MOLECULAR ECOLOGY

Molecular Ecology (2009) **18**, 4870–4887 doi: 10.1111/j.1365-294X.2009.04414.x

Recombination and lineage-specific gene loss in the aflatoxin gene cluster of *Aspergillus flavus*

GEROMY G. MOORE,* RAKHI SINGH,* BRUCE W. HORN† and IGNAZIO CARBONE*

Center for Integrated Fungal Research, Department of Plant Pathology, North Carolina State University, Raleigh, NC 27695, USA, †National Peanut Research Laboratory, Agricultural Research Service, US Department of Agriculture, Dawson, GA 39842, USA

Abstract

Aflatoxins produced by *Aspergillus flavus* are potent carcinogens that contaminate agricultural crops. Recent efforts to reduce aflatoxin concentrations in crops have focused on biological control using nonaflatoxigenic *A. flavus* strains AF36 (=NRRL 18543) and NRRL 21882 (the active component of afla-guard®). However, the evolutionary potential of these strains to remain nonaflatoxigenic in nature is unknown. To elucidate the underlying population processes that influence aflatoxigenicity, we examined patterns of linkage disequilibrium (LD) spanning 21 regions in the aflatoxin gene cluster of *A. flavus*. We show that recombination events are unevenly distributed across the cluster in *A. flavus*. Six distinct LD blocks separate late pathway genes *aflE*, *aflM*, *aflN*, *aflG*, *aflL*, *aflI* and *aflO*, and there is no discernable evidence of recombination among early pathway genes *aflA*, *aflB*, *aflC*, *aflD*, *aflR* and *aflS*. The discordance in phylogenies inferred for the *aflW/aflX* intergenic region and two noncluster regions, tryptophan synthase and acetamidase, is indicative of trans-species evolution in the cluster. Additionally, polymorphisms in *aflW/aflX* divide *A. flavus* strains into two distinct clades, each harbouring only one of the two approved biocontrol strains. The clade with AF36 includes both aflatoxigenic and nonaflatoxigenic strains, whereas the clade with NRRL 21882 comprises only nonaflatoxigenic strains and includes all strains of *A. flavus* missing the entire gene cluster or with partial gene clusters. Our detection of LD blocks in partial clusters indicates that recombination may have played an important role in cluster disassembly, and multilocus coalescent analyses of cluster and noncluster regions indicate lineage-specific gene loss in *A. flavus*. These results have important implications in assessing the stability of biocontrol strains in nature.

Keywords: ancestral recombination graph, balancing selection, coalescent, linkage disequilibrium

Received 20 May 2009; revision received 8 October 2009; accepted 9 October 2009

Introduction

Species in *Aspergillus* section *Flavi* commonly infect agricultural staples such as corn, cottonseed, peanuts and tree nuts, and produce an array of mycotoxins (Horn 2007). The most common mycotoxins are the aflatoxins, carcinogenic polyketides that pose a serious health risk to animals and humans (Eaton & Groopman 1994). *Aspergillus flavus* and *Aspergillus parasiticus* are the most abundant aflatoxin-producing species in agricultural soils and have the potential to cause considerable contamination of crops (Horn 2005). *Aspergillus flavus* is the dominant aflatoxin-producing species in the majority of crops, whereas *A. parasiticus* is more frequently associated with peanuts (Horn *et al.* 1994; Horn 2005). The two species also differ in the types of aflatoxins produced. Aflatoxin biosynthesis involves more than 25 enzymes encoded by genes that are clustered together in a 75-kb telomeric region on chromosome 3 (Carbone *et al.* 2007b). Four structurally related aflatoxins occur in nature, depending upon the presence of

Correspondence: Ignazio Carbone, Fax: 919-513-0024;
E-mail: ignazio_carbone@ncsu.edu

polyketide dihydro- (B_1 and G_1) or tetrahydro- (B_2 and G_2) bisfuran rings. Of these aflatoxins, B_1 is the most toxic and carcinogenic (Eaton & Groopman 1994). Although both *A. flavus* and *A. parasiticus* produce the B aflatoxins (B_1 and B_2), *A. parasiticus* also produces the G aflatoxins (G_1 and G_2) or rarely *O*-methylsterigmatocystin (OMST), which is the precursor of aflatoxin B_1 (Yu *et al.* 2004). Because of the high toxicity of aflatoxins to humans and animals, federally and internationally mandated laws require the destruction, decontamination or reprocessing of crops when the aflatoxin content exceeds specified levels, resulting in huge yield losses worldwide (van Edmond & Jonker 2005). The economic loss is worse when crops are exposed to drought and elevated temperatures, conditions that favour crop invasion and fungal growth by aflatoxigenic fungi (Cole *et al.* 1985).

Individuals within *A. flavus* populations vary widely in their ability to produce aflatoxins, ranging from those that are nonaflatoxigenic to those that are potent producers of aflatoxins (Horn & Dorner 1999). *Aspergillus flavus* is characterized by two morphotypes: the typical L strain with large sclerotia >400 μm in diameter and the S strain with abundant small sclerotia <400 μm (Cotty 1989). Nonaflatoxigenic strains are fairly common in *A. flavus* L strain (Horn & Dorner 1999; Chang *et al.* 2005), but are rare in *A. flavus* S strain and *A. parasiticus* (Horn & Dorner 1999). In addition to aflatoxins, *A. flavus* produces another unrelated mycotoxin, cyclopiazonic acid (CPA), an indol-tetramic acid that targets the liver, kidneys and gastrointestinal tract in animals (Burdock & Flamm 2000). Aflatoxins and CPA often co-contaminate agricultural products and several of the disease symptoms in animals can be attributed to CPA (Cole 1986). CPA biosynthesis involves ~3 to 5 enzymes encoded in a 50-kb mini-cluster adjacent to the aflatoxin gene cluster (Chang *et al.* 2009). The inability to produce aflatoxins and/or CPA in some strains of *A. flavus* is the result of various deletions in these gene clusters (Chang *et al.* 2005, 2009). Populations of *A. flavus* show a high level of variation in mycotoxin production, with individuals producing both aflatoxins and CPA, aflatoxins alone, CPA alone or neither mycotoxin (Horn *et al.* 1996; Horn & Dorner 1999).

Two nonaflatoxigenic strains of *A. flavus* (AF36 and NRRL 21882) are currently registered through the US Environmental Protection Agency (EPA) and applied to fields as biocontrol agents to outcompete indigenous aflatoxigenic strains (Dorner 2005). AF36 was originally isolated from a cotton field in Arizona (EPA 2003) and approved for use on cotton in Arizona and Texas. Ehrlich & Cotty (2004) reported that loss of aflatoxigenicity in AF36 is the result of a nonsense mutation in *pksA* (*aflC*), a critical early pathway gene in aflatoxin biosynthesis. AF36 otherwise has a full aflatoxin gene cluster and a functional CPA cluster. NRRL 21882 (the active component of afla-guard®) was isolated from a peanut seed in Georgia (EPA 2004) and is currently approved for commercial use on peanuts and corn in the USA. Chang *et al.* (2005) reported that NRRL 21882 is missing the entire aflatoxin and CPA gene clusters. The application of highly competitive nonaflatoxigenic strains of *A. flavus* to fields has been shown to be effective in reducing aflatoxin contamination in peanuts (Dorner 2009b), corn (Dorner 2009a) and cottonseed (Cotty 1994). Although the two biocontrol strains have been approved by the EPA and are used widely in the USA, there is uncertainty as to whether they are capable of reacquiring toxigenicity through genetic exchange and recombination with indigenous aflatoxigenic strains.

The possibility that recombination in *A. flavus* could potentially influence the stability of biocontrol strains was first suggested by Geiser *et al.* (1998), who reported a population structure in *A. flavus* indicative of recombination based on a lack of congruence of five gene genealogies. However, the method of recombination, how often recombination occurs, or when recombination occurred in the history of the species could not be determined. Similar evidence for recombination based on incongruence of gene genealogies for four genes was shown for *Aspergillus nomius* (Peterson *et al.* 2001), another aflatoxin-producing species. More recently, Carbone *et al.* (2007a) sequenced 21 regions across the aflatoxin gene cluster for a population of *A. parasiticus* sampled from a single peanut field in Georgia. Detection of significant linkage disequilibrium (LD) over the evolutionary history of *A. parasiticus* resolved five recombination blocks across the gene cluster, whereas coalescent analysis suggested that some recombination had occurred within the last 1 million years. Furthermore, recombination alone separated some aflatoxin cluster haplotypes (Carbone *et al.* 2007a).

Although genealogical approaches have provided indirect evidence of recombination, our discovery of a sexual state for both *A. flavus* (Horn *et al.* 2009a) and *A. parasiticus* (Horn *et al.* 2009b,c) has provided the first direct evidence that sexual recombination may be an important mechanism for generating diversity in aflatoxin production in these agriculturally important species. In *A. parasiticus*, sexual reproduction in nature, as suggested by a 1:1 distribution of *MAT* genes in populations (Ramirez-Prado *et al.* 2008), increases diversity in mycotoxin profiles and creates new vegetative compatibility groups (VCGs) (B.W. Horn, unpublished data), whereas balancing selection acts to maintain the ratio of aflatoxins G_1/B_1 (Carbone *et al.* 2007a). In the present study, we show that recombination and

balancing selection in *A. flavus* also work in parity to maintain nonaflatoxigenicity and to influence cluster disassembly in a lineage-specific fashion. These underlying evolutionary processes are important considerations when developing biocontrol strategies.

Materials and methods

Population sample

Our initial sample of 79 single-spore *A. flavus* isolates (L strains) was from a single peanut field (Herod) sampled in 1992 (Horn & Greene 1995). This sample was consolidated to a subset of 44 isolates (Table 1) based on vegetative compatibility in which isolates capable of forming stable hyphal fusions are assigned to the same VCG (Leslie 1993). Vegetative compatibility groups 6–30 contain two or more isolates and VCGs 31–63 each contain a single isolate. We expanded the sample to include two additional strains, if available, within each VCG. Two VCGs (24 and 32) are nonaflatoxigenic; VCG 24 (NRRL 21882; IC252–254) is missing all the genes in the aflatoxin and CPA clusters and the telomeric repeat sequence (TTAGGG) is adjacent to *hxtA* (Chang *et al.* 2005), whereas VCG 32 (IC277) has complete clusters. The production of aflatoxins B_1 and B_2 and CPA was determined in a previous study (Horn *et al.* 1996). Also examined were *A. flavus* L strains representative of deletion patterns A (IC309 = NPL MS5-6 = NRRL 35735), B (IC310 = NPL NC3-6 = NRRL 35736), C (IC311 = NPL AL3-9 = NRRL 35737), D (IC312 = NPL GA4-4 = NRRL 35738), E (IC313 = NPL TX13-5 = NRRL 35739), F (IC314 = NPL NC7-8 = NRRL 35740) and G (IC315 = NPL AL1-4 = NRRL 35741) reported by Chang *et al.* (2005) in which NRRL 21882 is pattern H; mycotoxin production was determined by Horn & Dorner (1999). To examine trans-species evolution, we included the 24 isolates of *A. parasiticus* sampled from the same Georgia field (Horn & Greene 1995) and analysed in a previous study (Carbone *et al.* 2007a). Also included were the biocontrol strain AF36 (=NRRL 18543) and the genome strains *A. flavus* NRRL 3357 (http://www.aspergillusflavus.org/) and *Aspergillus oryzae* NRRL 5590 (Machida *et al.* 2005); *A. nomius* NRRL 13137 (ex type) was the outgroup species for *Aspergillus* section *Flavi.*

DNA isolation, PCR amplification and sequencing

Sequences of oligonucleotide primers (Table 2) and procedures for isolation of DNA and for amplifications were described previously (Carbone *et al.* 2007a). Approximately 1–2 kb was sequenced per isolate for each of 21 regions within the *A. flavus* aflatoxin gene cluster positioned on chromosome 3 and for two noncluster regions, tryptophan synthase (*trpC*) and acetamidase (*amdS*), located on chromosomes 4 and 6 respectively. We focused on the same regions that were previously examined in *A. parasiticus* (Carbone *et al.* 2007a). This was important for comparing patterns of LD and rates of recombination in the gene cluster between the two species and for reconstructing trans-species evolution. We used the primers *aflC-F* (TTAGATCGGTCCCTTTACTTTC) and *aflC-R* (GGTGTCAGTCCTTGTCTCTGTA) to amplify and sequence a coding portion of the *aflC* gene to determine whether nonaflatoxigenic strains contained a nonsense mutation, as previously reported (Ehrlich & Cotty 2004).

Molecular sequence variation

DNA sequences for each locus were aligned and manually adjusted using Sequencher Version 4.5 (Gene Codes Corporation). The multiple sequence alignment for each region was trimmed and exported as NEXUS files for further analysis in SNAP Workbench (Price & Carbone 2005). We calculated π, the mean number of pairwise nucleotide differences per site (Nei 1987), for each of the 21 multiple sequence alignments and for *trpC* and *amdS* using the program SITES version 1.1 (Hey & Wakeley 1997). Alignment files for the 21 regions were then combined using SNAP Combine (Aylor *et al.* 2006) to create concatenated sequences for each isolate that span the cluster. The multiple sequence alignment was collapsed into haplotypes with the options of recoding insertions/deletions (indels) and excluding all variation that violates an infinite-sites mutation model. The latter was important for coalescent analyses (described below) that assume an infinite-sites model. The real positions of single nucleotide polymorphisms (SNPs) in the concatenated multiple sequence alignment were based on the physical mappings of the 21 regions within the 75-kb aflatoxin gene cluster of *A. flavus*, using AF36 (AY510455) as a reference sequence.

Linkage disequilibrium and compatibility analyses

We first examined LD across the 21 regions for the subset of 44 isolates representing distinct VCGs. After inferring LD blocks in this subsample, we examined the expanded VCG sample to determine if block boundaries were conserved within VCGs. We performed LD and compatibility analyses using SNAP Clade and Matrix (Bowden *et al.* 2008). Recombination blocks were based on informative SNPs and recoded indels that are infinite sites compatible; uninformative variation was included to identify nonrecombining regions possibly under the influence of balancing selection (Carbone *et al.* 2007a). For example, previous research showed

Table 1 Strain designations, haplotypes, vegetative compatibility groups and mean concentrations of mycotoxins in *Aspergillus flavus**

IC	NRRL	*MAT*	VCG	Haplotype	B_1	B_2	Total aflatoxins	CPA
IC229	29459	2	6	H2	39.2 (22.4)	0.8 (0.5)	40 (22.9)	22.5 (3.8)
IC230	29460	2	6	H2	15.8 (13.3)	0.3 (0.2)	16.1 (13.5)	26.9 (3.2)
IC231	29461	2	6	H2	33.6 (13.3)	0.7 (0.2)	34.3 (13.5)	23.9 (2.5)
IC234	29464	2	14	H14	85.6 (5.8)	1.4 (0.3)	87 (5.7)	60.7 (8.4)
IC235	29465	2	14	H14	104.2 (16.7)	1.7 (0.1)	105.9 (16.8)	65.6 (3.1)
IC236	29466	2	14	H14	105.5 (15.3)	1.6 (0.7)	107.1 (16)	53.2 (7.9)
IC244	29473	1	17	H15	104.3 (10.3)	1.7 (0.1)	106 (10.5)	64.9 (4.2)
IC245	29474	2	23	H10	96.3 (6.2)	1.9 (0.6)	98.3 (6.8)	48.5 (3.9)
IC246	29475	2	23	H10	100 (8.6)	2.2 (0.2)	102.2 (8.8)	54.4 (5.7)
IC247	29476	2	23	H10	94.2 (12.5)	2 (0.3)	96.2 (12.8)	48.2 (0.6)
IC252	29481	2	24	—	0 (0)	0 (0)	0 (0)	0 (0)
IC253	29482	2	24	—	0 (0)	0 (0)	0 (0)	0 (0)
IC254	29483	2	24	—	0 (0)	0 (0)	0 (0)	0 (0)
IC258	29487	2	25	H15	168.3 (28.7)	4.4 (0.6)	172.7 (29.2)	155.7 (51.9)
IC259	29488	2	25	H15	134.3 (5)	3.8 (0.8)	138.1 (5.8)	92.9 (2.8)
IC260	29489	2	25	H15	210.6 (35.5)	6 (0.8)	216.6 (36.3)	115.6 (25.6)
IC263	29492	2	26	H26	0.8 (0.1)	0 (0)	0.8 (0.1)	22 (3.4)
IC264	29493	2	26	H26	0.5 (0.1)	0 (0)	0.5 (0.1)	25.2 (9.5)
IC265	29494	2	26	H26	0.1 (0.1)	0 (0)	0.1 (0.1)	22.4 (3)
IC267	29496	2	27	H6	109 (41.4)	2.5 (1.3)	111.4 (42.7)	93.1 (12.9)
IC268	29497	2	27	H6	141.2 (50)	3.8 (2)	145 (52)	117.2 (18.3)
IC269	29498	2	27	H6	79 (37)	2.5 (1.3)	81.5 (38.3)	79.9 (4.3)
IC270	29499	2	28	H20	72.5 (7.2)	3.3 (0.7)	75.8 (6.6)	44.9 (2.6)
IC271	29500	2	28	H20	89.4 (6.5)	2.1 (1)	91.5 (7.5)	53.9 (3.1)
IC272	29501	2	29	H22	64.2 (6.7)	1.2 (0.2)	65.5 (6.7)	85.8 (12.3)
IC273	29502	2	29	H22	53.7 (10.7)	0.4 (0.1)	54.1 (10.8)	74.8 (6.6)
IC274	29503	2	30	H19	4.8 (1.5)	0.1 (0.1)	4.9 (1.5)	53.4 (8.2)
IC275	29504	2	30	H19	2.4 (1.3)	0.1 (0)	2.5 (1.3)	55.5 (2.5)
IC276	29505	2	31	H11	0.4 (0.1)	0 (0)	0.4 (0.1)	55.3 (6.4)
IC277	29506	2	32	H1	0 (0)	0 (0)	0 (0)	99.6 (20.7)
IC278	29507	1	33	H15	99.4 (46.2)	1.8 (1.1)	101.2 (47)	76.8 (4.8)
IC279	29508	2	34	H2	12.9 (1.1)	0.1 (0)	13 (1.2)	53.2 (2.9)
IC280	29509	2	35	H7	36.9 (7.8)	1.3 (0.5)	38.1 (8)	46 (7.3)
IC281	29510	2	36	H14	154 (8.6)	2.6 (0.5)	156.6 (8.6)	85 (3.8)
IC282	29511	1	37	H19	164.6 (26.8)	5.5 (1)	170 (27.8)	46.4 (3.2)
IC283	29512	1	38	H24	9.6 (1.4)	0.1 (0)	9.8 (1.4)	10.2 (1.1)
IC284	29513	2	39	H22	24.6 (3.3)	0.4 (0)	25 (3.3)	126 (6.5)
IC285	29514	2	40	H26	185.3 (47.1)	1.5 (0.8)	186.8 (47.8)	18.9 (3.5)
IC286	29515	2	41	H2	24.7 (4)	0.7 (0.1)	25.4 (4.1)	20.3 (1.1)
IC287	29516	2	42	H18	104.1 (27.5)	2.3 (1.2)	106.4 (28.7)	75 (0.9)
IC288	29517	2	43	H5	97.4 (18.3)	1.8 (0.6)	99.2 (18.8)	87.7 (4.4)
IC289	29518	1	44	H4	40.2 (26.1)	0.7 (0.4)	40.9 (26.5)	85.9 (14.2)
IC290	29519	2	45	H7	15.3 (4.5)	0.3 (0.1)	15.6 (4.6)	48.2 (3.4)
IC291	29520	1	46	H25	47.6 (13.9)	0.3 (0.1)	47.9 (14.1)	43.7 (5)
IC292	29521	2	47	H21	105.6 (24.5)	2.8 (0.4)	108.3 (25)	97.5 (14.6)
IC293	29522	1	48	H8	54.2 (12.7)	1 (0.2)	55.2 (12.9)	161.5 (20.1)
IC294	29523	2	49	H23	19.6 (5)	0.1 (0.1)	19.7 (5)	76.7 (2.7)
IC295	29524	2	50	H19	164.6 (17.4)	4.7 (0.8)	169.3 (18.2)	60.8 (2.8)
IC296	29525	1	51	H12	0.3 (0.1)	0 (0)	0.3 (0.1)	58.6 (5.5)
IC297	29526	1	52	H9	25.7 (10.9)	0.8 (0.4)	26.5 (11.3)	39.4 (5.7)
IC298	29527	1	53	H22	18.7 (4.3)	0.2 (0)	18.8 (4.3)	44.6 (8.1)
IC299	29528	2	54	H17	36.2 (19.2)	0.8 (0.5)	37 (19.7)	125.7 (12.1)
IC300	29529	1	55	H14	51.4 (10.9)	0.6 (0.2)	52 (11.1)	52.3 (4.4)
IC301	29530	1	56	H22	16.8 (3.3)	0.1 (0)	17 (3.3)	100.3 (17.6)
IC302	29531	2	57	H17	83.2 (16.1)	1.7 (0.6)	84.9 (16.6)	55.6 (3.1)
IC303	29532	1	58	H16	94.9 (15)	1.8 (0.2)	96.7 (15.1)	46.1 (2.2)

Table 1 *Continued*

IC	NRRL	*MAT*	VCG	Haplotype	B_1	B_2	Total aflatoxins	CPA
IC304	29533	2	59	H15	69.6 (12.2)	1.5 (0.7)	71.1 (12.9)	114.2 (5.3)
IC305	29534	1	60	H27	37.4 (10.5)	0.5 (0.2)	37.9 (10.7)	17.5 (1.6)
IC306	29535	2	61	H2	11.5 (0.7)	0.2 (0.1)	11.7 (0.8)	12.9 (1.9)
IC307	29536	2	62	H13	139.8 (5.1)	3.1 (0.1)	142.9 (5.1)	77.2 (1.5)
IC308	29537	1	63	H3	39.8 (6.5)	0.6 (0.1)	40.4 (6.6)	183.5 (18.3)

*Isolates were sampled from a Georgia peanut field in 1992 (Horn & Greene 1995). Culture collection designations: IC (I Carbone), NRRL (Agricultural Research Service Culture Collection, Peoria, Illinois). Mating-type designations are from Ramirez-Prado *et al.* (2008) and vegetative compatibility groups are based on Horn & Greene (1995). Mean mycotoxin concentrations (μg/mL; $n = 3$) are shown; standard deviations are indicated in parentheses. Strains were grown on yeast extract–sucrose broth in vials at 30 °C and cultures were analysed for mycotoxins with high-performance liquid chromatography (Horn *et al.* 1996). Isolates in vegetative compatibility group (VCG) 24 (IC252, IC253 and IC254) are missing aflatoxin cluster genes, precluding inference of multilocus haplotypes.

Table 2 Oligonucleotide primer sets used for amplification and sequencing of 21 cluster regions in *Aspergillus flavus*

Target region	Primer name	Primer sequence	Intergenic/ gene* (bp)
aflT/aflC	*aflTF-XIR / aflCR-XIR*	AACTGGTCCAACCGGAGTAC / CGGCAGATACAGTCATGGAC	390/0
aflC/aflD	*aflCF-XIR / aflDR-XIR*	GGTTCGAGCCAACTTGTGAT / GACGTTGGAGAAAAGCTTCAAT	340/0†
aflD/aflA	*aflDF-XIR / aflAR-XIR*	AACAACTGTCGCAGACAGTGT / TGAGATCGAGCATGGAGGTA	265/0
aflA/aflB	*aflAF-XIR / aflBR-XIR*	AATTGCCGTCTTCAGCTTTC / GTGTTGGATGCCACGTCTAG	534/200 (*aflA*)
aflB/aflR	*aflBF-XIR / aflRR-XIR*	GAGGGGAGATTGAGCCTTATC / CCAGTCGCTGGTGAACCTAT	183/0
aflR/aflS	*aflRF-XIR / aflSR-XIR*	CCTGGCTGAAGGAAGACTCT / CTGGCGAGGGCTAATACTTG	529/192 (*aflR*)
aflS/aflH	*aflSF-XIR / aflHR-XIR*	GGTCAGTCTGAGCGATCTCTC / AGCAGACGTAGTGGACGTGT	311/8 (*aflS*)
aflH/aflJ	*aflHF-XIR / aflJR-XIR*	ACCTTCTTGCTCCTTGGTTC / CCGTAGCGCGTAGCTAATGTA	523/63 (*aflH*)
aflJ/aflE	*aflJF-XIR / aflER-XIR*	AAGGCTCCTGAGACTCGCTA / GTTCGACTTGATCTTTTGCG	0/456 (*aflJ*)
aflE/aflM	*aflEF-XIR / aflMR-XIR*	GAACCATTTGACGTCGGATT / GTTGGCCTTGATCTGTTGAA	336/59 (*aflE*)
aflM/aflN	*aflMF-XIR / aflNR-XIR*	GCTTGGCTCTCTCCTTTGAA / GCTGCTGAGGGAGTTGAAAC	546/506 (*hypE*)‡
aflN/aflG	*aflNF-XIR / aflGR-XIR*	TATTACGCCAGCATACGATGA / CGGTTGATCCTAGTCAAGCTT	428/5 (*aflN*) †
aflG/aflL	*aflGF-XIR / aflLR-XIR*	ATAGCTCATTGGGTGCGATT / AGGAACCGTACAAGTACGACAA	454/163 (*aflG*)
aflL/aflI	*aflLF-XIR / aflIR-XIR*	AGAGAACTGGCTCGCCATAG / GCCACTGGTAGTGTCCATCA	450/103 (*hypB*)§
aflI/aflO	*aflIF-XIR / aflOR-XIR*	GCACTATCCTGGCACAATGT / TTCTTCGAGTCACAGCCTATTC	358/278 (*aflO*)
aflO/aflP	*aflOF-XIR / aflPR-XIR*	CTGCTCAATCGCATAACCAC / CAGCGCTTCATGAATCAGAT	350/12 (*aflO*)
aflP/aflQ	*aflPF-XIR / aflQR-XIR*	GTGGAGGATGGAGTCCTCTCT / CCAGCAGCTTCTCCAGAATC	349/20 (*aflP*)
aflQ/aflK	*aflQF-XIR / aflKR-XIR*	GGTCAGCTGCTTCTTCATCTC / CGTAGTCGAATGACTGTCCC	284/0
aflK/aflV	*aflKF-XIR / aflVR-XIR*	CTATGGGCAAAGCAGATGATT / CAGAAGAGGCGAAAATGTCA	409/0
aflV/aflW	*aflVF-XIR / aflWR-XIR*	GGGACTGAAATATGCGGTTT / GGAGCAAAGGGGTAGGTGTAG	331/133 (*aflV*)
aflW/aflX	*aflWF-XIR / aflXR-XIR*	GCACACGGTGTGGAAAGATA / GACTAGTGCACGATGTGCAAC	529/296 (*aflW*)

*The ratio of the intergenic region to gene (exons plus introns).
†*hypC* (GenBank accession XM-002379909) and *hypD* (GenBank accession XM-002379897) do not overlap with our sequenced regions.
‡About 506 bp of the total length overlap with a hypothetical gene, *hypE* (GenBank accession XM-002379899). Of these 506 bp, only 414 bp are coding and 92 bp are noncoding (intron).
§About 103 bp of the total length overlap with *hypB* (GenBank accession XM-002379894), and 61 bp of an intron.

evidence of trans-species evolution in *hypE*, a protein-encoding gene of unknown function between *aflM* and *aflN* in the aflatoxin gene cluster of *A. parasiticus* (Carbone *et al.* 2007a). The size of recombination blocks for the concatenated multiple DNA sequence alignment depended on the number of contiguous pairs of sites that were both strongly correlated ($0.8 < r^2 < 1$) and significantly linked ($P < 0.01$). The actual sizes of the blocks were determined by performing LD analysis on complete cluster sequences for *A. flavus* AF13 (L strain, AY510451), AF36 (L strain, AY510455) and NRRL 3357 (L strain; http://www.aspergillusflavus.org/), and for *A. oryzae* NRRL 5590 (Machida *et al.* 2005).

We further examined patterns of incompatibility among pairs of sites using the four-gamete test (Hudson & Kaplan 1985) implemented in SNAP Clade. This

analysis may reveal one or more blocks with distinct evolutionary histories that need to be examined separately in coalescent analyses that assume no recombination (described below). Recombination hotspots, defined as regions in the cluster that are more prone to recombination, were deduced by mapping the minimum number of recombination events (R_h) calculated using the program RECMIN (Myers & Griffiths 2003). An estimate of the population recombination rate, per base pair, within the cluster was obtained using Hey and Wakeley's γ estimator (Hey & Wakeley 1997) as implemented in the SITES program. A minimal ancestral recombination graph (ARG) for all SNPs spanning the cluster was inferred using the branch and bound method implemented in Beagle (Lyngsø *et al.* 2005); if there were too many recombination events for branch and bound to go to completion, we used the program KWARG (http://www.stats.ox.ac.uk/~lyngsoe/section26/), which is a heuristic implementation of the Beagle algorithm. The ARG was useful in examining the ancestral history of haplotypes and provided a rough estimate of haplotype age; in this case, we define older haplotypes as requiring more mutations and recombination events before coalescing to a recent common ancestor in the ARG.

Linkage disequilibrium and compatibility analyses were performed for the 21 sequenced regions in complete clusters and for homologous regions in partial gene clusters. The latter was important to determine if recombination block boundaries are conserved among complete and partial clusters. This would indicate that complete and partial clusters share a history of recombination and possibly that partial clusters are derivatives of full cluster ancestors.

Telomeric regions and G + C content

Several studies report on the importance of recombination and G + C content in the maintenance and disassembly of gene clusters (Howlett *et al.* 2007; Patron *et al.* 2007). We examined G + C content distribution along the right arm of chromosome 3 (200 kb) for the sequenced genome strain of *A. flavus* NRRL 3357 to identify regions or islands that undergo an abrupt change in G + C content (Zhang & Zhang 2004). In bacteria, sharp transitions in G + C content have been associated with horizontal transfer of genomic islands (Zhang & Zhang 2004). We used a windowless method to display the cumulative GC profile (or z' curve), which is reported to have better resolution than window-based methods in identifying putative boundaries of genomic islands (Zhang & Zhang 2004). The z' curve displays variations in G + C content along a chromosome based on the cumulative count of G and C bases. A z' curve that is approximately a straight line indicates a constant G + C content; a jump (positive slope) in the z' curve indicates a decrease in G + C content, whereas a drop (negative slope) in the z' curve indicates an increase in G + C content.

Trans-species evolution

To reconstruct patterns of trans-species evolution, we expanded the *A. flavus* sample to include 24 *A. parasiticus* isolates examined previously (Carbone *et al.* 2007a), two representative nonaflatoxigenic strains with complete clusters (*A. flavus* AF36 and *A. oryzae* NRRL 5590), and eight nonaflatoxigenic *A. flavus* strains that represent distinct deletion patterns in the aflatoxin cluster (Chang *et al.* 2005). Because we are interested in detecting the targets of balancing selection in *A. flavus*, we initially restricted our analysis to regions in the *A. flavus* cluster harbouring uninformative polymorphisms, which are easily identifiable as strongly correlated pairs of sites in the LD plot with no significant linkage. This is important as recombination will uncouple polymorphisms, thus making it difficult to identify chemotype-specific haplotypes. The two noncluster genes, *trpC* and *amdS*, are orthologous and provided independent estimates of the species phylogeny. Phylogenies were inferred using unweighted parsimony in PAUP* 4.0 (Swofford 1998) and rooted by specifying *A. nomius* as the outgroup (Peterson 2008). If parsimony searches yielded more than one equally parsimonious tree, support for branches was assessed using 500 bootstrap samples. We also inferred a strict consensus tree and resolved polytomies by reconstructing ARGs. The ARG for the aflatoxin cluster region under balancing selection was examined more closely to determine if the effects of mutation and recombination result in significantly different toxin profiles, which would be expected if the region is important in maintaining chemotype-specific differences in *A. flavus*. Recombination between chemotype-specific lineages throughout the cluster is the basis for our detection of recombination blocks, as was shown in *A. parasiticus* (Carbone *et al.* 2007a). We examined the ARGs for evidence of recombination between chemotype-specific lineages, or in the case of *A. flavus*, between aflatoxigenic and nonaflatoxigenic lineages.

Neutrality tests and coalescent analyses

Departures from neutrality and population-size constancy were tested separately for each of the 21 cluster regions using Tajima's D (Tajima 1989), Fu and Li's D^* and F^* (Fu & Li 1993) and Fu's F_S (Tajima 1989; Fu & Li 1993; Fu 1997). Significance thresholds were Bonferroni-corrected by dividing by the total number of neutrality tests ($n = 84$). Significant values of Tajima's D and Fu &

Li's D^* and F^* and nonsignificant values of Fu's F_S (Fu 1997) are indicative of balancing selection. One effect of balancing selection is to preserve polymorphisms in two or more lineages over an extended period, resulting in strong genetic differentiation and positive values of D and F_S, which should not be interpreted as population subdivision or reproductive isolation. Negative values of D and F_S would indicate population growth or a selective sweep.

To reconstruct the ancestral history of complete and partial aflatoxin gene clusters, we examined molecular sequence variation in conserved cluster regions as well as in adjacent noncluster genes. This is particularly relevant because the genes required for the biosynthesis of CPA are immediately adjacent to the aflatoxin cluster (Chang *et al.* 2009). Specifically, we wanted to ascertain whether gene loss was occurring only for genes in the aflatoxin cluster or included genes in the CPA cluster and possibly other telomeric regions. Identifying conserved genes that flank complete and partial clusters would indicate a process of gene loss in complete clusters giving rise to partial clusters. This was investigated using two approaches. First, we focused on cluster genes and reconstructed the ancestral history of complete and partial clusters by simulating mutational histories using the coalescent. Second, we inferred rooted gene genealogies for two adjacent noncluster regions that are conserved in complete and partial aflatoxin clusters. This was important to further corroborate inferences of gene gain or loss in the cluster. We determined the largest nonrecombining partition for complete and partial aflatoxin gene clusters in *A. flavus* using SNAP CladeEx (Bowden *et al.* 2008). As in *A. parasiticus*, the recovery of *A. flavus* isolates belonging to the same VCG across the USA (Horn & Dorner 1999) supports a population model with migration and random mating (Bayman & Cotty 1993). Assuming panmixis and constant population size, coalescent analysis was performed using GENETREE version 9.0 (http://www.stats.ox.ac.uk/~griff/software.html). This involved generating all possible rooted gene genealogies for the largest nonrecombining partition in the cluster and then calculating the relative probabilities of all rooted genealogies by performing 1–10 million simulations of the coalescent assuming Watterson's (1975) estimate of θ, panmixis and constant population size.

Results

Molecular sequence variation

We sequenced a total of 8351 bp for 21 regions of the aflatoxin gene cluster, 323 bp for the 5′-untranslated region of *trpC* and 333 bp (285 bp exon and 48 bp intron) for *amdS* in our subsample of 44 *A. flavus* strains. For the additional strains within VCGs, we examined only 14 of the 21 regions. Examination of the additional strains excluded most of the early pathway regions between *aflT* and *aflJ* (see Fig. 1), which had low nucleotide diversities per site (π = 0.00041–0.00496) comparable to noncluster genes, *trpC* (0.00458) and *amdS* (0.00342). The late pathway regions from *aflE* to *aflW* had ~10-fold higher nucleotide diversities per site (0.00253–0.0489). Similar nucleotide diversity estimates were obtained when alignments for late pathway genes included aflatoxin deletion pattern strains. We also sequenced 700 bp of *aflC* for four nonaflatoxigenic isolates (IC277, IC309, IC310 and IC312). Only one isolate (IC309) showed the same nonsense mutation as in AF36; five other polymorphisms encoded three synonymous and two replacement substitutions, and were present in both aflatoxin- and nonaflatoxin-producing strains. All sequences have been deposited in GenBank under Accession numbers FJ877157–FJ878623, GQ456072–GQ456075 and GQ479324–GQ479333. A total of 27 haplotypes were inferred for the 21 aflatoxin-cluster regions (Tables 1 and 2). The 14 regions from *aflS* to *aflJ* and *aflE to aflW* examined in additional representative strains within VCGs yielded 24 distinct haplotypes, indicating that 89% of all haplotype-specific variation in the aflatoxin cluster of *A. flavus* resides in the late pathway genes.

Linkage disequilibrium and compatibility analyses

Linkage was examined for 453 polymorphisms distributed across the 21 regions of the aflatoxin gene cluster in *A. flavus* (Table 2). We identified six distinct blocks spanning the intergenic regions of late-pathway genes: *aflE/aflM*, *aflM/aflN (hypE)*, *aflN/aflG*, *aflG/aflL*, *aflL/aflI (hypB)* and *aflI/aflO* (Table 2; Fig. S1). The early pathway regions from *aflT* to *aflE* (not inclusive) accounted for only 8% (36/453) of the total nucleotide sequence variation. Only one 3′-untranscribed region, *aflW/aflX*, showed a large number of fixed polymorphisms for a nonaflatoxigenic haplotype H1 (IC277); all other strains shared the consensus sequence (Fig. S1). In Fig. S2, the six putative blocks identified from the 21 sequenced regions are shaded in the LD plot of complete cluster sequences in which only the lower triangular portion of the LD plot, showing correlations among pairs of sites, can be used for inference of block structure as the significance of linkage cannot be assessed with only four strains. The upper triangular portion shows compatibility relationships for all pairs of sites. As expected, blocks are larger in size and may span several intergenic regions and genes when based on full cluster sequences (Fig. S2). For example, block 1 includes all

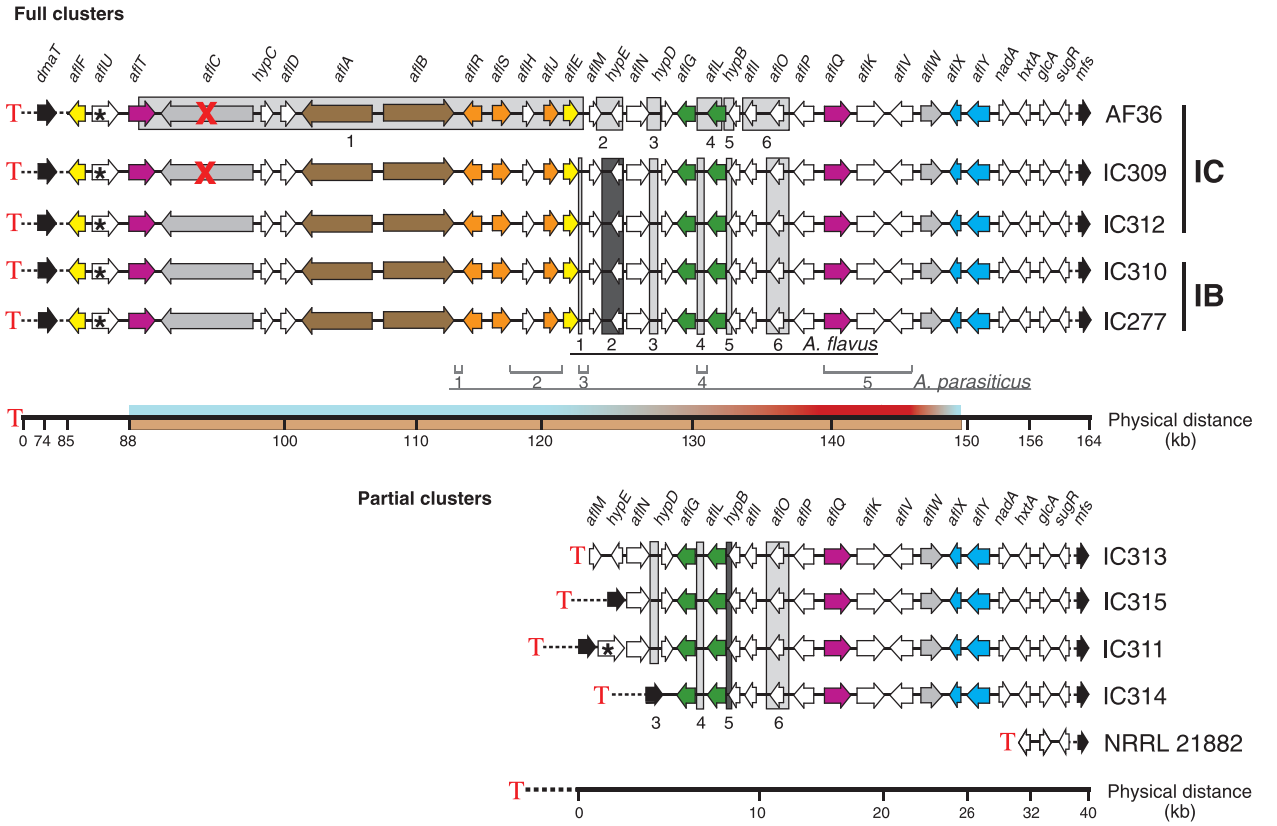

Fig. 1 Schematic representation of full and partial aflatoxin gene clusters in *Aspergillus flavus*. The aflatoxin gene cluster starting with *aflF* is ~85 kb from the telomeric end of chromosome 3R, which is indicated with a 'T'. A total of 18 putative transcripts have been identified in this telomeric region (Chang *et al.* 2009). One of these transcripts, encoding for dimethylallyl tryptophan synthase (*dmaT*) and involved in CPA biosynthesis, is ~10 kb from *aflF* and is present in all complete *A. flavus* gene clusters, as well as in three of the partial cluster deletion strains (IC311, IC314 and IC315). As reported by Chang *et al.* (2005), only one of these partial cluster strains (IC311) synthesizes CPA; IC314 and IC315 are reported to be missing two other genes important in CPA biosynthesis: monoamine oxidase (*maoA*) and the hybrid polyketide and nonribosomal peptide synthase gene (*pks–nrps*). The nonsense mutation in *aflC* for AF36 and IC309 is designated with a 'X'. Genes shown in the same colour arise from gene duplication (Carbone *et al.* 2007b). The six linkage disequilibrium (LD) blocks in *A. flavus* from the concatenation of 21 cluster regions are boxed and numbered; the position of the five *A. parasiticus* blocks is shown below the *A. flavus* blocks and is from a previous study (Carbone *et al.* 2007a). The real size of LD blocks in *A. flavus* is shown in AF36 (e.g. block 1 spans *aflT-aflE*). LD blocks that have a different evolutionary history (i.e. incompatible partitions) are shown in a darker shade of grey (e.g. block 2 in complete clusters and block 5 in partial clusters). The upper scale bar shows real distance (kb) from the telomere across the CPA (*dmaT* only) and aflatoxin gene clusters, and is based on the genome strain of *A. flavus* NRRL 3357. The dashed segments between the telomere and *dmaT* in full and partial cluster strains denote variation in the length of the subtelomeric region. The scale bar for the partial cluster strains shows only real distance between contiguous aflatoxin cluster genes. A recombination hotspot in full clusters is shown in red above the scale bar and recombination coldspots are shown in blue. Full cluster strains belonging to groups IC and IB (see Fig. 4) are indicated; all partial cluster strains are in IB.

the early pathway genes from *aflT* to *aflE*, blocks 2–5 each have one boundary in a gene and the other in an intergenic region, and block 6 extends from the middle of *hypB* in *aflL/aflI* to the *aflO/aflP* intergenic region (Fig. S2). There is a rapid decay in LD with physical distance between *aflP* and *aflW*, as revealed in the LD plot for the 21 concatenated regions (Fig. S1) and for the four complete *A. flavus* clusters (Fig. S2). LD analysis for homologous regions in full and partial clusters could only be performed for regions spanning blocks 3–6; only block 5 was incompatible with the other blocks (Fig. S3; Table S1).

Compatibility analysis using RECMIN revealed a minimum of 24 recombination events across the 21 concatenated regions. Recombination events were unevenly distributed, with 3, 6, 13 and 2 inferred breaks spanning *aflT-aflE*, *aflE-aflP*, *aflP-aflW* and *aflW-aflX* regions respectively (Fig. S1). This results in an overall population recombination rate of 0.000394 per generation per base pair for intergenic regions. Based on significance

of linkage and compatibility analysis with CladeEx, the six blocks could be grouped into two distinct evolutionary histories: block 2 (shaded in Fig. S1) and combined blocks 1, 3, 4, 5 and 6. LD breaks down completely between *aflP* and *aflW* but is restored from *aflW* to *sugR* (Figs S1 and S2).

The ARG indicates extensive recombination in the ancestral history of the aflatoxin cluster in *A. flavus* (Fig. 2). Although many recombination and mutational events are required to interconnect all haplotypes to a most recent common ancestor in the past, it is clear from the ARG that many haplotypes are very closely related and recently evolved, often arising from only single crossovers and one to three mutations. For example, proceeding from left to right in the ARG: clades (H15, H16), (H19, H14, H20), (H22, H23) and (H24, H26) are separated by mutation alone; haplotypes H21, H18 and H25 arise from a single crossover; and H6, H5, H12 and H11 have evolved by both mutation and recombination.

Telomeric regions and G + C content

The distribution of G + C content along 200 kb of the right arm of chromosome 3 is shown in Fig. 3. There is a strong bias for higher G + C content as the distance from the telomere increases and G + C content reaches a maximum with the first gene in the aflatoxin cluster, *aflF*, shown in the *z'* curve as a sharp downward peak. The downward peak is followed by a sharp decrease in G + C content seen as an upward slope in the *z'* curve until reaching *aflE*, after which G + C content stabilizes and follows a sinusoidal *z'* curve ending with *aflY*. The end of the cluster marks another GC-rich region. Mapping the distribution of recombination blocks and hotspots below the *z'* curve (shown as a highlighted bar in Fig. 3) shows that the early pathway genes (*aflF* to *aflE*), with reduced sequence diversity and significant LD, have higher G + C content than the late pathway regions (*aflE* to *aflY*), which harbour the six LD blocks and a recombination hotspot. There is a localized increase in G + C content in the region of the recombination hotspot from *aflI* to *aflW*. Similarly, the G + C content of the genes involved in CPA biosynthesis (*maoA*, *dmaT*, *pks–nrps*) is relatively constant until after the *maoA* gene where it reaches a maximum with *aflF*.

Trans-species evolution

A total of 17, 11 and 12 haplotypes were inferred for the *aflW/aflX*, *trpC* and *amdS* regions respectively, for the expanded *A. flavus* and *A. parasiticus* sample. The haplotype distribution in these regions is shown in Fig. 4 (Table S2). Phylogenetic analysis of the *aflW/aflX* region revealed two distinct *A. flavus* lineages (Fig. 4). One lineage containing both aflatoxigenic and nonaflatoxigenic strains, including AF36, shares a more recent common ancestor with *A. parasiticus* than a second distinct *A. flavus* lineage, which is completely nonaflatoxigenic and includes complete/partial cluster strains and NRRL 21882. These two *A. flavus* aflatoxin gene lineages, or clades, correspond to groups IC and IB respectively (Geiser *et al.* 2000). As expected, these patterns of trans-speciation were not observed in *trpC* and *amdS* phylogenies. The partitioning of *A. flavus* into clades IB and IC in *aflW/aflX* was incongruent with the monophyly of *A. flavus* in the *trpC* and *amdS* organismal genealogies. Figure S4 (Table S2) shows the strict consensus phylogenies and ARGs inferred using branch and bound (beagle) and heuristic (kwarg) search methods. Although we observed three distinct clades in *aflW/aflX* when we assume no recombination, the ARGs indicate a history of genetic exchange and recombination. For example, in the branch and bound and heuristic ARGs for *aflW/aflX* (Fig. S4, Table S2), haplotype H9 arises from a single crossover with putative parents haplotype H8, belonging to the strictly nonaflatoxigenic clade, and haplotype H14, which is a member of the *A. flavus* clade sharing a recent common ancestor with *A. parasiticus* (paths to parents determined by lack of mutations and additional recombination events are shown in bold in both ARGs in Fig. S4). Also, *aflW/aflX* shows evidence of an *A. parasiticus* G_1-dominant (high G_1/B_1 ratio) haplotype H5 acting as one of the parents in a recombination event in the ancestor of all *A. parasiticus* haplotypes represented in H2, H3 and H4. The *A. parasiticus* OMST-producing haplotypes H9 and H10 show a recent history of recombination with *A. flavus* in the ARGs for *amdS* and *trpC* respectively.

Neutrality tests and coalescent analyses

All neutrality tests were nonsignificant after Bonferroni correction. Negative values of D and F_S were obtained for early pathway genes spanning *aflT* to *aflE* (13 of 18 tests) whereas positive values of D and F_S were observed among late pathway genes from *aflE* to *aflX* (18 of 24 tests). Collectively, this reversal in values indicates that recombining regions in the cluster are under strong balancing selection for maintaining chemotype-specific differences – in this case aflatoxigenicity vs. nonaflatoxigenicity – whereas regions of low recombination spanning the early pathway genes are swept to fixation by selection.

The largest nonrecombining partition for complete and partial clusters included the intergenic regions: *aflI/aflO*, *aflO/aflP*, *aflP/aflQ*, *aflQ/aflK*, *aflK/aflV*, *aflV/aflW* and *aflW/aflX*. This corresponds to a total concatenated sequence length of 2605 bp, 184 polymorphic

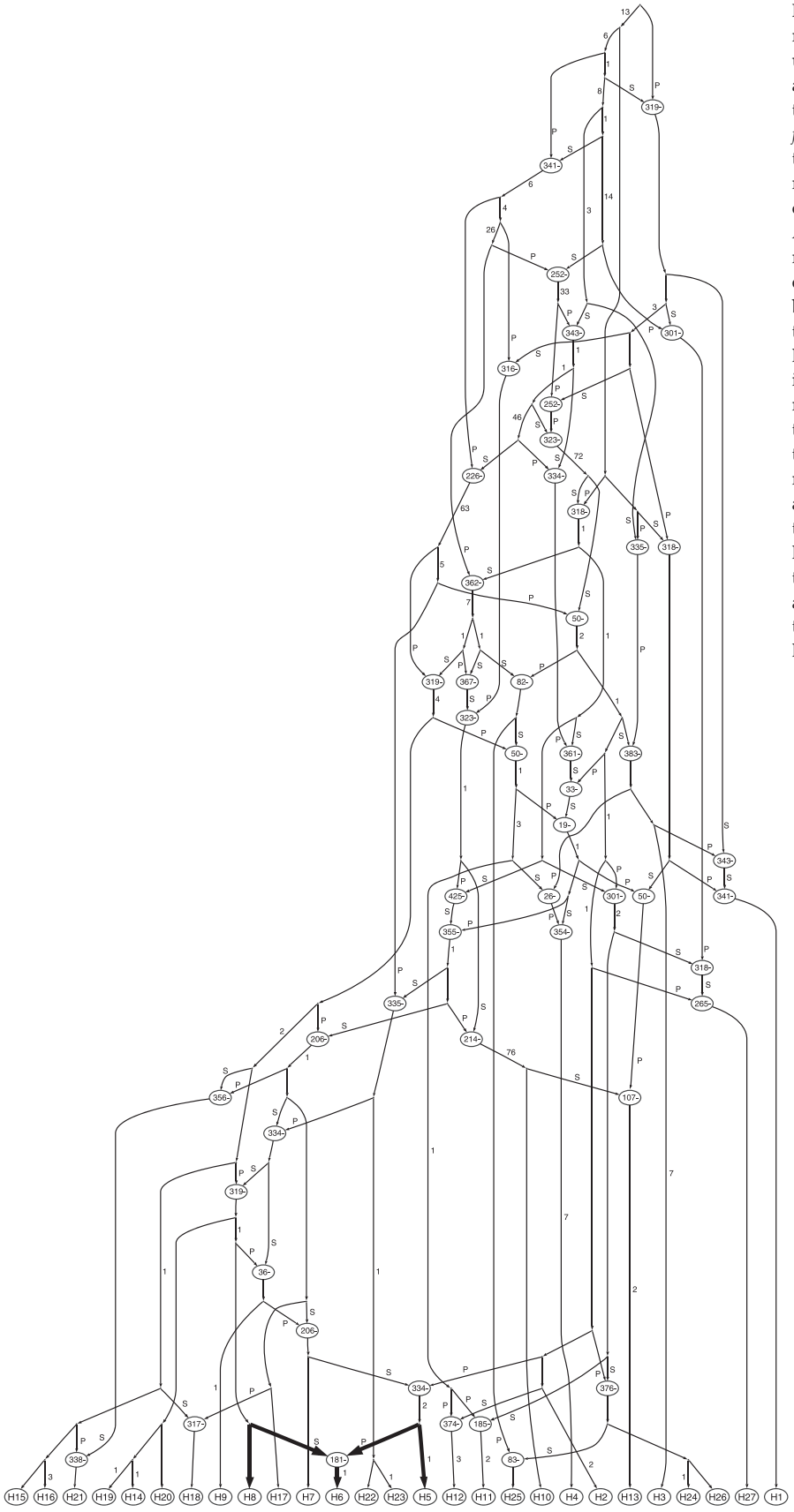

Fig. 2 One possible minimal ancestral recombination graph (ARG) inferred using the Kwarg heuristic method for all polymorphisms in the 21 regions of the aflatoxin gene cluster in *Aspergillus flavus*. The ARG shows possible mutation and recombination paths giving rise to the sampled haplotypes. The direction of paths is from the top of the ARG (past) to the bottom (present); moving backwards in time one of three events (mutation, coalescence or recombination) is possible. The paths leading to the recombination nodes (ovals) are labelled with a P (prefix) or S (suffix), indicating the 5′ and 3′ segments of the recombinant sequence respectively; the number labels on the paths indicate the number of polymorphisms. The number in the ovals indicates the variable position immediately to the left of the recombination breakpoint. The paths leading to the putative parental haplotypes (H8, H5) of the recombinant (H6) are thickened. The ancestral recombination graph (ARG) was rooted with haplotype H1 (IC277).

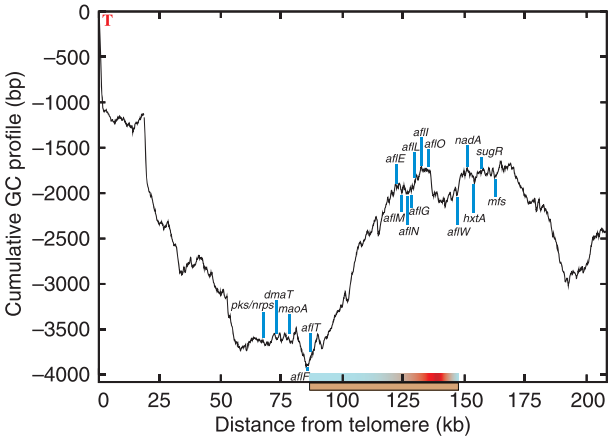

Fig. 3 The cumulative GC profile (or *z*′ curve) for 200 kb of chromosome 3R in *Aspergillus flavus* NRRL 3357. The AT-rich telomere is indicated with a 'T'. On the *y*-axis, −4000 bp indicates the lowest point in the *z*′ curve at a distance of 85 kb from the telomere, and signifies that ~4000 G + C bases are in excess of the average G + C content calculated for this subregion. G + C content increases with distance from the telomere and reaches a maximum with the first gene in the aflatoxin cluster, *aflF* at ~85 kb, and then a minimum at ~133 kb with *aflI*. The higher G + C content at ~85 kb also coincides with the 3′ end of the CPA cluster, after *maoA*; the other maximum at 190 kb marks the end of the aflatoxin and sugar clusters. The scale bar shows the distribution of recombination hotspots and coldspots in the aflatoxin cluster; the recombination hotspot is localized in a region of relatively higher G + C content at ~140 kb in the late pathway genes.

sites and 20 haplotypes (Fig. 5; Table S3). The *aflG/aflL* region could not be amplified in IC314 and was excluded in the concatenation. The *aflL/aflI* region was incompatible with the largest nonrecombining partition and excluded from the analysis. The genealogy with the highest root probability (Fig. 5; Table S3) shows two distinct evolutionary lineages (IB and IC) as observed in the *aflW/aflX* region (Fig. 4). The strong genetic differentiation observed in the cluster is the result of balancing selection acting on a nonaflatoxigenic lineage (IB, haplotypes in bold type in Fig. 5 and Table S3) that includes all the aflatoxin cluster deletion strains from the sample we examined. The rooted genealogy shows that partial clusters are recent descendents of full clusters. For example, IC311 (H13) and IC315 (H14) from Alabama share a recent common ancestor and descended from a full cluster ancestor with a mutation configuration similar to IC277 (H17). Similarly, partial deletion strain IC314 (H10, which includes *A. oryzae* NRRL 5590) and full cluster strain IC310 (H8) from North Carolina share a recent common ancestor (Fig. 5).

Similar patterns of descent among full and partial aflatoxin clusters were observed in regions flanking the aflatoxin gene cluster (Fig. 6; Table S4). In the *dmaT* gene genealogy, there was no clear separation of partial and full aflatoxin cluster strains; however, the genealogy does reveal two distinct evolutionary lineages similar to the balancing selection observed in the aflatoxin late pathway genes. In the *mfs* gene genealogy, the full-cluster nonaflatoxigenic haplotype H8 (IC277) is ancestral to partial cluster haplotype H9 (IC315) and to haplotype H1, which includes a mixture of partial (IC313, IC314) and full cluster strains (*A. oryzae* NRRL 5590, IC310). In rooting genealogies, haplotypes with no mutations along the edges of the genealogy (branches) occupy interior positions and serve as a root from which all mutation paths to descendant haplotypes can be retraced (Griffiths & Tavaré 1994). The immediate descendents of H1 are partial cluster haplotypes H6 (IC311) and H11 (IC252 and NRRL 21882). There is no signature of balancing selection in *mfs* positioned at the 3′ distal end of the cluster.

Discussion

The early portion of the cluster from *aflT* to *aflE* (~35 kb) showed 10-fold lower nucleotide diversity than the late pathway segment from *aflE* to *aflP* (~13 kb). The reduced sequence diversity could be the result of functional constraints on pathway genes such as *aflC*, selection acting on adjacent noncluster genes such as *dmaT*, or differences in gene regulation or chromatin organization. We cannot rule out genetic hitchhiking effects due to the proximity of the CPA and aflatoxin gene clusters independent of whether they are co-regulated or not. In the absence of recombination, we would predict adjacent genes or clusters to be dragged through the selection process. With recombination between *maoA* and *aflF*, as well as in the late aflatoxin pathway genes starting with *aflE/aflM*, we would predict that hitchhiking would break down at these junctions. This is consistent with the observed variation in mycotoxin production, with individuals producing both aflatoxins and CPA, aflatoxins alone, CPA alone or neither mycotoxin. Only 25% of nonaflatoxigenic strains examined in this study had the nonsense mutation in *aflC*. In the study by Criseo *et al.* (2008), gene loss and nucleotide variation in the early pathway genes *aflD* and/or *aflR* comprised ~20% of the 134-nonaflatoxigenic *A. flavus* strains that they examined. This suggests that the majority of nonaflatoxigenic strains have early pathway genes. The abrupt transition in G + C content in this region (Fig. 3) indicates the potential for the entire segment from *aflT* to *aflE* to be exchanged or lost as a single block during the shuffling process (Fig. 1). Only 12% (3/24) of recombination events fall in the early part of the pathway, which further supports the tight linkage of these genes.

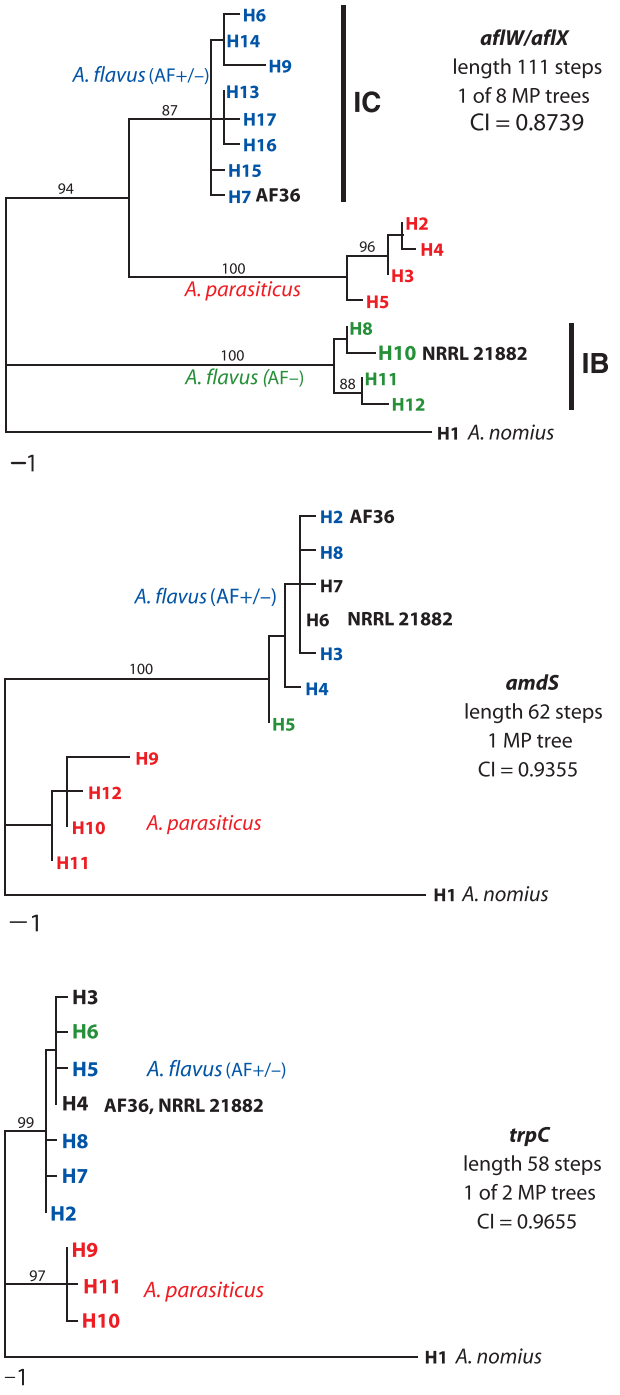

Fig. 4 One most parsimonious (MP) phylogeny inferred for *aflW/aflX* and two noncluster loci (*amdS* and *trpC*). Taxa include *Aspergillus flavus* and *Aspergillus parasiticus* sampled from Georgia, the two biocontrol *A. flavus* strains AF36 and NRRL 21882, the genome strains of *A. flavus* NRRL 3357 and *A. oryzae* NRRL 5590, eight nonaflatoxigenic *A. flavus* strains that represent distinct deletion patterns in the aflatoxin cluster, and *A. nomius* NRRL 13137 (GenBank Accessions AY510454, GQ479323 and GQ479322 for *aflW/aflX*, *trpC* and *amdS* respectively) as the outgroup species. Each phylogeny is labelled with its respective gene name, tree length, number of MP trees and consistency index (CI). The scale bar below each tree represents a single character state change. For each phylogeny, the distribution of isolates among haplotypes and whether isolates are aflatoxin producers (+) or nonproducers (–) are shown in Table S2. The *aflW/aflX* phylogeny shows two distinct *A. flavus* aflatoxin gene clades: one clade (IC, shown in blue) includes aflatoxigenic and nonaflatoxigenic strains (AF+/–), including AF36, and shares a more recent common ancestor with *A. parasiticus* (shown in red); a second clade (IB, shown in green) is completely nonaflatoxigenic (AF–) and includes complete/partial cluster *A. flavus* strains as well as NRRL 21882. With the exception of *A. nomius*, haplotypes shown in black for *amdS* and *trpC* include strains from both *A. flavus* lineages. The AF36 and NRRL 21882 biocontrol strains show divergent origins only in *aflW/aflX*. As expected, the noncluster genes are orthologous and show no evidence of trans-speciation. Bootstrap values are shown for branches with >80% support.

Linkage disequilibrium analysis for homologous regions in full and partial clusters revealed that polymorphisms in deletion strains align with blocks 3–6 (Fig. S3; Table S1), indicating a shared history of recombination in complete and partial clusters; however, block 1, which spans *aflT* to *aflE*, is the largest block and is missing in all partial clusters (Chang *et al.* 2005; Yin *et al.* 2009). Two recent studies examining nonaflatoxigenic *A. flavus* strains isolated from peanut fields in China (Yin *et al.* 2009) and from food and feed commodities in Italy (Criseo *et al.* 2008) report a high frequency of cluster deletions, which could account for more than 60% of the nonaflatoxigenic trait (Criseo *et al.* 2008). All deletion patterns reported so far are consistent with the block structure we report. In all cases, missing genes in partial clusters are flanked by recombination blocks in complete clusters; for example, loss of *aflN* in deletion pattern D in Yin *et al.* (2009) is flanked by *A. flavus* blocks 2 and 3 (Fig. 1). Similar deletion patterns have been reported in *A. oryzae* groups 2 and 3 (Kusumoto *et al.* 2000) and observed in *A. oryzae* IC902 (=RIB 430) (I. Carbone, unpublished data). Collectively, these patterns are consistent with the hypothesis of directed stepwise deletions, which postulates that gene loss begins at the *aflF* end of the aflatoxin cluster and ends in the late pathway segment (Kusumoto *et al.* 2000). Because the position of recombination block boundaries coincides with deletion patterns, it might be possible to predict the existence of other partial aflatoxin clusters by using these boundaries as a guide.

The recombination blocks we identified in *A. flavus* coincide with the sequence breakpoints in deletion variants reported previously (Chang *et al.* 2005). For example, deletion strain IC313 is missing all the early

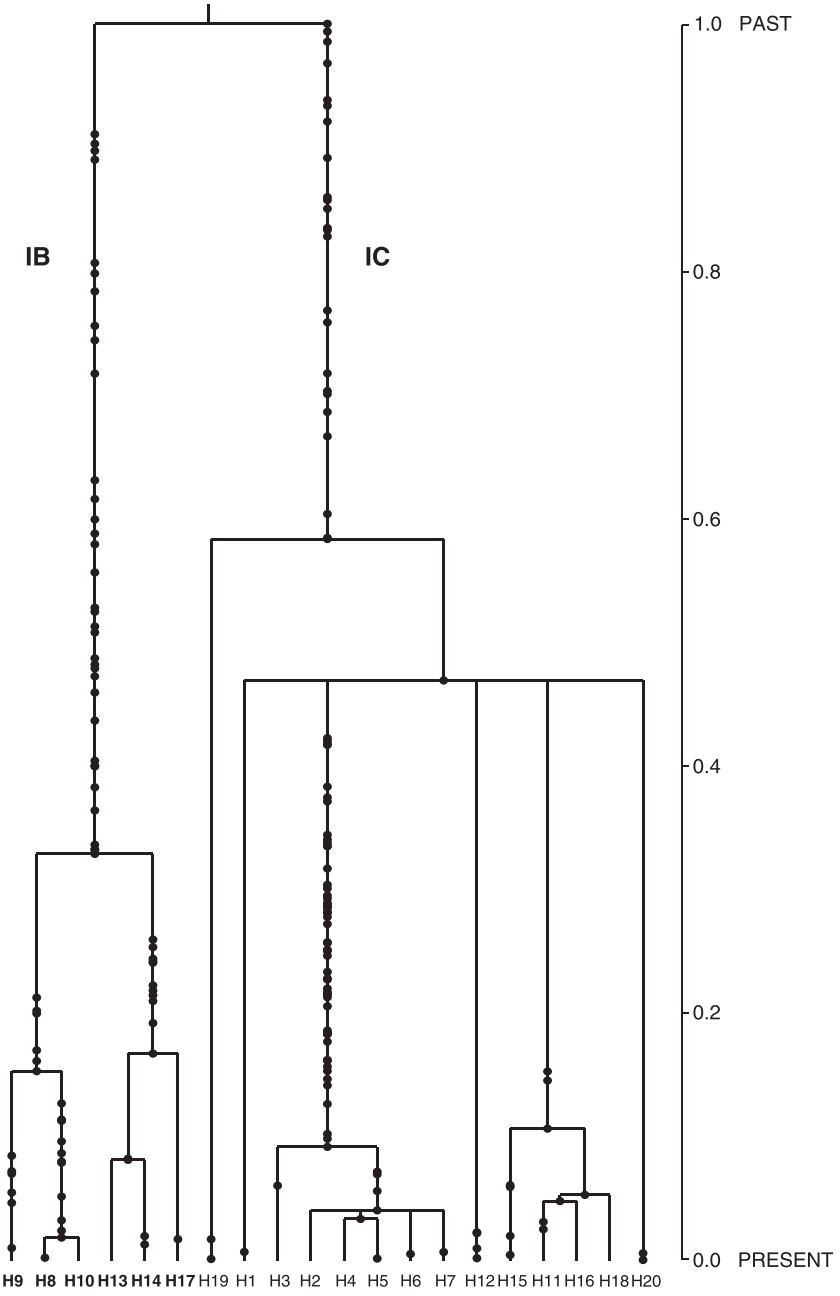

Fig. 5 The rooted coalescent-based gene genealogy for the largest nonrecombining partition in full and partial gene clusters. Mutations are shown as dots along the branches. The estimated mean time to the most recent common ancestor (TMRCA) in coalescent time units is 2.07 (SD = 0.02). The scale at the right shows time rescaled to a TMRCA of 1.0. All partial cluster strains examined are in a distinct nonaflatoxigenic evolutionary clade (IB), which comprises haplotypes shown in bold. The other distinct clade (IC) includes AF+/- strains. The gene genealogy with the highest root probability was selected based on maximum-likelihood estimates of all possible rooted trees using genetree (likelihood of best tree = 1.16×10^{-80}, SD = 3.64×10^{-77}, $\theta = 40.72$). Convergence to the best tree was achieved within six independent runs of 5 million simulations each and one 10 million simulation, assuming panmixia and constant population size. The distribution of strains among haplotypes is shown in Table S3.

pathway genes upstream of *aflM*, consistent with a double-strand break (DSB) in the *aflE/aflM* region (block 1). Similarly, strain IC315 could arise from a DSB in the *aflM/aflN* region (block 2) and IC314 from a DSB in the *aflN/aflG* region (block 3) (Fig. 1). Most DSBs, which presumably give rise to deletion variants, appear to be located in intergenic regions containing transcriptional promoters at the junction of recombination blocks 1, 2 and 3. Although DSBs can occur in genes, as observed in Fig. S2 for blocks 2 (*aflM*), 3 (*aflN*), 4 and 5 (*aflL*), most are located in intergenic promoter-containing regions (Baudat & Nicolas 1997). There are no reports of deletion variants associated with blocks 4, 5 and 6, but it is possible that variants exist as these blocks are in a region of increasing G + C content.

The clear transition in G + C content between the CPA and aflatoxin gene clusters (Fig. 3) suggests a possible recombination hotspot between the two gene clusters. This might explain why gene loss in the aflatoxin gene cluster sometimes involves only aflatoxin cluster genes, as seen in IC311, which makes CPA but not aflatoxins as a result of a deletion of the *aflT-aflM* region

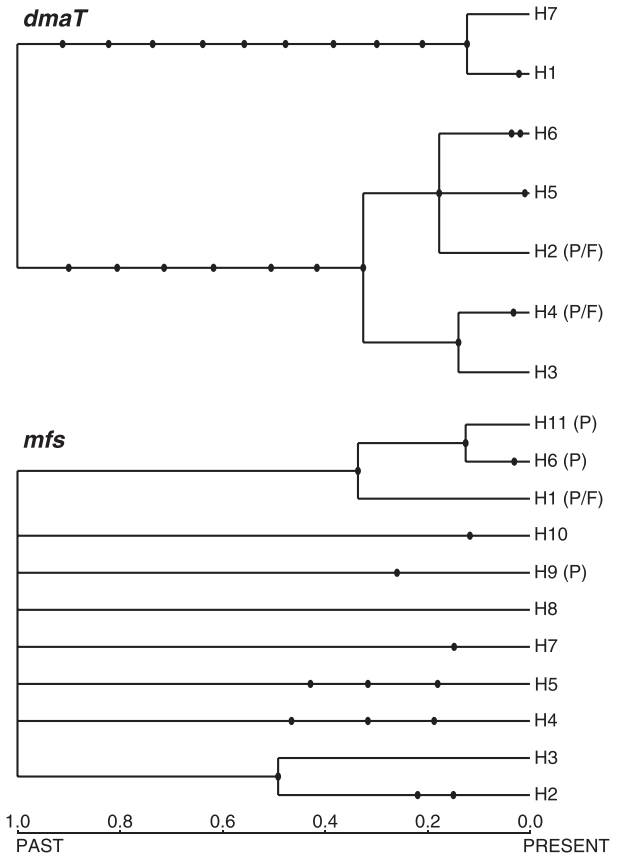

Fig. 6 Coalescent-based gene genealogies for *dmaT* and *mfs*. Mutations are indicated with dots along branches and time is scaled to the most recent common ancestor (TMRCA) of 1.0 for each locus. The mean TMRCA for *dmaT* is 3.32 (SD = 0.69) and for *mfs* is 1.67 (SD = 0.48). The haplotype structure for each genealogy is shown in Table S4. Haplotypes that contain partial (P) or full (F) aflatoxin cluster strains are indicated in the genealogies. The *dmaT* rooted genealogy (likelihood of best tree = 1.0265×10^{-13}, SD = 4.4358×10^{-12}, $\theta = 5.22$) indicates two divergent lineages similar to, but topologically discordant with, the aflatoxin cluster-based genealogy in Fig. 5. In the *mfs* rooted genealogy (likelihood of best tree = 5.9185×10^{-13}, SD = 6.5180×10^{-12}, $\theta = 3.69$), haplotype H8 (IC277) is the most recent common ancestor of the sample, and partial cluster strains in haplotypes H6 and H11 are the most recently evolved. At least three independent runs of 1–10 million simulations each were examined to ensure convergence to the best tree.

(Fig. 1). A recombination hotspot between the two gene clusters might also explain the observed diversity in mycotoxin production in populations, from which it is possible to isolate strains that are AF+/CPA+, AF+/CPA−, AF−/CPA+ and AF−/CPA− (Horn & Dorner 1999; Chang *et al.* 2009).

Although the mechanism that gives rise to deletion variants in *A. flavus* is unclear, crossing over between full and partial clusters during sexual reproduction is one possibility. Alternatively, intra-genomic translocations may move cluster genes to other chromosomes. This may involve translocations of entire telomeric/subtelomeric regions between chromosomes (Smith *et al.* 2007) or translocations within a single gene such as *stcW* (=*aflW*) of the sterigmatocystin gene cluster that disrupts the open reading frame (Hodges *et al.* 2000). In addition, movement of cluster genes may be due to duplication followed by gene loss (Carbone *et al.* 2007b; Powell *et al.* 2008); a partial cluster duplication of genes from *aflR* to *aflM* and *aflO* has been reported for several *A. parasiticus* strains (Chang & Yu 2002). Several late pathway genes, including the *aflW/aflX* region, have been amplified and sequenced in NRRL 21882, and *aflA*, *aflB*, *aflR* and *aflS* genes have also been sequenced in other representative strains within the same VCG 24 (I. Carbone, unpublished data). Although these genes are present, they are not associated with the aflatoxin gene cluster on the telomere end of chromosome 3R. Whether these cluster genes are the result of a translocation to another chromosomal location in a full cluster ancestor or a duplication of the original cluster to another genomic location followed by gene loss needs to be investigated further.

In previous work, we showed that gene modularity arising from duplications of a single gene is a common feature across many genes in aflatoxin biosynthesis, but such modularity is clearly more pronounced in the *aflT-aflE* region of the cluster (Carbone *et al.* 2007b). We now provide evidence that LD blocks and recombination hotspots in *A. flavus* may be responsible for the observed modularity, such that regions in the cluster with low recombination rates and reduced sequence diversity show greater modularity than regions that harbour distinct LD blocks and recombination hotspots. In Fig. 1, the 3′ terminal end of the cluster starting with *aflW*, which flanks the recombination hotspot, shows significant LD and a return to modularity (Carbone *et al.* 2007b). In *A. parasiticus*, block 5 spans *aflQ* to *aflV* (Carbone *et al.* 2007a), which in turn spans with the recombination hotspot in *A. flavus* (Figs 1 and S1). At least two blocks in *A. parasiticus* span two or more cluster genes (blocks 2 and 5); these blocks are potentially larger and more encompassing of the cluster but this cannot be determined without the sequencing of full gene clusters in *A. parasiticus*. Only two small LD regions (blocks 3 and 4) were detected in the span of *aflE* to *aflP* in *A. parasiticus*; the same spanned region in *A. flavus* harbours six distinct LD blocks (Fig. 1). The reduced recombination in *aflE* to *aflP* and *aflQ* to *aflW* simplifies the ARG in *A. parasiticus* (Fig. 2B in Carbone *et al.* 2007a) compared to the ARG in *A. flavus* (Fig. 2). Although recombination rates are ~10-fold lower in *A. parasiticus* than *A. flavus* (0.000394 vs. 0.0011), the

existence of common blocks (*A. flavus* block 1 with *A. parasiticus* block 3, and *A. flavus* block 4 with *A. parasiticus* block 4) indicates shared common ancestry in the history of these species. The different rates of recombination in *A. flavus* and *A. parasiticus* may be the result of a larger effective mating population in *A. flavus* than in *A. parasiticus*. VCG diversity in the single peanut field in Georgia was higher for *A. flavus* than *A. parasiticus*, with 44 and 17 VCGs respectively (Horn & Greene 1995). Because mating in *A. flavus* and *A. parasiticus* is between parents of different VCGs (Horn *et al*. 2009a,b,c) and results in progeny that belong to new VCGs (B.W. Horn, unpublished data), ongoing sexual reproduction would continue to yield disproportionate numbers of VCGs and, as a consequence, different rates of recombination in these fungi.

In *A. parasiticus*, trans-speciation and balancing selection acting on a G_1-dominant lineage was observed in *aflM/aflN* (*hypE*), which is a relatively large nonrecombining region of the *A. parasiticus* cluster. Similarly, in *A. flavus*, the nonrecombining *aflW/aflX* shows a strong signature of balancing selection, which is acting on a nonaflatoxigenic lineage (Fig. S1). *AflW* is a flavin-dependent oxidase and *aflX* is an NAD(P)H-dependent oxidoreductase that may be involved in electron transfer to the flavin (Cary *et al*. 2006). Positive selection for these genes may explain the retention of a basal cluster even in nonaflatoxin-producing strains of *A. flavus*. All six LD blocks in *A. flavus* show two highly divergent haplotypes; however, within each block, there is no specific association of haplotype with either an aflatoxigenic or a nonaflatoxigenic trait. Only the *aflW/aflX* region in our population study, and possibly the genes and intergenic regions from *aflW* to *hxtA* (Fig. S2), show two distinct haplotypes in which one of the two haplotypes is a nonaflatoxigenic strain (IC277 in Fig. S1, and *A. oryzae* NRRL 5590 in Fig. S2). Geiser *et al.* (1998) also showed disproportionately more polymorphisms in *omtA* (=*aflP*) than in four noncluster genes within *A. flavus* group I. Moreover, *aflP*, which is located in the recombination hotspot, has been reported to harbour a greater number of phylogenetically informative sites than other genes in the cluster (Chang *et al.* 2006). Both distinct *A. flavus* lineages that we report in this paper are in group I of Geiser *et al.* (1998). The existence of a strictly nonaflatoxigenic clade (IB) in the *aflW/aflX* region (Fig. 4) suggests that selection is acting to maintain this phenotype in the population. The strains examined with deletion patterns B (IC310), C (IC311), E (IC313), F (IC314) and G (IC315) share the same nonaflatoxigenic clade in *aflW/aflX*. The specific genes under balancing selection for the nonaflatoxigenic phenotype are not obvious when examining other cluster regions, as gene conversion or double crossovers will shuffle highly divergent sequences among aflatoxigenic and nonaflatoxigenic strains. This process results in alternating haplotype patterns among blocks such that one common haplotype in one block transitions to a different one in the adjacent block (see haplotype polymorphism map at the bottom of Fig. S1).

If selection were favouring the nonaflatoxigenic phenotype, then *aflC* (*pksA*) and *aflM* (*ver-1*) would seem to be the ideal candidates to target. The aflatoxin precursor, versicolorin A, has been shown to accumulate in strains that are nonaflatoxigenic but have a functional *aflC*, and both *aflX* and *aflY* are integral in the conversion of versicolorin A to demethylsterigmatocystin (Ehrlich *et al.* 2005; Cary *et al.* 2006) as are *aflM* and *aflN* (Yu *et al.* 2004). Versicolorin A also accumulates in toxigenic strains, leading us to believe that it is the rate-limiting step before the split into B_1 or B_2 production (D. Bhatnagar, personal communication). Although we identified a distinct nonaflatoxigenic clade, the actual target of selection may be ecological adaptiveness rather than the loss of aflatoxin production *per se*. *Aspergillus flavus* strains vary considerably in factors influencing aggressiveness in crop invasion, such as the production of plant-degrading enzymes (Cotty *et al.* 1990; Brown *et al*. 1992; Shieh *et al.* 1997). There appears to be a positive correlation between regions with a high frequency of drought and *A. flavus* soil populations with high aflatoxigenicity (Horn & Dorner 1999; Dorner & Horn 2007). As crops are susceptible to *A. flavus* under drought-stress conditions, this suggests that nonaflatoxigenic *A. flavus* strains may be less aggressive than aflatoxigenic strains in invading crops. Alternatively, nonaflatoxigenic strains might be better adapted to other as yet undefined ecological niches. This warrants further investigation.

Balancing selection in the nonaflatoxigenic lineage was not evident in the 5′ end of the cluster corresponding to the early pathway genes, which suggests that these genes are more prone to loss. Further upstream, a phylogenetic signature of balancing selection was observed in *dmaT* (Fig. 6), which encodes a critical enzyme in CPA biosynthesis (Chang *et al.* 2009); however, the *dmaT* genealogy is incongruent with the aflatoxin cluster genealogy and, as expected, loss of genes in the 5′ end of the aflatoxin cluster does not always result in a concurrent loss of genes in the CPA cluster. With the exception of IC313 and NRRL 21882, all deletion strains examined contain an intact *dmaT* gene as well as several additional genes between the telomere and *dmaT*. It appears that *dmaT* is functional for at least one deletion strain (IC311) and one nonaflatoxigenic strain with a full cluster (IC277).

Comparative genome hybridization of *A. oryzae* NRRL 5590 and *A. flavus* IC277 to the *A. flavus* NRRL

3357 reference genome reveals deletions of ~75 and 45 kb in size respectively, from the telomere ends of these strains (I. Carbone, unpublished data). This indicates that gene loss is also occurring in NRRL 5590 and IC277 full cluster strains, at the distal end of chromosome 3R. Deletions in the telomere of IC277 do not extend to the CPA and aflatoxin clusters, whereas in NRRL 5590, the breakpoint is in the CPA gene cluster. One possible explanation is that gene clusters slow down the deletion process, perhaps through the actions of balancing selection in these clusters. The only exception is NRRL 21882 in which all telomeric genes are missing and the telomeric sequence is immediately adjacent to *hxtA*.

Results from coalescent simulations indicate that aflatoxin cluster deletions in *A. flavus* lineage IB are recent evolutionary events and that deletion strains are descendants of full cluster ancestors (Fig. 5). This has important implications for biological control using nonaflatoxigenic *A. flavus* strains. Although our population study cannot exclude the possibility of a nonaflatoxigenic strain with a full gene cluster reacquiring toxicity genes, there is no evidence that deletion strains can reassemble a full cluster via crossing over or gene conversion in the cluster. A single cross-over between a partial cluster strain and a full cluster strain in the recombination hotspot region would give rise to progeny with full clusters that have the *aflW/aflX* region of partial clusters and vice versa. Alternatively, progeny from either AF36 or NRRL 21882 can become aflatoxigenic by acquiring chromosome 3 from an aflatoxigenic parent, through the independent assortment of chromosomes during meiosis. We are currently investigating both possibilities in laboratory crosses.

Coalescent simulations of nonaflatoxin cluster *dmaT* and *mfs* gene genealogies show that the full cluster nonaflatoxigenic strains occupy interior positions in the genealogies and are therefore older than derived deletion strains (Fig. 6). Haplotypes that include deletion variants in interior positions such as H2 in *dmaT* and H1 in *mfs* also include full cluster strains (Fig. 6), which suggest a lack of resolution. These noncluster gene genealogies further support our observation of lineage-specific gene loss in the *aflW/aflX* region (Fig. 4) and late pathway genes (Fig. 5). The existence of common LD blocks in full and partial cluster strains indicates that the process of crossing over in full clusters may be driving gene loss. Because only one or two mutations separate *A. flavus* partial deletion (IC314 [H10]) and full cluster (IC310 [H8]) strains (Fig. 5), it is possible that sexual recombination may be accelerating gene loss. The close similarity between full and partial cluster strains argues against the hypothesis that gene loss (partial or entire deletion) is due to genetic drift, when the aflatoxin cluster is no longer needed, long after the recombination events have ceased. In contrast, the shared ancestry and geographic proximity among IC310 and IC314 strains indicate that cluster disassembly is recent and may be influenced by local population dynamics or environmental conditions. Future work will examine recombination, gene loss and heritability of toxin production in *A. flavus* offspring from sexual matings.

Acknowledgements

We thank two anonymous reviewers for their insightful comments and suggestions that helped us improve this manuscript. This work is funded by the North Carolina Cooperative State Research, Education and Extension Service, Grant Nos 2006-34500-17032, 2008-34500-19396, and by the National Research Initiative of the USDA Cooperative State Research, Education and Extension Service, Grant No. 2005-35319-16126 to I. C.

References

Aylor DL, Price EW, Carbone I (2006) SNAP: combine and map modules for multilocus population genetic analysis. *Bioinformatics*, **22**, 1399–1401.

Baudat F, Nicolas A (1997) Clustering of meiotic double-strand breaks on yeast chromosome III. *Proceedings of the National Academy of Sciences, USA*, **94**, 5213–5218.

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

Bowden LC, Price EW, Carbone I (2008) *SNAP Clade and Matrix, Version 2. Distributed Over the Internet*. Department of Plant Pathology, North Carolina State University. Available at: http://www.cals.ncsu.edu/plantpath/faculty/carbone/home.html (last accessed October 2009)

Brown RL, Cleveland TE, Cotty PJ, Mellon JE (1992) Spread of *Aspergillus flavus* in cotton bolls, decay of intercarpellary membranes, and production of fungal pectinases. *Phytopathology*, **82**, 462–467.

Burdock GA, Flamm WG (2000) Safety assessment of the mycotoxin cyclopiazonic acid. *International Journal of Toxicology*, **19**, 195–218.

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

Carbone I, Ramirez-Prado JH, Jakobek JL, Horn BW (2007b) Gene duplication, modularity and adaptation in the evolution of the aflatoxin gene cluster. *BMC Evolutionary Biology*, **7**, 111.

Cary JW, Ehrlich KC, Bland JM, Montalbano BG (2006) The aflatoxin biosynthesis cluster gene, *aflX*, encodes an oxidoreductase involved in conversion of versicolorin A to demethylsterigmatocystin. *Applied and Environmental Microbiology*, **72**, 1096–1101.

Chang P-K, Yu J (2002) Characterization of a partial duplication of the aflatoxin gene cluster in *Aspergillus parasiticus* ATCC 56775. *Applied Microbiology and Biotechnology*, **58**, 632–636.

Chang P-K, 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.

Chang P-K, Ehrlich KC, Hua SS (2006) Cladal relatedness among *Aspergillus oryzae* isolates and *Aspergillus flavus* S and L morphotype isolates. *International Journal of Food Microbiology*, **108**, 172–177.

Chang P-K, Horn BW, Dorner JW (2009) Clustered genes involved in cyclopiazonic acid production are next to the aflatoxin biosynthesis gene cluster in *Aspergillus flavus*. *Fungal Genetics and Biology*, **46**, 176–182.

Cole RJ (1986) Etiology of turkey ''X'' disease in retrospect: a case for the involvement of cyclopiazonic acid. *Mycotoxin Research*, **2**, 3–7.

Cole RJ, Sanders TH, Hill RA, Blankenship PD (1985) Mean geocarposphere temperatures that induce preharvest aflatoxin contamination of peanuts under drought stress. *Mycopathologia*, **91**, 41–46.

Cotty PJ (1989) Virulence and cultural characteristics of two *Aspergillus flavus* strains pathogenic on cotton. *Phytopathology*, **79**, 808–814.

Cotty PJ (1994) Influence of field application of an atoxigenic strain of *Aspergillus flavus* on the populations of *A. flavus* infecting cotton bolls and on the aflatoxin content of cottonseed. *Phytopathology*, **84**, 1270–1277.

Cotty PJ, Cleveland TE, Brown RL, Mellon JE (1990) Variation in polygalacturonase production among *Aspergillus flavus* isolates. *Applied and Environmental Microbiology*, **56**, 3885–3887.

Criseo G, Racco C, Romeo O (2008) High genetic variability in non-aflatoxigenic *A. flavus* strains by using Quadruplex PCR-based assay. *International Journal of Food Microbiology*, **125**, 341–343.

Dorner JW (2005) Biological control of aflatoxin crop contamination. In: *Aflatoxin and Food Safet* (ed. Abbas HK), pp. 333–352. Taylor and Francis, Boca Raton, FL.

Dorner JW (2009a) Biological control of aflatoxin contamination in corn using a nontoxigenic strain of *Aspergillus flavus*. *Journal of Food Protection*, **72**, 801–804.

Dorner JW (2009b) Development of biocontrol technology to manage aflatoxin contamination in peanuts. *Peanut Science*, **36**, 60–67.

Dorner JW, Horn BW (2007) Separate and combined applications of nontoxigenic *Aspergillus flavus* and *A. parasiticus* for biocontrol of aflatoxin in peanuts. *Mycopathologia*, **163**, 215–223.

Eaton DL, Groopman JD (1994) *The Toxicology of Aflatoxins: Human Health, Veterinary, and Agricultural Significance*. Academic Press, San Diego.

van Edmond HP, Jonker MA (2005) Worldwide regulations on aflatoxins. In:*Aflatoxin and Food Safet* (ed Abbas HK), pp. 77–93. Taylor and Francis, Boca Raton, FL.

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 B, Boue SM, Bhatnagar D (2005) An aflatoxin biosynthesis cluster gene encodes a novel oxidase required for conversion of versicolorin A to sterigmatocystin. *Applied and Environmental Microbiology*, **71**, 8963–8965.

EPA (2003) *Aspergillus flavus AF36*. 3 July 2003. Biopesticide Registration Action Document. Available at: http://www.epa.gov/pesticides/biopesticides/ingredients/tech_docs/brad_006456.pdf

EPA (2004) *Aspergillus flavus NRRL 21882*. 24 March 2004. Biopesticide Registration Action Document. http://www.epa.gov/oppbppd1/biopesticides/ingredients/tech_docs/brad_006500.pdf.

Fu Y-X (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics*, **147**, 915–925.

Fu Y-X, Li W-H (1993) Statistical tests of neutrality of mutations. *Genetics*, **133**, 693–709.

Geiser DM, Pitt JI, Taylor JW (1998) Cryptic speciation and recombination in the aflatoxin-producing fungus *Aspergillus flavus*. *Proceedings of the National Academy of Sciences, USA*, **95**, 388–393.

Geiser DM, Dorner JW, Horn BW, Taylor JW (2000) The phylogenetics of mycotoxin and sclerotium production in *Aspergillus flavus* and *Aspergillus oryzae*. *Fungal Genetics and Biology*, **31**, 169–179.

Griffiths RC, Tavaré S (1994) Ancestral inference in population genetics. *Statistical Science*, **9**, 307–319.

Hey J, Wakeley J (1997) A coalescent estimator of the population recombination rate. *Genetics*, **145**, 833–846.

Hodges RL, Kelkar HS, Xuei X *et al.* (2000) Characterization of an echinocandin B-producing strain blocked for sterigmatocystin biosynthesis reveals a translocation in the *stcW* gene of the aflatoxin biosynthetic pathway. *Journal of Industrial Microbiology & Biotechnology*, **25**, 333–341.

Horn BW (2005) Ecology and population biology of aflatoxigenic fungi in soil. In: *Aflatoxin and Food Safet* (ed. Abbas HK), pp. 95–116. CRC Press, Boca Raton, FL.

Horn BW (2007) Biodiversity of *Aspergillus* section *Flavi* in the United States: a review. *Food Additives and Contaminants*, **24**, 1088–1101.

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 and Environmental Microbiology*, **65**, 1444–1449.

Horn BW, Greene RL (1995) Vegetative compatibility within populations of *Aspergillus flavus, Aspergillus parasiticus,* and *A. tamarii* from a peanut field. *Mycologia*, **87**, 324–332.

Horn BW, Dorner JW, Greene RL, Blankenship PD, Cole RJ (1994) Effect of *Aspergillus parasiticus* soil inoculum on invasion of peanut seeds. *Mycopathologia*, **125**, 179–191.

Horn BW, Greene RL, Sobolev VS *et al.* (1996) Association of morphology and mycotoxin production with vegetative compatibility groups in *Aspergillus flavus, A. parasiticus,* and *A. tamarii*. *Mycologia*, **88**, 574–587.

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

Horn BW, Ramirez-Prado JH, Carbone I (2009b) Sexual reproduction and recombination in the aflatoxin-producing fungus *Aspergillus parasiticus*. *Fungal Genetics and Biology*, **46**, 169–175.

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

Howlett BJ, Idnurm A, Heitman J (2007) Fungal pathogenesis: gene clusters unveiled as secrets within the *Ustilago maydis* code. *Current Biology*, **17**, R87–R90.

Hudson RR, Kaplan NL (1985) Statistical properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics*, **111**, 147–164.

Kusumoto K, Nogata Y, Ohta H (2000) Directed deletions in the aflatoxin biosynthesis gene homolog cluster of *Aspergillus oryzae*. *Current Genetics*, **37**, 104–111.

Leslie JF (1993) Fungal vegetative compatibility. *Annual Review of Phytopathology*, **31**, 127–150.

Lyngsø RB, Song YS, Hein J (2005) Minimum recombination histories by branch and bound. *Proceedings of the 5th International Workshop on Algorithms in Bioinformatics Lecture Notes in Bioinformatics*, **3692**, 239–250.

Machida M, Asai K, Sano M *et al.* (2005) Genome sequencing and analysis of *Aspergillus oryzae*. *Nature*, **438**, 1157–1161.

Myers SR, Griffiths RC (2003) Bounds on the minimum number of recombination events in a sample history. *Genetics*, **163**, 375–394.

Nei M (1987) *Molecular Evolutionary Genetics*. Columbia University Press, New York.

Patron NJ, Waller RF, Cozijnsen AJ *et al.* (2007) Origin and distribution of epipolythiodioxopiperazine (ETP) gene clusters in filamentous ascomycetes. *BMC Evolutionary Biology*, **7**, 174.

Peterson SW (2008) Phylogenetic analysis of *Aspergillus* species using DNA sequences from four loci. *Mycologia*, **100**, 205–226.

Peterson SW, Ito Y, Horn BW, Goto T (2001) *Aspergillus bombycis*, a new aflatoxigenic species and genetic variation in its sibling species, *A. nomius*. *Mycologia*, **93**, 689–703.

Powell AJ, Conant GC, Brown DE, Carbone I, Dean RA (2008) Altered patterns of gene duplication and differential gene gain and loss in fungal pathogens. *BMC Genomics*, **9**, 147.

Price EW, Carbone I (2005) SNAP: workbench management tool for evolutionary population genetic analysis. *Bioinformatics*, **21**, 402–404.

Ramirez-Prado JH, Moore GG, Horn BW, Carbone I (2008) Characterization and population analysis of the mating-type genes in *Aspergillus flavus* and *Aspergillus parasiticus*. *Fungal Genetics and Biology*, **45**, 1292–1299.

Shieh MT, Brown RL, Whitehead MP *et al.* (1997) Molecular genetic evidence for the involvement of a specific polygalacturonase, P2c, in the invasion and spread of *Aspergillus flavus* in cotton bolls. *Applied and Environmental Microbiology*, **63**, 3548–3552.

Smith CA, Woloshuk CP, Robertson D, Payne GA (2007) Silencing of the aflatoxin gene cluster in a diploid strain of *Aspergillus flavus* is suppressed by ectopic *aflR* expression. *Genetics*, **176**, 2077–2086.

Swofford DL (1998) *PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4.0*. Sinauer Associates, Sunderland, MA.

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **123**, 585–595.

Watterson GA (1975) On the number of segregating sites in genetic models without recombination. *Theoretical Population Biology*, **7**, 256–276.

Yin Y, Lou T, Yan L, Michailides TJ, Ma Z (2009) Molecular characterization of toxigenic and atoxigenic *Aspergillus flavus* isolates, collected from peanut fields in China. *Journal of Applied Microbiology*; doi: 10.1111/j.1365-2672.2009.04356.x.

Yu J, Chang P-K, Ehrlich KC *et al.* (2004) Clustered pathway genes in aflatoxin biosynthesis. *Applied and Environmental Microbiology*, **70**, 1253–1262.

Zhang R, Zhang CT (2004) A systematic method to identify genomic islands and its applications in analyzing the genomes of *Corynebacterium glutamicum* and *Vibrio vulnificus* CMCP6 chromosome I. *Bioinformatics*, **20**, 612–622.

Geromy Moore is a PhD student interested in the evolution of the aflatoxin gene cluster. His work is elucidating the global population structure and mycotoxin diversity of species in *Aspergillus* section *Flavi*. Rakhi Singh is a researcher in the Carbone laboratory and is interested in the molecular biology of aflatoxigenic fungi. Bruce Horn is a Research Microbiologist interested in the ecology and population biology of aflatoxigenic fungi. Ignazio Carbone's laboratory focuses on the population genomics of agriculturally important *Aspergillus* species.

Supporting information

Additional Supporting Information may be found in the online version of this article.

Table S1 Haplotypes and strain designations for Fig. S3

Table S2 Haplotypes and strain designations for *Aspergillus flavus* and *Aspergillus parasiticus* phylogenies in Figs 4 and S4

Table S3 Haplotypes and strain designations for multilocus gene genealogy in Fig. 5

Table S4 Haplotypes and strain designations for gene genealogies in Fig. 6

Fig. S1 Linkage disequilibrium plot for 453 polymorphisms in 21 regions of the aflatoxin gene cluster in *Aspergillus flavus*.

Fig. S2 Linkage disequilibrium plot for 2188 polymorphisms across four full aflatoxin cluster sequences: *Aspergillus flavus* AF13, AF36, NRRL 3357, and *Aspergillus oryzae* NRRL 5590.

Fig. S3 Linkage disequilibrium plot based on 228 polymorphisms spanning intergenic regions from *aflN* to *aflP* for the 43 Georgia isolates (excluding IC252, which is missing the entire cluster) plus three partial deletion strains (IC311, 313, and 315) and three nonaflatoxigenic full cluster strains (IC309, IC310, and IC312).

Fig. S4 The corresponding strict consensus phylogenies (assuming no recombination) and minimal ARGs (assuming recombination) inferred for the phylogenies in Fig. 4.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.