PHYLOGENETIC ANALYSIS OF ASPERGILLUS FLAVUS POPULATIONS IN MEXICO, TEXAS, AND PAKISTAN

By

AUSTIN TAYLOR KLEIN

_____

A Thesis Submitted to The Honors College

In Partial Fulfillment of the Bachelors degree
With Honors in

Molecular
and
Cellular
Biology

THE UNIVERSITY OF ARIZONA

M A Y 2 0 1 8

Approved by:

_____

Dr. Peter Cotty
School of Plant Science

Abstract

Aflatoxins are carcinogenic secondary metabolites that are produced mainly by members of the *Aspergillus* section *Flavi* species. Maize has long been associated with *A. flavus*, for maize seems to be especially prone to *A. flavus* infection.. Through simple-sequence repeat (SSR) data, it can be shown that the population of L-morphotype *A. flavus* in Mexico is not extremely diverse, with most individuals sharing many alleles across 17 loci of the *A. flavus* genome. However, in Pakistan, a country that has only had maize for a couple of centuries, the L-morphotype *A. flavus* community is diverse and possesses many unique alleles. For this reason, we are able to sort fungal individuals into two populations: a Modern Population (Population M) and an Ancient Population (Population A). This study shows that the diversity among the *A. flavus* community in Pakistan is the result of a native *A. flavus* population taking control of a new maize crop rather than a quick evolution of the *A. flavus* already present on the maize. This study will also show that the Ancient Population is older than the Modern Population through diversity measures and SplitsTree analysis.

Introduction

Maize is theorized to have first appeared in central Mexico, but since then, the maize crop has spread throughout the world. However, maize was not the only organism sent around

the world: a fungus, *Aspergillus flavus*, was often sent with it. *Aspergillus flavus* (*A. flavus*) is a fungus that resides in the warmest regions of the Earth and is of particular concern to the international community due to the fact that it produces aflatoxin. Aflatoxin is a known carcinogen, and consumption of aflatoxin can lead to liver cancer and even death depending on the dose of aflatoxin (Wu, 2015). Therefore, the tracking of *A. flavus* populations, especially as it relates to maize, can help us better understand the fungal communities present throughout the world and those parts of the world that are more susceptible to aflatoxin related issues. A better understanding of these communities would then allow us to better treat these communities using biocontrol products.

These biocontrol products work by first establishing what fungal individuals are most prevalent in a certain geographical region. From this fungal community, it is likely that there will be some individuals native to the field that produce no aflatoxin (atoxigenic). From the atoxigenic individuals, the strongest individual is selected, in terms of survivability, to be used in the biocontrol. The chosen atoxigenic individual is added to colored, water-activated pellets that farmers can spread on their fields. This, in effect, inoculates a field with a strong atoxigenic population that will compete for resources with the already present toxigenic population. Over

time, the atoxigenic population will overtake the toxigenic population, leading to an overall reduction in aflatoxin levels (Brown et al., 1991).

The focus of this study is on the divergence of *A. flavus* populations on maize in Mexico, Texas, and Pakistan. Specifically, this study will focus solely on *A. flavus* L-morphotype isolates (Cotty 1989). *A. flavus* has existed for millions of years on maize in Mexico and Texas, however, Pakistan has only had maize for a couple of centuries. Yet, through the use of simple sequence repeat (SSR) data, we can show that the *A. flavus* populations in Pakistan are divergent from those in Mexico and Texas (Grubisha and Cotty 2009). This could be the result of the founder effect, where by pure probability, a divergent population of *A. flavus* already present in Mexico was transplanted to Pakistan. More likely, though, is that a population of *A. flavus* was present in Pakistan for millions of years so that it could evolve separately from the population in Mexico and Texas. Then, once the maize arrived in Pakistan, the native *A. flavus* population simply overtook the fungal population already present on the imported maize crops as a result of environmental advantage.

Also of concern in this study is the diversity within each population separately. For both Texas and Pakistan, we have access to SSR data from multiple years so we are able to evaluate how each of the populations may have diverged from itself. For Pakistan, it seems that the alleles of each locus are actually increasing in comparison to alleles from Texas and Mexico. This is

most likely due to the unique environmental pressures present in Pakistan that are very different from those in Mexico or Texas.

Materials and Methods

Sample Isolation

Samples of maize from Pakistan, Mexico, and Texas were all isolated in the same way. Each maize sample was ground in a laboratory mill and then 10g of each sample was mixed with 50mL of 0.1% Tween-80 solution and shaken for 20 minutes. Dilutions of the maize/Tween solution were then spread on rose-Bengal agar plates (Cotty, 1994). The fungus was allowed to incubate for three days before fungal individuals (up to 8 per plate) were moved to 5-2 agar (Cotty, 1989) . After five to seven days of incubation, eight plugs from each plate were taken and saved in a water vial for future use (Cotty, 1988).

DNA Extraction

5-2 with 5% salt was inoculated with a fungal isolate and allowed to grow for 5 to 7 days. Spores were then harvested using a 0.1% Tween solution and spun down to create a pellet of spores.

After removing the Tween solution, Lysis buffer with SDS was added to the spore pellet and shaken for one hour at 60^{o}C and 8000rpm. The tubes are then spun down for 30 minutes at 14000rpm before the supernatant is removed from the tubes and transferred to new sterile tubes. After adding ammonium acetate and ethanol to the supernatant, the tubes are inverted several times and left at -20^{o}C for 30 minutes. Tubes are then spun down at maximum speed for 5 minutes and the supernatant is poured off. Finally, the tubes are allowed to dry out for at least an hour before the DNA is suspended in 25uL of water. This stock DNA is preserved in a -20^{o}C freezer for future use. For lab use, the concentration of DNA is determined via Nanodrop for each sample and then diluted until each sample has a DNA concentration of 5 ng/uL.

Running SSR Panels

First, four primer mixes are created that contain the 17 primer pairs. (Grubisha and Cotty, 2009). These primer mixes are then used to run four panels: A4, B7, D4, E2. In order to run panels, for each sample, a mix of 5 uL of Accustart, 2.5 uL of water, and 0.5 uL of one of the primer mixes

is added to 2 uL of sample DNA. Then each sample is run through a three-step PCR before being sent off for further genetic analysis.

Inoculation Experiment

10 g of clean corn is brought to 30% moisture content before being transferred into a 250 mL Erlenmeyer flask and autoclaved with a bug-stopper and foil for 20 minutes at 121 ^{o}C. Fungal isolates were grown on 5-2 with NaCl for seven days before being harvested with a cotton swab and stored in a vial with 10 mL of sterilized water. After using a turbidimeter to measure the spore content of each vial, 1,000,000 spores are added to each flask of corn with each flask having a different fungal isolate inoculated into it (Probst et al., 2010). The inoculated flasks with bug-stoppers are allowed to grow for seven days in an incubator at 31 ^{o}C. 50 mL of 70% ethanol is then added to each flask in order to halt fungal growth and extract aflatoxin. Each flask is then ground in a laboratory blender on high speed for 20 seconds. After allowing the liquid and solid layers to settle, 4 uL of the ethanol layer of each flask is spotted on a TLC plate and run against a known aflatoxin standard (Kachapulula et al., 2017). Plates are then scanned in density

spectrometer in order to determine intensity of fluorescence of the aflatoxin bands on the plate. Aflatoxin concentrations are then calculated off of these values in comparison to the standard.

Selection of Fungal Individuals

For all of the following analysis techniques, fungal isolates were sampled randomly using random number assignment via Excel. This sampling technique was used for the samplings in Table 1 and for choosing the isolates to be used for the inoculation experiments.

SplitsTree Analysis

Genetic distance was calculated for the SSR data via the START2 program (Jolley et al., 2001). From there, the genetic distance file was uploaded into the SplitsTree4 program for the drawing of the final SplitsTree (Huson, 1998).

PCoA Analysis and Allele Analysis

From the SSR data, the genetic distance was calculated through the GenAlEx program (Peakall and Smouse, 2006). GenAlEx was also used to perform the Principal Components Analysis and

the results were graphed within Excel. GenAlEx was also used in order to find the unique alleles and allele frequencies in the SSR data.

STRUCTURE Analysis

SSR data was imported into STRUCTURE version 2.3.2 and analysis was run for k values from 1 to 10 in triplicate (Pritchard et al., 2000 ; Falush et al., 2003, 2007). STRUCTURE data results were then analyzed for the best k value by looking at delta k, which was calculated by the mean of the second derivative of the likelihood of k over the standard deviation of k. This calculation was performed with the STRUCTURE Harvester using the Evanno method (Earl DA, 2012).

Results

The first analyses run were comparing small samplings from each of the three regions in order to observe any patterns in genetic distribution. This was done by randomly taking 50 individuals each from Pakistan, Mexico, and Texas and grouping the 150 individuals together in a single “sampling.” After repeating this procedure 5 times, the genetic data for each sampling was uploaded into the STRUCTURE program. All five samplings showed that the individuals clearly sorted out into two separate populations (Fig 2a). The k plot of the STRUCTURE results further confirmed that k=2 was the most likely of the population sorting (Fig 2b). This separation

of two populations is also confirmed by the PCoA analysis performed on the five samplings and the SplitsTree analysis of the five samplings (Figs 1 and 3). One consistency from all the samplings and analyses was that all of the Pakistan individuals would sort into one of the populations with the Mexico and Texas individuals sorting into both populations. Therefore, the population in which all of the Pakistan isolates sorted will be further referred to as the Ancient Population (Pop A) and the other population will be referred to as the Modern Population (Pop M). The results of the separation of the five samplings in these two populations can be found in Figure 1. Notably, it seems that the Mexico isolates stay mostly within the Modern Population while the Texas isolates seem to have a significant percentage of isolates (30.8 %) in the Ancient Population.

Diversity within these regions and within the two populations were also of interest. The regional diversity of the data was evaluated and is reflected in Table 1. It is clear from Shannon's and Nei's diversity indices that the individuals sampled from Pakistan are significantly more diverse than those from Mexico and Texas. The population diversity is noted by the extreme difference between the genetic distances of the individuals in the Ancient and Modern Populations (Table 2). The fact that the Ancient Population genetic distances are much longer than the Modern Population Distances also supports the diversity calculations from the random samplings. Along with evaluating the genetic distances the alleles at each SSR locus for each

population were analyzed (Table 4). For the majority of the loci, there are more alleles, more unique alleles, and a larger range of alleles for the Ancient Population than the Modern Population (Table 4). A final piece of evidence for the diversity among the two populations is the PCoA analysis which shows that the Ancient Population is much more spread out in the graph than the Modern Population (Fig 1).

Aflatoxin analysis was also performed on 45 individuals from Mexico, Texas, and Pakistan via an inoculation experiment (Table 3). Regionally, Pakistan individuals produced significantly more toxin than Texas individuals and Mexico individuals produced similar amounts of aflatoxin to both Texas and Pakistan. More interestingly, however, is that the those individuals belonging to the Ancient Population seemed to produce significantly more toxin than those individuals from the Modern Population. These results are especially important when considering the creation of a potential biocontrol for Pakistan.

Discussion

It is apparent from the data above that there are two populations present within the *Aspergillus flavus* communities in Pakistan, Mexico, and Texas. The remaining question then remains as to why this separation would be occurring in such a dramatic way. This difference in genetics between Pakistan and Mexico must derive from evolution. One of the first possibilities

for this divergence in the *A. flavus* species would be the founder effect. This would imply that when maize was brought to Pakistan merely a few centuries ago, the *A.* flavus that was on that maize just happened to be part of the small percentage of Mexican individuals that would fall into our classification of the Ancient Population. In this case, the founder effect seems to be an unlikely explanation.

Another explanation for the separation would be that the isolates in Pakistan evolved in order to adapt to the Pakistani environment. Actually, the location of Pakistan itself also helps to explain why the isolates from Texas seemed to be more closely related to the Pakistan isolates. Texas and Pakistan reside along the same latitude on the Earth, so they share similar environmental zones. Therefore, the shared environmental factors further give evidence toward more shared genetics between the Texas and Pakistan isolates. However, this explanation is insufficient to explain why the two populations exist because it would be impossible for *A. flavus* from Mexico to evolve in only a few centuries in order to adapt to the Pakistani environment.

A stronger explanation of the data would be to associate the divergence of the different populations with time. Assuming that different individuals of *A. flavus* mutate at the same rate, the only way that one population of *A. flavus* could be more diverse than another population would be if one of the populations is older than the other population. For this study, it has been shown that the Ancient Population is more diverse than the Modern Population (Table 1,2,4).

Therefore, the Ancient Population is most likely older than the Modern Population. This finding could shed light on the geographic origins of *A. flavus* itself, but further research would be necessary to come to such a conclusion.

One of the most immediate uses of the findings in this study would be towards the creation of a biocontrol for Pakistan. The fact that the individuals from the Ancient Population produced significantly more aflatoxin than those individuals from the Modern Population could mean that producing a biocontrol for Pakistan could be especially challenging (Table 3). This means that whichever individual(s) are chosen for the Pakistan biocontrol will have to compete with high-aflatoxin producing individuals in the field. However, atoxigenic individuals of Pakistan were found in this study, so there are possible candidates for a future Pakistan biocontrol.

References

Boyd ML, Cotty PJ. 2001. Aspergillus flavus and aflatoxin contamination of leguminous trees of the Sonoran Desert in Arizona. Phytopathology 91:913-919.

Brown R, Cotty P, Cleveland T. 1991. Reduction in Aflatoxin Content of Maize by Atoxigenic Strains of Aspergillus-Flavus. J Food Prot 54:623-626.

Cotty PJ. 1994. Comparison of 4 Media for the Isolation of Aspergillus-Flavus Group Fungi. Mycopathologia 125:157-162.

Cotty PJ. 1989. Virulence and Cultural-Characteristics of 2 Aspergillus-Flavus Strains Pathogenic on Cotton. Phytopathology 79:808-814.

Cotty PJ. 1988. Aflatoxin and Sclerotial Production by Aspergillus-Flavus - Influence of Ph. Phytopathology 78:1250-1253.

Cotty PJ, Cardwell KF. 1999. Divergence of West African and North American communities of Aspergillus section Flavi. Appl Environ Microbiol 65:2264-2266.

Earl DA, vonHoldt BM. 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation Genetics Resources 4:359-361.

Falush D, Stephens M, Pritchard J. 2003. Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. Genetics 164:1567-1587.

Falush D, Stephens M, Pritchard JK. 2007. Inference of population structure using multilocus genotype data: dominant markers and null alleles. Molecular Ecology Notes 7:574-578.

Grubisha LC, Cotty PJ. 2009. Twenty-four microsatellite markers for the aflatoxin-producing fungus Aspergillus flavus. Molecular Ecology Resources 9:264-267.

Hubisz MJ, Falush D, Stephens M, Pritchard JK. 2009. Inferring weak population structure with the assistance of sample group information. Molecular Ecology Resources 9:1322-1332.

Huson D. 1998. SplitsTree: analyzing and visualizing evolutionary data. Bioinformatics 14:68-73.

Jolley K, Feil E, Chan M, Maiden M. 2001. Sequence type analysis and recombinational tests (START). Bioinformatics 17:1230-1231.

Kachapulula PW, Akello J, Bandyopadhyay R, Cotty PJ. 2017. Aspergillus section Flavi community structure in Zambia influences aflatoxin contamination of maize and groundnut. Int J Food Microbiol 261:49-56.

Mehl HL, Jaime R, Callicott KA, Probst C, Garber NP, Ortega-Beltran A, Grubisha LC, Cotty PJ. 2012. Aspergillus flavus diversity on crops and in the environment can be exploited to reduce aflatoxin exposure and improve health. Advances Against Aspergillosis Ii 1273:7-17.

Peakall R, Smouse P. 2006. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Molecular Ecology Notes 6:288-295.

Peakall R, Smouse PE. 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. Bioinformatics 28:2537-2539.

Pritchard J, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. Genetics 155:945-959.

Probst C, Schulthess F, Cotty PJ. 2010. Impact of Aspergillus section Flavi community structure on the development of lethal levels of aflatoxins in Kenyan maize (Zea mays). J Appl Microbiol 108:600-610.

Wu F. 2015. Global impacts of aflatoxin in maize: trade and human health. World Mycotoxin Journal 8:137-142.

Table 1 Diversity Analysis of the Five Samplings

Community	Population	% Population in Community[a]	Shannon's[b]	Nei's[c]
Mexico	A	8.4	1.526	0.723
	M	91.6	0.833	0.381
Texas	A	32.4	2.033	0.793
	M	67.6	0.884	0.392
Pakistan	A	100	2.168	0.827
	M	0	N/A	N/A

[a] Percentage of each population within a given community. Values calculated from five random samplings (N=50 for each) from each community

[b] Shannon's diversity index calculated using SSR data via GenAlEx (Peakall and Smouse, 2006)

[c] Nei's diversity index calculated using SSR data via GenAlEx (Peakall and Smouse, 2006)

Table 2 Genetic Distance Comparison

	Average Genetic Distance[a]	
Community	Population A	Population M
Pakistan	12957.65^{A}	N/A
Mexico	11962.80^{A}	1427.97^{B}
Texas	10844.55^{A}	1530.66^{B}
Average	11926.76^{A}	1479.27^{B}

[a]Genetic distances calculated for L-morphotype *A.flavus* using GenAlEx for individuals from Pakistan in Population A (N = 601), Mexico in Population A (N= 75) and M (N= 1193), and Texas in Population A (N= 575) and M (N= 1191). Individuals were distributed into populations based off of STRUCTURE analysis. Averages were compared using paired t-test ($\alpha = 0.05$). Capital letters A/B represent significantly different values.

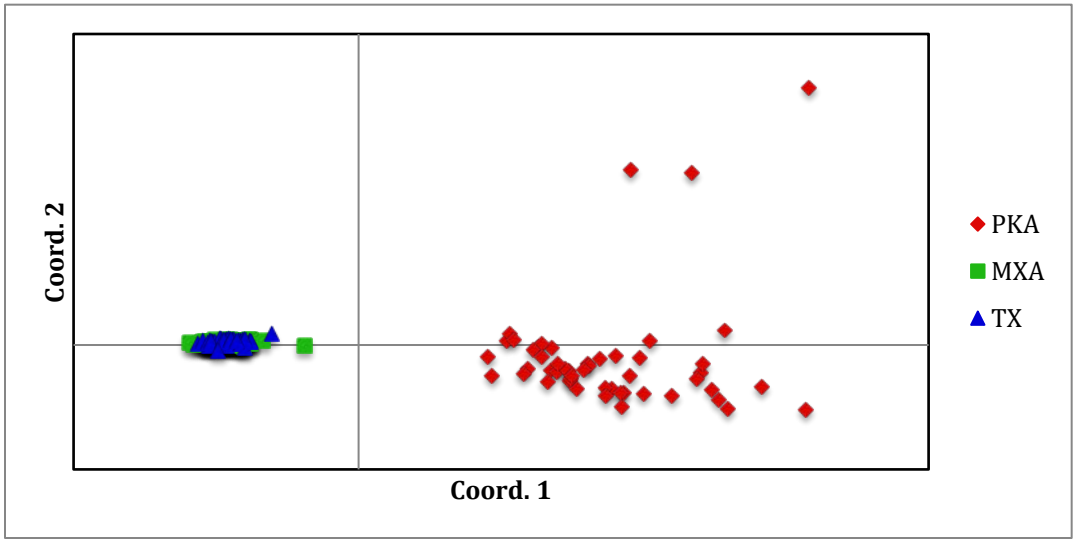
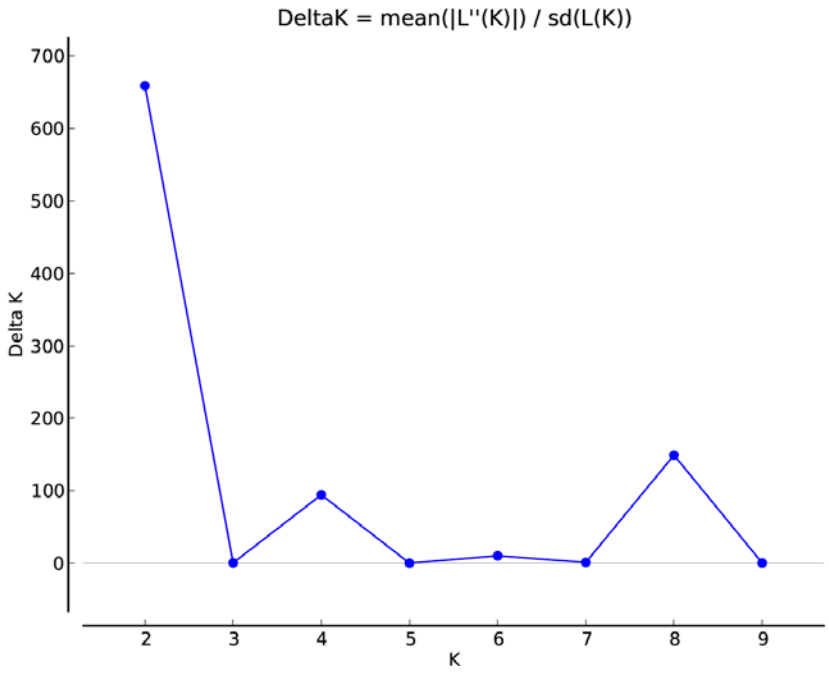


Fig. 1 Principal Components Analysis (PCoA) made with GenAlEx (Peakall and Smouse, 2006, 2012) for one of five random samplings of L-morphotype *A. flavus* (N = 50 for each community). The PCoA axis 1 and axis 2 explained 48.01% and 13.25% , respectively, of the total variation of the 17 loci of the SSR data. Left two quadrants represent Population M and right two quadrants represent Population A. Other PCoA analyses for the four other random samplings showed similar results.

a)



b)

Fig. 2 Results from STRUCTURE analysis based on one of five random samplings of L-morphotype *A. flavus* run for each K (K= 1-10) (Pritchard et al., 2000 ; Falush et al., 2003, 2007) . Other four samplings showed similar results. a) Calculation of delta K using STRUCTURE Harvester for STRUCTURE results, showing highest likelihood separation of K = 2 (Earl DA, 2012). b) Graphical representation of K = 2 plot for one of the five samplings with shading representing the estimated proportion of membership for each individual in the sampling. Red represents Population A and green represents Population M.

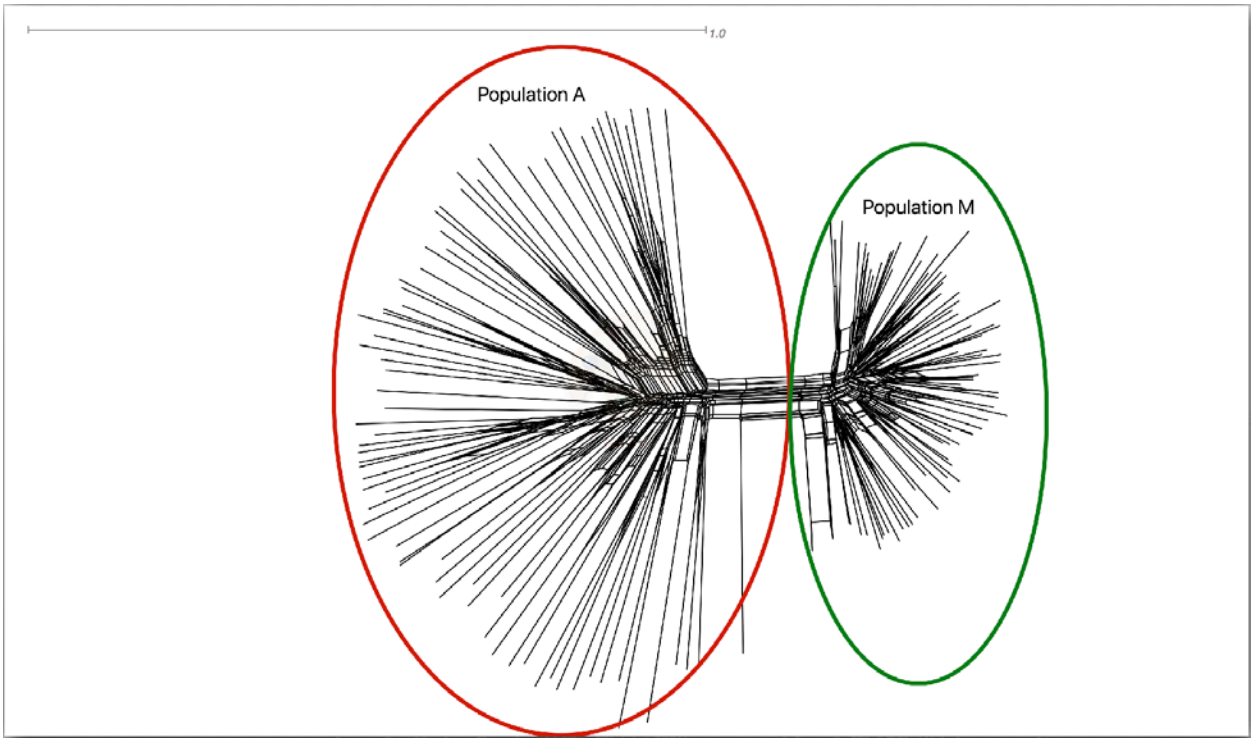


Fig. 3 SplitsTree for one of five random samplings of L-morphotype *A. flavus* (Huson, 1998). Other four samplings showed similar results with all SplitsTree's separating into two populations. Population A contains mostly Pakistani individuals with some Mexican and Texan individuals as well. Population M contains no Pakistani individuals and is entirely made up of individuals from Texas and Mexico.

Table 3 Total Aflatoxin Production (mg/kg)

	Ancient Population						Modern Population					
Trial	N	% 0[a]	% 0.1-1[b]	% 1-10[c]	% 10 - 100[d]	% >100[e]	N	% 0	% 0.1-1	% 1-10	% 10-100	% >100
1	22	8.69	0	4.55	40.9	54.54	20	45	10	30	15	0
2	18	N/A	0	0	38.88	38.88	28	N/A	0	25	25	3.57

[abcde] Percentage of L-morphotype *A. flavus* individuals from both populations who produced total aflatoxin (B1 and B2) in the ranges presented above. Individuals were inoculated onto a 250 mL Erlenmeyer flask with 10g of maize with 30% moisture for 7 days at 31ºC. Contents of each flask were ground in a laboratory blender with 50mL of 70% ethanol for aflatoxin extraction. Liquid layer of each flask was then spotted onto TLC plates and plates were run in a 96:3:1 ratio of ether : methanol : water for 45 minutes. Plates were then scanned by a CAMAG TLC plate scanner and total aflatoxin was calculated from areas under the curve of the peaks present from the scanner.

Fig. 4 Birds-eye view of L-morphotype *A. flavus* inoculated onto maize after 7 days at 31^{o}C

Table 4 Unique Allele Comparison

	Ancient Population			Modern Population			
Locus	# of Alleles[a]	% Unique Alleles[b]	Range of Alleles[c]	# of Alleles	% Unique Alleles	Range of Alleles	Total Unique Alleles[d]
1	18	75	55	5	25	29	8
2	30	55	273	18	45	79	20
3	22	17	266	44	83	199	24
4	14	67	190	4	33	44	6
5	36	100	255	3	0	6	7
6	25	69	172	7	31	98	16
7	22	39	258	16	61	100	18
8	55	74	254	14	26	243	23
9	37	67	107	19	33	77	15
10	50	86	267	7	14	20	35
11	12	50	196	7	50	23	10
12	22	38	168	23	62	139	13
13	44	57	147	16	43	48	14
14	16	31	44	13	69	139	13
15	55	79	139	8	21	80	19
16	10	50	18	6	50	8	8
17	24	58	55	12	42	17	12

[a] Total number of alleles present at a given locus within the given population

[b] Percentage of unique alleles present within a given population for a given locus

[c] Smallest allele at a given locus subtracted from the largest allele at a given locus for a given population

[d] Total number of unique alleles at a given locus for both populations