatoxins costs the peanut industry between 25 and 58 million dollars annually (Lamb and Sternitzke 2001; Leidner 2012). In addition, worldwide efforts have not yet produced aflatoxin-resistant breeding lines of peanut (Nigam et al. 2009).

The aflatoxin biosynthesis cluster (ABC) includes 25 identified genes clustered within a 70-kb DNA region. Despite the high level of conservation of gene order and intergenic distances (Ehrlich et al. 2005), deletions have been observed and show different patterns (Chang et al. 2005). Callicott and Cotty (2015) proposed a practical method for monitoring deletions that affect aflatoxin biosynthesis using cluster amplification patterns (CAPs). This multiplex polymerase chain reaction (PCR) amplified insertion/deletion (InDel) regions associated with atoxigenicity in *A. flavus*. Nucleotide sequence data were obtained for characterizing intraspecific variability in this species. The method proposed was neither expensive nor time-consuming. InDel markers, together with a sensitive analytical tool such as fluorescent capillary electrophoresis, would allow quantifying single-base-pair changes to obtain large amounts of high-quality data. These data are valuable in biotechnology programs, especially for RNA interference (RNAi) methods, where the development of fungal resistance in plants is possible by silencing fungal genes through the generation of siRNAs in the plant host (Tinoco et al. 2010; Arias et al. 2015).

Genetic variability in section *Flavi* is expected, especially in this biosynthesis cluster, because natural populations are diverse (Bayman and Cotty 1993), there is evidence of sexual reproduction (Horn et al. 2009a, 2009b) and deletions are common in the ABC (Chang et al. 2005). By sequencing 21 intergenic regions, Carbone et al. (2007) found five distinct recombination blocks within the entire ABC from 24 *A. parasiticus* isolates from a single Georgia field. Moore et al. (2009) also examined patterns of linkage disequilibrium spanning 21 regions in the same ABC but of *A. flavus*, showing that recombination events are randomly distributed across the cluster in this species. They emphasized the importance of population studies to elucidate the specific mechanisms that regulate and maintain aflatoxigenicity. Ehrlich et al. (2003) compared the nucleotide sequences of the *aflJ*/*aflR* genes from different *Aspergillus* species and found variability in elements that may affect the expression of genes required for aflatoxin accumulation. A single nucleotide mutation in the *pksA-nrps* gene was identified as responsible for the lack of production of aflatoxins and cyclopiazonic acid (CPA) in the biocontrol *A. flavus* K49 (Chang et al. 2012). High-throughput DNA sequencing technologies allow analyzing genetic variation at the nucleotide level; however, their cost makes them impractical for analyzing large numbers of fungal isolates. Here, we propose a workflow using InDel markers to identify the predominant aflatoxigenic strains within a particular geographic region, with the objective of sequencing the whole genomes of a few representative isolates and generating an *Aspergillus* spp. database. This genetic information can be used to determine the predominant genetic variants to develop control strategies to reduce aflatoxin contamination in susceptible crops.

MATERIALS AND METHODS

A total of 151 peanut samples were obtained from the Georgia Federal-State Inspection Service (www.gafsis.com). These seeds were collected from the entire state of Georgia, USA, during the 2014 harvest season. For fungal isolation, samples were shelled and approximately 10 mL seed volume was placed into sterile 50-mL centrifuge tubes. Sterile water was added to cover the seeds, and tubes were placed in a KLECO (Visalia, California, USA) tissue pulverizer for 2 min. The liquid obtained was plated on two modified dichloran Rose Bengal (MDRB) agar plates (50 and 100 µL each) and incubated at 37 C for 72 h (Horn and Dorner 1999). Spores from *Aspergillus* section *Flavi* colonies were streak plated on fresh MDRB plates and then transferred to Czapek-Dox (OXOID, Hampshire, UK) agar slants (Raper and Fennell 1965) for identification according to Horn and Dorner (1998). A total of 562 isolates were collected and grown in 1 mL of modified YES broth (20 g yeast extract; 150 g sucrose; 10 g soytone; 1 L distilled water) and screened for aflatoxins and CPA production by thin-layer chromatography (TLC) (Hicks et al. 1997). Finally, single-spore culture of 238 isolates were randomly selected for aflatoxin B_1, B_2, G_1, and G_2 quantifications, including 14 *Aspergillus* spp. strains for reference (NRRL 50429 [TX21-5], NRRL 3357, NRRL 21882, NRRL 50428 [NC7-8], NRRL 50427 [AL4-7], NRRL 29487 [F10], NRRL 29459 [F35], NRRL 29473 [F86], NRRL 29537 [F90], NRRL 18543 [AF36], 99-8Q, 51-1EP, 144EP). After a cleanup procedure (Sobolev and Dorner 2002), aflatoxins were quantified using the following equipment (all from Waters): Acquity ultra-high-performance (UPLC) instrument equipped with a matching UPLC H-class quaternary solvent manager, UPLC sample manager, UPLC fluorescence detector (FLR), and an Acquity UPLC BEH C18 2.1 × 50 mm, 1.7 µm column. The

mobile phase was composed of a H_2O/MeOH/CH_3CN (64:23:13, *v*/*v*/*v*) mixture, and the flow rate was 0.30 mL/min. The column was maintained at 35 C in the system column heater. Concentrations of aflatoxins were determined by reference to peak areas of corresponding commercial standards (Sigma, St. Louis, Missouri, USA). The results were expressed in µg/mL of culture broth. Detection limit for aflatoxins G_1 and B_1 was 0.1 µg/mL and for aflatoxins G_2 and B_2 was 0.01 µg/mL.

Insertion-deletion (InDel) primers were designed based on seven ABCs of *Aspergillus* spp. arbitrarily chosen from the National Center for Biotechnology Information (NCBI) GenBank database for the purpose of observing InDels; accession numbers were AY371490.1, AB196490.1, AY510454.1, AY510451.1, AY510453.1, AY510452.1, and AY510455.1. The ABC sequences were aligned, and a consensus was generated and realigned with the seven sequences to visualize InDels especially in conserved areas adjacent to the distal end and telomeric region. This alignment rendered approximately 28 random InDel areas for which we could design primers that, including these conserved areas, would generate 100–500-bp amplicons, suitable for our capillary electrophoresis platform. Alignments and primer design were performed using Clone Manager Professional 9 (Sic-Ed Software, Morrisville, North Carolina). A total of 25 InDel primer sets (TABLE 1) were designed (FIG. 1) and tested on the 248 *Aspergillus* spp. isolates, including the 14 reference strains. Forward primers were 5′ tailed with the sequence 5′-CAGTTTTCCCAGTCACGAC-3′ and labeled with 6-carboxyfluorescein (6-FAM) (Integrated DNA Technologies, Coralville, Iowa) for product labeling (Waldbieser et al. 2003). Reverse primers were tailed at the 5′ end with the sequence 5′-GTTT-3′ to

Table 1. InDel markers designed throughout the aflatoxin biosynthesis cluster to discriminate *Aspergillus* species.

Marker name	Forward 5′ → 3′	Reverse 5′ → 3′	Range of amplicon sizes (bp)	Number of alleles (mean ± SE)	Total number of amplicons/ marker
AFLC01	CCGACCTCACGACGCATTAT	CCGGCTAGCTTCAACAGACG	143–370	0.99 ± 0.01	8
AFLC02	GGTTGGCGGATTGAGAGGTA	GGAGATCAGCCGAGAAGACA	104–296	1.04 ± 0.02	11
AFLC03	TCCGCCGAGAGCCATAATAG	GGATGCTGACACCTCGATAG	121–161	1.06 ± 0.02	4
AFLC04	ACAGCTGGCATGCTCCGTAT	ATTGCTGCGCACGACGCTTA	197–370	0.98 ± 0.01	3
AFLC05	GTGGATGGACTGCCACTTAG	AGACCACAGTGAGTGCTTCT	154–175	0.98 ± 0.01	3
AFLC06	GCTGTCCTGGACGGATAGTA	CATCGGTCAACGACGAAGTA	230–232	0.93 ± 0.02	3
AFLC07	GTCAGCAAGAGGAGCCTTCA	GGTCACGGAGATCCTCCATA	161–407	0.60 ± 0.03	4
AFLC08	CGCCAGCACGGAGATCGAAT	CGTCTCCTCAGGCGGTCTAT	224–399	1.06 ± 0.03	6
AFLC09	AACACTCCGCTGCTCAACTA	AACGCTCAGGCAACGTCGAA	113–158	1.04 ± 0.01	3
AFLC10	GACGTTGCCTGAGCGTTAAT	TGACTGGTCGTCGCCAGAAT	202–208	1.00 ± 0.01	3
AFLC11	CTCGACGTAGCGTTGAACAG	AACGCATGGCCAGCTAATCT	215–468	0.98 ± 0.01	3
AFLC12	CGCAAGGAGCTCGACCAATA	TTCAGCTCAGCGACGAGAGT	240–360	1.06 ± 0.03	4
AFLC13	TCGGTTCAATGCTCGAACAC	TCCAACCTTCGGCCTAGTCT	139–417	1.00 ± 0.01	6
AFLC14	GACGCCTCGGCTTGTCAAGA	CTCCAACCTTCGGCCTAGTC	95–107	0.92 ± 0.02	4
AFLC15	GCTCTACAGGCTGATTCAAG	TCGACAGTCCGACAATATGC	206–370	1.02 ± 0.01	4
AFLC16	ATCGCAGCGGAAGCTTGGAA	AGTCTCGGACTCCGGTGACA	147–416	1.15 ± 0.04	12
AFLC17	GCACAACTCGTACAGCTATC	TCTAAGTGCGAGGCAACGAA	127–391	1.13 ± 0.03	10
AFLC18	GGCAGCCAGACCAAGGAATA	CCTTCTCGTAGCCGCTCATC	133–412	1.01 ± 0.01	5
AFLC19	ACAGGACCGCACGGATCAAT	AGGAGCGGATGTCGAAGTCT	262–490	0.98 ± 0.01	6
AFLC20	GCCTAGCGCTCCATTCTCAG	CCATCGTATCCGGCTCTATC	116–370	0.98 ± 0.02	6
AFLC21	TACCTTACTCCGCTAAGCAG	GCGGTCACCTACCAATGAAT	151–371	1.01 ± 0.01	5
AFLC22	TTCGCAGGAGTGTAGCCAAG	GTTGGAACACGCTCCATAGG	121–371	0.98 ± 0.01	5
AFLC23	GGCGTCAGTGGATTCCGGAT	CGTGGTCCGCAGCAATAGTG	125–149	1.05 ± 0.01	2
AFLC24	GAACGAGATAACGGCTGCAT	ATCAATCCACGGACCGTTGT	102–427	0.99 ± 0.02	8
AFLC25	CAGTGCGACCGGATGGTACA	CGGCTGAACGCGATGACTCT	182	1.00 ± 0.01	1

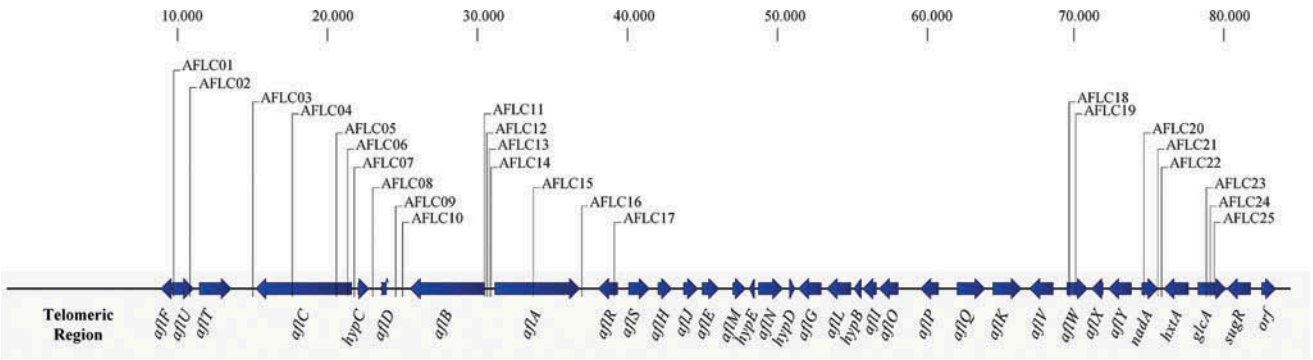

Figure 1. Location of InDel markers in the aflatoxin biosynthesis cluster of *Aspergillus flavus* NRRL 3357.

promote nontemplate adenylation (Brownstein et al. 1996). DNA was extracted from single-spore fungal isolates by using the DNeasy Plant Mini Kit (Qiagen, Valencia, California) following the manufacturer's instructions. Fungal material was vigorously disrupted twice for 40 s at room temperature in a bead mill Omni Bead Ruptor 24 homogenizer (Omni International, Kennesaw, Georgia). DNA was isolated using an automated QIAcube robotic station (Qiagen), and DNA concentrations were determined on a NanoDrop (Thermo Scientific, Waltham, Massachusetts). The PCR amplification used 10 ng DNA template and Titanium *Taq* DNA polymerase (Clontech, Mountain View, California) in 5-µL reactions on an M&J thermal cycler (Bio-Rad Laboratories, Hercules, California). Amplification conditions were 95 C for 1 min, 60 C for 1 min (two cycles), 95 C for 30 s, 60 C for 30 s, 68 C for 30 s (for 27 cycles), and final extension at 68 C for 4 min. Fluorescently labeled PCR fragments were analyzed on an ABI 3730XL DNA analyzer, and data were processed using GeneMapper version 4.0 (both from Applied Biosystems, Foster City, California). Presence of alleles was converted to a binary matrix, where different size amplicons (bp) at each locus were scored as "1" and absence of an amplicon was scored as "0"; samples that showed no amplification were considered to have "null allele or part of a larger deletion" and scored as "0." Isolate relationships according to InDel data and aflatoxin production were studied using a principal coordinate analysis (PCoA) on the NTSYSpc software (Rohlf 2000).

Cluster analysis was performed for the results of 25 InDels and 248 *Aspergillus* DNA samples; genetic distances and neighbor joining (Saitou and Nei 1987) were calculated using NTSYSpc software (Rohlf 2000). The confidence levels for the dendrograms were assessed by bootstrap resampling (100 replicates) (Felsenstein 1985; Efron et al. 1996) by using WinBoot (Yap and Nelson 1996).

An assignment test of the 248 isolates was made with Structure 2.1 (Pritchard et al. 2000) using the admixture model with 200,000 burn-ins and 200,000 iterations to allow the Markov chain to reach stationarity. A total of 10 independent simulations were run for each value of K tested, ranging from 1 to 10. The data generated were used to obtain the ideal K with the method of Evanno et al. (2005) using the Structure Harvester program (Earl and vonHoldt 2011).

Genomic DNA of representative isolates was processed for whole-genome sequencing using Next-Generation Sequencing (Illumina Hiseq2500; San Diego, California, USA) at the High-Throughput Genomics Centre, University of Washington. The sequenced reads obtained were processed as published in Faustinelli et al. (2016) using Geneious 8.1.7 (Kearse et al. 2012). The ABC sequences from the 10 genomes were aligned to the ABC (92,078 bp) of *A. flavus* NRRL 3357 (Nierman et al. 2015) using the Clone Manager multiway alignment tool (Clone Manager 9 Professional Edition; Denver, Colorado, USA) to generate, according to sequence similarity, the corresponding dendrogram.

RESULTS

More than 500 *Aspergillus* isolates were collected from peanut seeds in Georgia during 2014. The predominant species from section *Flavi* were *A. flavus* (95.5%), *A. parasiticus* (3.3%), and *A. caelatus* (1.2%). The 25 InDel markers designed were used to assess the genetic variation within the ABC of 234 randomly selected *Aspergillus* isolates as well as 14 reference *Aspergillus* strains. The InDel markers detected 129 alleles, ranging from 1 to 12 alleles per marker. Some primers identified deletion patterns associated with a deficiency of aflatoxin production. TABLE 1 includes sequences of InDel primers and amplicon sizes, alleles per sample, and number of amplicons obtained per marker.

The dendrogram obtained from InDel markers (FIG. 2) revealed, with a 25% cutoff distance criterion (0.4025), that the section *Flavi* isolates sampled in this study belonged to three main groups; five groups were represented by one isolate (25-5, 55-2, 69-1, NC7-8 [NRRL 50428], TX21-5 [NRRL 50429]). Group I comprised 10 non-aflatoxigenic producers, including one used as a biocontrol agent (NRRL 21882). Group II included *A. parasiticus* (aflatoxin B and G producers). Group III, the largest, included mostly aflatoxigenic *A. flavus* and the three non-aflatoxigenic *A. caelatus* isolates that formed a subgroup (GIII-a). Four additional subclusters were distinguished based on their capability to produce either large or small amounts of aflatoxin B_1. Half of subgroup GIII-b, which represented 18.5% of the Group III total, produced between 50 and 100 µg/mL of aflatoxin B_1, whereas the rest produced less. GIII-c represented almost 43% of the Group III total and comprised the higher aflatoxin B_1 producers, e.g., isolate 53-2 with 407.5 µg/mL, 23-1 with 313.4 µg/mL, and 24-1 with 254.8 µg/mL. The other two *A. flavus* subgroups (GIII-d and GIII-e) produced average aflatoxin B_1 concentrations of 57.24 ± 44.8 (±SD; n = 17) and 6.47 ± 17.9 (n = 17) µg/mL, respectively.

Estimates of ancestral components of the 248 *Aspergillus* isolates using Structure program are shown in FIG. 3. The value that captures the major structure in our data is K = 10 (ΔK = 14.2271). Color coding was

Figure 2. Cluster analysis of 248 *Aspergillus* isolates from peanut seeds from Georgia. Dendrogram obtained by DNA samples and 25 InDel markers coded as presence/absence of amplification using Neighbour Joining algorithm with NTSYSpc software. GI to GVII represent groups with different aflatoxin-producing capability. (An enlarged version of FIGURE 2 is included with the online supplemental files available at www.tandfonline.com/umyc.)

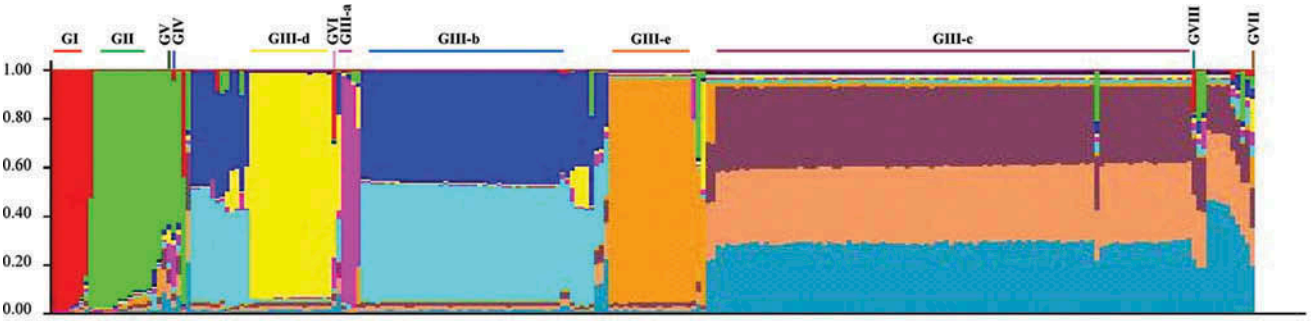

Figure 3. Estimates of ancestral components of 248 isolates of *Aspergillus* section *Flavi* from Georgia peanut seeds, using Structure program (k = 10) and 16 InDel markers. GI to GVII represent groups with different ancestral components.

used to show the probability of belonging to a group (Peña-Malavera et al. 2014). Each individual is represented by a vertical line partitioned into colored segments whose lengths are proportional to the contributions of the ancestral components. This structure clustering provided data comparable with the dendrogram obtained with InDels (FIG. 2). Group I (GI), comprising the 10 non-aflatoxigenic isolates, was clearly differentiated in red, and Group II (GII), *A. parasiticus*, was shown in green. Group III was heterogeneous as expected and showed subpopulations in yellow (GIII-d) and orange (GIII-e). We can also identify two admixed subpopulations, GIII-b (blue and light blue) and GIII-c (purple, light orange, and turquoise), and the subgroup comprising *A. caelatus* (GIII-a) was shown in fuchsia.

The data obtained by UPLC showed that all the isolates from Group I and subgroup GIII-a (*A. caelatus*) were non-aflatoxigenic. In Group II, *A. parasiticus* produced aflatoxin B_1 ranging from 1.41 to 122.15 µg/mL, aflatoxin G_1 from 0.5 to 137.92 µg/mL, and trace amounts of B_2 and G_2. The 223 *A. flavus* isolates in Group III produced aflatoxin B_1 ranging from not detected to 407.49 µg/mL (SUPPLEMENTARY TABLE 1). The principal coordinate analysis (PCoA) based on the aflatoxin production data showed that the first three coordinates explained more than 89% of the total observed variation. The first (PCo1) and second (PCo2) coordinates distinguished most of the isolates that do not produce aflatoxins, including Group I. The third coordinate (PCo3) contributed to the differentiation of Group II (*A. parasiticus*) isolates, which produced B and G aflatoxins (SUPPLEMENTARY FIG. 1).

The sensitivity of capillary electrophoresis for detecting insertions/deletions allowed grouping of individuals that do not produce aflatoxins. Isolates from Group I, for example, were characterized by a deletion of 40 bp between the *aflD* and *aflB* genes (marker AFLC09) and an insertion of a 173-bp segment into the *aflC* gene region (marker AFLC04). In Group II, the presence of a 135-bp insertion was detected, which separated all of the *A. parasiticus* isolates into one cluster (marker AFLC21) except for isolate 206-3, which was the highest aflatoxin B_2 (4.1 µg/mL) and G_2 (6.6 µg/mL) producer. Group III was more heterogeneous in exhibiting distinct subclusters. The non-aflatoxigenic *A. caelatus* was clearly discriminated, forming a homogeneous group with marker AFLC17 (SUPPLEMENTARY FIG. 2).

These analyses allowed for the selection of 10 representative section *Flavi* isolates from various groups for whole-genome sequencing using Illumina. The DNA dendrogram generated by multiple alignments of the entire ABC of the 10 isolates (FIG. 4) showed a clustering comparable to the one obtained with the InDels (FIG. 2). Alignment of the ABC of *A. flavus* 26-3 (aflatoxin and CPA nonproducer) with *A. flavus* NRRL 3357 showed 95% base matching in the 92,078-bp total length of aligned sequences, whereas *A. parasiticus* 68-5 showed an 81% base matching with *A. flavus* NRRL 3357. Group III included two isolates (78-6, 54-2) that only produced low aflatoxin B_1 (0.002 and 0.004 µg/mL, respectively) and that shared 98–99% of bases with *A. flavus* NRRL 3357. Despite the high percentage of ABC homology, genetic differences were enough to differentiate section *Flavi* isolates in their capability to produce toxins.

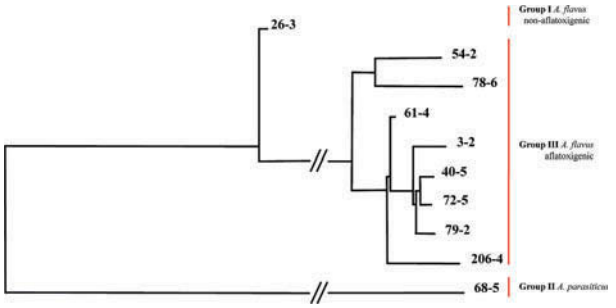

Figure 4. Cluster analysis of *Aspergillus* isolates from Georgia. Dendrogram generated according to aflatoxin biosynthesis cluster sequence similarity of 10 isolates using Geneious 8.1.7 and Clone Manager 9 software. Group I to Group III are groups with different capability to produce aflatoxins.

The genome sequences of eight *A. flavus* isolates and one *A. parasiticus* isolate have been submitted to NCBI and the nucleotide sequence accession numbers are as follows: LOAN00000000, LLET00000000, LOAO00000000, LOAM00000000, LIZI00000000, LIZJ00000000, LOAK00000000, LOAL00000000, and LOAP00000000 (Faustinelli et al. 2016).

DISCUSSION

Although the cost of sequencing entire genomes of microorganisms has significantly decreased in recent years (Baym et al. 2015; Previte et al. 2015; Rowan et al. 2015), it is still impractical to sequence hundreds or thousands of individuals to generate genetic data. Here, we propose a workflow that allows prescreening the predominant aflatoxigenic strains as a means to obtain DNA sequencing information without sequencing all the individuals. This approach involves fingerprinting with InDel markers within the ABC using capillary electrophoresis, UPLC detection of aflatoxin production, and whole-genome sequencing of few representatives within groups. The genetic data were

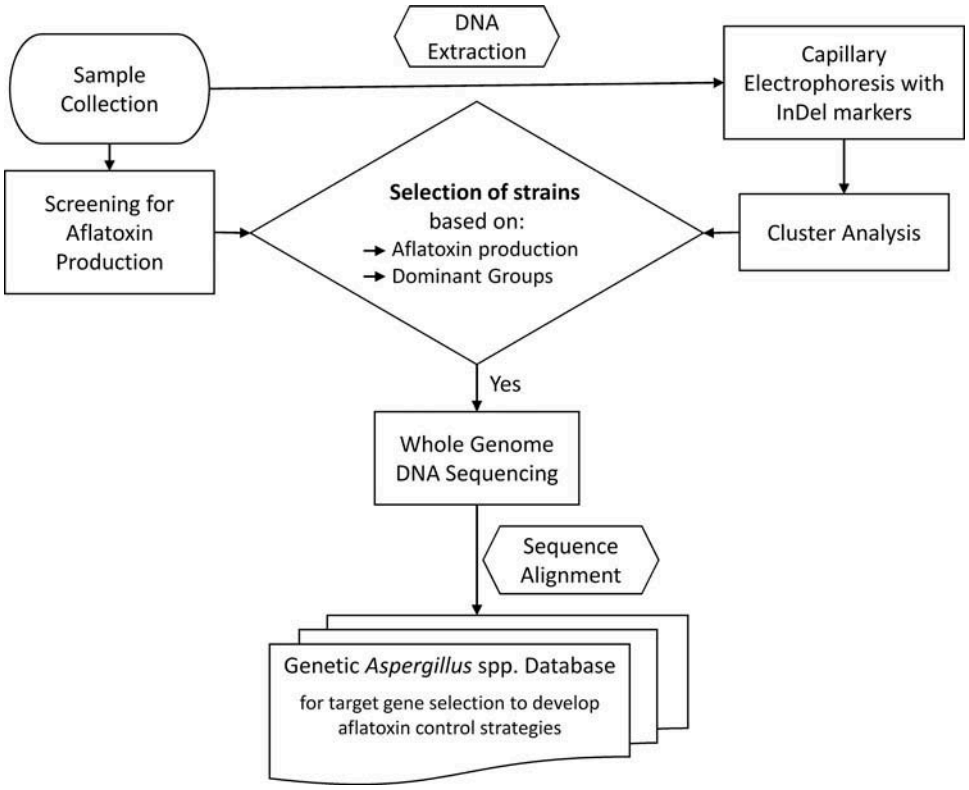

Figure 5. Workflow to carry out screening of *Aspergillus* spp. isolates for aflatoxin-cluster genotype variations (using InDels) and aflatoxin production.

organized into a genomic *Aspergillus* spp. database that will be useful for future target gene selection to develop aflatoxin control strategies (FIG. 5).

Callicott and Cotty (2015) used CAP markers to characterize the ABC as a practical tool for assessing intraspecific variability in *A. flavus*. However, they indicated that one disadvantage of CAP markers is the presence of multiple genotypes with the same CAP profile and/or lesion. In our studies, InDel fingerprinting efficiently identified genetic biodiversity within the section *Flavi*. Using fluorescent capillary electrophoresis, these markers were able to locate insertions and deletions within the ABC that may result in the inability of the fungus to produce aflatoxins. For example, the insertion detected in the *aflC* gene region may be associated with the inability of Group I isolates to produce aflatoxins. A large deletion in *A. oryzae* (Lee et al. 2006) and a premature stop codon in the *A. flavus* biocontrol AF36 *pksA* gene (Ehrlich and Cotty 2004) are responsible for their lack of aflatoxin-producing capability. Ehrlich et al. (2008) showed that biosynthesis of AFG_1 involves *nadA* reduction in *A. parasiticus*, and it was confirmed by Cai et al. (2008) that the *nadA* gene encodes the cytosol enzyme required for the last biosynthesis step of G aflatoxins. The insertion we found between *nadA* and *hxtA* in *A. parasiticus* from Group II may contribute to its capability to produce G aflatoxins.

All analyses (neighbor-joining, three-dimensional principal coordinate, Structure) revealed that the *Aspergillus* isolates sampled in this study were grouped by their capability to produce aflatoxins. InDel markers were capable of distinguishing non-aflatoxigenic from aflatoxigenic *Aspergillus* spp. as well as *Aspergillus* spp. by mycotoxin profile and amount of aflatoxins produced. Interestingly, these markers differentiated groups within the toxigenic *A. flavus* group (GIII) and showed some association with the amount of aflatoxin produced. Although higher concentration of aflatoxin are expected from S strains (Horn and Dorner 1999), in the year 2014 only L strains were observed in the dilutions plated; using the same sampling protocol, we usually recover L and S strains from other geographical areas (Mohammed et al. 2016). Since point mutations (Ehrlich and Cotty 2004) or small deletions (Calvo et al. 2004) in regulatory genes are sufficient to affect aflatoxin production, further studies should be conducted to determine if the insertions/deletions observed are related to the capability to produce large or small amounts of aflatoxins.

Finally, we selected two representative isolates from the main groups to perform whole-genome sequencing using Illumina. The dendrogram generated according to genomic DNA sequences matched the cluster analysis obtained with InDels. Despite the high percentage of ABC homology observed, the genetic differences obtained allowed discrimination between species (*A. flavus, A. parasiticus*) and an association of isolates with their aflatoxin-production capability. The genetic information generated here is required for the design of control strategies where the identification of conserved regions within genomes is essential. The effectiveness of RNAi-mediated resistance is difficult to predict. For example, in transgenic tobacco plants, silencing of the glucuronidase (*GUS*) gene translocated from the plant to the pathogen *Fusarium verticillioides* was 62–97% effective (Tinoco et al. 2010). In transgenic wheat, silencing of a gene within the phytopathogenic fungus *Puccinia striiformis* f. sp. *tritici* infecting wheat was 70–90% effective (Yin et al. 2011). Elbashir et al. (2001) performed extensive analysis of small interfering RNA (siRNA) duplexes, concluding that the target recognition is a highly sequence-specific process. The *Aspergillus* database generated using InDel markers and DNA sequencing will provide a full view of relevant targets from the genome of this pathogen.

In conclusion, InDel markers based on the aflatoxin biosynthesis cluster (ABC) have the ability to distinguish groups within section *Flavi*, showing their potential application for the selection of the most abundant genotypes in a sampled area and minimizing the number of entire genomes to DNA sequences. In this work, we developed a workflow that involves fingerprinting with 25 new InDel markers using capillary electrophoresis. Although InDels in *Aspergillus* have been used by other research groups, this is the first time that the cluster analysis resulting from fingerprinting was followed by whole-genome sequencing of representative isolates. In our study, the cluster analysis of the 10 ABC sequences validated the results obtained with the InDels. This shows that the 25 InDels used here can predict similarities at the genome sequence level. Our results also revealed a relationship between groups and their capability to produce aflatoxins. The database generated of *Aspergillus* spp. genomes will provide valuable information to design RNAi targets and to evaluate the effectiveness of the technology to reduce aflatoxin contamination in peanut.

ACKNOWLEDGMENTS

The authors thank Imana Power and Valerie Orner for critical suggestions.

FUNDING

This work was supported by USDA-ARS project number 6604-21000-003-00D, the Peanut Mycotoxin Innovation Laboratory at the University of Georgia, and USAID Feed the Future grants. Mention of trade names or commercial products in this article is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the US Department of Agriculture.

ORCID

Paola C. Faustinelli http://orcid.org/0000-0002-6925-4610
Bruce W. Horn http://orcid.org/0000-0002-0858-5424

LITERATURE CITED

Arias RS, Dang PM, Sobolev VS. 2015. RNAi-mediated control of aflatoxins in peanut: method to analyze mycotoxin production and transgene expression in the peanut/*Aspergillus* pathosystem. Journal of Visualized Experiments 106:e53398.

Azziz-Baumgartner E, Lindblade K, Gieseker K, Rogers HS, Kieszak S, Njapau H, Schleicher R, McCoy LF, Misore A, DeCock K, Rubin C, Slutsker L. 2005. Case–control study of an acute aflatoxicosis outbreak, Kenya, 2004. Environmental Health Perspectives 113:1779–1783.

Baym M, Kryazhimskiy S, Lieberman TD, Chung H, Desai MM, Kishony R. 2015. Inexpensive multiplexed library preparation for megabase-sized genomes. PLoS ONE 10:e0128036.

Bayman P, Cotty PJ. 1993. Genetic diversity in *Aspergillus flavus*: association with aflatoxin production and morphology. Canadian Journal of Botany 71:23–31.

Brownstein MJ, Carpten JD, Smith JR. 1996. Modulation of non-templated nucleotide addition by TaqDNA polymerase: primer modifications that facilitate genotyping. BioTechniques 20:1004–1010.

Cai J, Zeng H, Shima Y, Hatabayashi H, Nakagawa H, Ito Y, Adachi Y, Nakajima H, Yabe K. 2008. Involvement of the *nadA* gene in formation of G-group aflatoxins in *Aspergillus parasiticus*. Fungal Genetics and Biology 45:1081–1093.

Callicott KA, Cotty PJ. 2015. Method for monitoring deletions in the aflatoxin biosynthesis gene cluster of *Aspergillus flavus* with multiplex PCR. Letters in Applied Microbiology 60:60–65.

Calvo AM, Bok J, Brooks W, Keller NP. 2004. veA is required for toxin and sclerotial production in *Aspergillus parasiticus*. Applied and Environmental Microbiology 70:4733–4739.

Carbone I, Jakobek JL, Ramirez-Prado JH, Horn BW. 2007. Recombination, balancing selection and adaptive evolution in the aflatoxin gene cluster of *Aspergillus parasiticus*. Molecular Ecology 16:4401–4417.

Chang PK, Abbas HK, Weaver MA, Ehrlich KC, Scharfenstein LL, Cotty PJ. 2012. Identification of genetic defects in the atoxigenic biocontrol strain *Aspergillus flavus* K49 reveals the presence of a competitive recombinant group in field populations. International Journal of Food Microbiology 154:192–196.

Chang PK, Horn BW, Dorner JW. 2005. Sequence breakpoints in the aflatoxin biosynthesis gene cluster and flanking regions in nonaflatoxigenic *Aspergillus flavus* isolates. Fungal Genetics and Biology 42:914–923.

Earl DA, vonHoldt BM. 2011. Structure Harvester: a website and program for visualizing Structure output and implementing the Evanno method. Conservation Genetics Resources 4:359–361.

Efron B, Halloran E, Holmes S. 1996. Bootstrap confidence levels for phylogenetic trees. Proceedings of the National Academy of Sciences of the United States od America 93:7085–7090.

Ehrlich KC, Cotty PJ. 2004. An isolate of *Aspergillus flavus* used to reduce aflatoxin contamination in cottonseed has a defective polyketide synthase gene. Applied Microbiology and Biotechnology 65:473–478.

Ehrlich KC, Montalbano BG, Cotty PJ. 2003. Sequence comparison of *aflR* from different *Aspergillus* species provides evidence for variability in regulation of aflatoxin production. Fungal Genetics and Biology 38:63–74.

Ehrlich KC, Scharfenstein LL, Montalbano BG, Chang P-K. 2008. Are the genes *nadA* and *norB* Involved in formation of aflatoxin G_1? International Journal of Molecular Sciences 9:1717–1729.

Ehrlich KC, Yu J, Cotty PJ. 2005. Aflatoxin biosynthesis gene clusters and flanking regions. Journal of Applied Microbiology 99:518–527.

Elbashir SM, Martinez J, Patkaniowska A, Lendeckel W, Tuschl T. 2001. Functional anatomy of siRNAs for mediating efficient RNAi in *Drosophila melanogaster* embryo lysate. The EMBO Journal 20:6877–6888.

Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Molecular Ecology 14:2611–2620.

Faustinelli PC, Wang XM, Palencia ER, Arias RS. 2016. Genome sequences of eight *Aspergillus flavus* spp. and one *A. parasiticus* sp., isolated from peanut seeds in Georgia. Genome Announcements 4:e00278–16.

Felsenstein J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783–791.

Gong Y, Hounsa A, Egal S, Turner PC, Sutcliffe AE, Hall AJ, Cardwell K, Wild CP. 2004. Postweaning exposure to aflatoxin results in impaired child growth: a longitudinal study in Benin, West Africa. Environmental Health Perspectives 112:1334–1338.

Hicks JK, Yu J-H, Keller NP, Adams TH. 1997. *Aspergillus* sporulation and mycotoxin production both require inactivation of the FadA Gα protein-dependent signaling pathway. The EMBO Journal 16:4976–4923.

Horn BW. 2007. Biodiversity of *Aspergillus* section *Flavi* in the United States: a review. Food Additives & Contamination 24:1088–1101.

Horn BW, Dorner JW. 1998. Soil populations of *Aspergillus* species from section *Flavi* along a transect through peanut-growing regions of the United States. Mycologia 90:767–776.

Horn BW, Dorner JW. 1999. Regional differences in production of aflatoxin B_1 and cyclopiazonic acid by soil isolates of *Aspergillus flavus* along a transect within the United States. Applied Environmental Microbiology 65:1444–1449.

Horn BW, Moore GG, Carbone I. 2009a. Sexual reproduction in *Aspergillus flavus*. Mycologia 101:423–429.

Horn BW, Ramirez-Prado JH, Carbone I. 2009b. The sexual state of *Aspergillus parasiticus*. Mycologia 101:275–280.

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A. 2012. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. BioInformatics 28:1647–1649.

Khlangwiset P, Wu F. 2010. Costs and efficacy of public health interventions to reduce aflatoxin-induced human disease. Food Additives & Contaminants, Part A: Chemistry, Analysis, Control, Exposure & Risk Assessment 27:998–1014.

Lamb MC, Sternitzke DA. 2001. Cost of aflatoxin to the farmer, buying point, and sheller segments of the Southeast United States peanut industry. Peanut Science 28:59–63.

Lee YH, Tominaga M, Hayashi R, Sakamoto K, Yamada O, Akita O. 2006. *Aspergillus oryzae* strains with a large deletion of the aflatoxin biosynthetic homologous gene cluster differentiated by chromosomal breakage. Applied Microbiology and Biotechnology 72:339–345.

Leidner J. 2012. The peanut genomics initiative. Southeastern Peanut Farmer (May–June):15.

Liao CD, Wong JW, Zhang K, Hayward DG, Lee NS, Trucksess MW. 2013. Multi-mycotoxin analysis of finished grain and nut products using high-performance liquid chromatography-triple-quadrupole mass spectrometry. Journal of Agriculture and Food Chemistry 61:4771–4782.

Mohammed A, Chala A, Dejene M, Fininsa Ch, Hoisington DA, Sobolev VS, Arias RS. 2016. *Aspergillus* and aflatoxin in groundnut (*Arachis hypogaea* L.) and groundnut cake in Eastern Ethiopia. Food Additives & Contaminants, Part B: Surveillance 9:290–298.

Moore GG, Singh R, Horn BW, Carbone I. 2009. Recombination and lineage-specific gene loss in the aflatoxin gene cluster of *Aspergillus flavus*. Molecular Ecology 18:4870–4887.

Murugavel KG, Naranatt PP, Shankar EM, Mathews S, Raghuram K, Rajasambandam P, Jayanthi V, Surendran R, Murali A, Srinivas U, Palaniswamy KR, Srikumari D, Thyagarajan SP. 2007. Prevalence of aflatoxin B_1 in liver biopsies of proven hepatocellular carcinoma in India determined by an in-house immunoperoxidase test. Journal of Medical Microbiology 56:1455–1459.

Nierman WC, Yu J, Fedorova-Abrams ND, Losada L, Cleveland TE, Bhatnagar D, Bennett JW, Dean R, Payne GA. 2015. Genome sequence of *Aspergillus flavus* NRRL 3357, a strain that causes aflatoxin contamination of food and feed. Genome Announcements 3: e00168–15.

Nigam SN, Waliyar F, Aruna R, Reddy SV, Kumar PL, Craufurd PQ, Diallo AT, Ntare BR, Upadhyaya HD. 2009. Breeding peanut for resistance to aflatoxin contamination at ICRISAT. Peanut Science 36:42–49.

Peña-Malavera A, Bruno C, Fernandez E, Balzarini M. 2014. Comparison of algorithms to infer genetic population structure from unlinked molecular markers. Statistical Applications in Genetics and Molecular Biology 13:391–402.

Previte MJ, Zhou C, Kellinger M, Pantoja R, Chen CY, Shi J, Wang B, Kia A, Etchin S, Vieceli J, Nikoomanzar A, Bomati E, Gloeckner C, Ronaghi M, He MM. 2015. DNA

sequencing using polymerase substrate-binding kinetics. Nature Communications 6:5936.

Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus Genotype Data. Genetics 155:945–959.

Raper KB, Fennell DI. 1965. The genus *Aspergillus*. Baltimore, Maryland: Williams & Wilkins. 686 p.

Roca WM, Mroginski LA. 1991. Cultivo de tehidos en la agricultura. Fundamentos y aplicaciones. Cali, Colombia: CIAT. xii + 970 p.

Rohlf FJ. 2000. NTSYSpc Numerical taxonomy and multivariate analysis system; version 2.1. User guide. New York, New York: Exeter Software, Applied Biostatistics Inc. 38 p.

Rowan BA, Patel V, Weigel D, Schneeberger K. 2015. Rapid and inexpensive whole-genome genotyping-by-sequencing for crossover localization and fine-scale genetic mapping. G3 Journal 5:385–398.

Saitou N, Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Molecular Biology and Evolution 4:406–425.

Sobolev VS, Dorner JW. 2002. Cleanup procedure for determination of aflatoxins in major agricultural commodities by liquid chromatography. Journal of AOAC International 85:642–645.

Spanjer MC, Rensen PM, Scholten JM. 2008. LC-MS/MS multi-method for mycotoxins after single extraction, with validation data for peanut, pistachio, wheat, maize, cornflakes, raisins and figs. Food Additives & Contaminants, Part A: Chemistry, Analysis, Control, Exposure & Risk Assessment 25:472–489.

Strosnider H, Wilson D, Williams JT, Wild C, Stroka J, Shephard G, Schaafsma A, Sabino M, Rubin C, Rogers HS, Pronczuk J, Pineiro M, Phillips T, Patel M, Park D, Page SW, Onsongo MTK, Ong C-N, Njapau H, Misore A, Miraglia M, Mensah P, McCoy L, Luber G, Liu X, Lewis L, Kibata GN, Jolly P, Jolly C, Jeffers D, Henry SH, Hell K, Groopman J, Dilley A, DeCock K, Brune M-N, Breiman R, Bhat RV, Banziger M, Azziz-Baumgartner E. 2006. Workgroup report: public health strategies for reducing aflatoxin exposure in developing countries. Environmental Health Perspectives 114:1898–1903.

Tinoco MLP, Dias BB, Dall'Astta RC, Pamphile JA, Aragão FJ. 2010. *In vivo* trans-specific gene silencing in fungal cells by *in planta* expression of a double-stranded RNA. BMC Biology 8:27.

Waldbieser GC, Quiniou SMA, Karsi A. 2003. Rapid development of gene-tagged microsatellite markers from bacterial artificial chromosome clones using anchored TAA repeat primers. BioTechniques 35:976–979.

Wang J-S, Huang T, Su J, Liang F, Wei Z, Liang Y, Luo H, Kuang S-Y, Qian G-S, Sun G, He X, Kensler TW, Groopman JD. 2001. Hepatocellular carcinoma and aflatoxin exposure in Zhuqing Village, Fusui County, People's Republic of China. Cancer Epidemiology, Biomarkers & Prevention 10:143–146.

Wu F, Groopman JD, Pestka JJ. 2014. Public health impacts of foodborne mycotoxins. Annual Review of Food Science and Technology 5:351–372.

Yap IV, Nelson RJ. 1996. WinBoot: a program for performing bootstrap analysis of binary data to determine the confidence limits of UPGMA-based dendrograms. IRRI Discussion Paper Series 14. Manila, Philippines: International Rice Research Institute. 32 p.

Yin Ch, Jurgenson JE, Hulbert SH. 2011. Development of a Host-Induced RNAi System in the Wheat Stripe Rust Fungus *Puccinia striiformis* f. sp. *tritici*. Molecular Plant-Microbe Interactions 24:554–561.

Yu J. 2012. Current understanding on aflatoxin biosynthesis and future perspective in reducing aflatoxin contamination. Toxins 4:1024–1057.