Cryptic speciation and recombination in the aflatoxin-producing fungus *Aspergillus flavus* (population genetics/mycotoxin/biological species/microbial evolution)

DAVID M. GEISER†, JOHN I. PITT‡, AND JOHN W. TAYLOR†

†Department of Plant and Microbial Biology, University of California, Berkeley, CA 94720; ‡Commonwealth Scientific and Industrial Research Organization, Division of Food Science and Technology, P.O. Box 52, North Ryde, NSW 2113, Australia

Communicated by Robert L. Metzenberg, Stanford University, Stanford, CA, November 11, 1997 (received for review July 21, 1997)

ABSTRACT *Aspergillus flavus*, like approximately one-third of ascomycete fungi, is thought to be cosmopolitan and clonal because it has uniform asexual morphology. *A. flavus* produces aflatoxin on nuts, grains, and cotton, and assumptions about its life history are being used to develop strategies for its biological control. We tested the assumptions of clonality and conspecificity in a sample of 31 Australian isolates by assaying restriction site polymorphisms from 11 protein encoding genes and DNA sequences from five of those genes. *A. flavus* isolates fell into two reproductively isolated clades (groups I and II). The lack of concordance among gene genealogies among isolates in one of the clades (group I) was consistent with a history of recombination. Our analysis included five strains of the closely related industrial fungus *A. oryzae*, all of which proved to be clonally related to group I.

We applied the tools of evolutionary biology to analyze three closely related fungal species of the genus *Aspergillus* that are important in food safety and production, addressing questions about reproductive mode and phylogenetic relationships that relate to biological control and regulation in the food industry. *Aspergillus* flavus and *A. parasiticus* produce aflatoxins, among the most carcinogenic compounds known (1), which are a serious problem worldwide in agricultural commodities such as maize, peanuts, tree nuts, and cotton seed. On the other hand, *A. oryzae* is a nontoxicogen species that has been used for over 4,000 years to produce Asian foods and beverages such as soy sauce and sake, as well as for the production of industrial enzymes. *A. flavus* and *A. oryzae* are distinguished only by subtle morphological differences (2) and appear nearly identical by DNA/DNA hybridization (3), but their identity as the same species or sister species is still controversial, partly because of the regulatory confusion that conspecificity would generate.

In many countries, strict controls are placed on aflatoxin levels in food and feed products, but the cost of lost crops and animals, reduced yield, and monitoring aflatoxin levels is staggering (4). In countries without strict controls, aflatoxin certainly remains a significant public health problem (5). Our ability to reduce or eliminate aflatoxin contamination via biological control methods requires an understanding of the life history of the fungus, which is currently rudimentary. In the laboratory, *A. flavus* is known to reproduce exclusively asexually, with some potential for nonsexual recombination. Biological control methods currently being tested, including seeding fields with large quantities of natural nontoxigenic strains (9), either assume that the fungus is clonal or fail to consider the effects of potential outcrossing. The introduced strains necessarily must be effective competitors with respect to the native strains. In addition, aflatoxin production is a complex secondary metabolite phenotype that involves a large number of genes. Most of the specific enzymatic activities required for aflatoxin production are encoded in a single gene cluster (10), but additional unlinked loci also are required (11). Because aflatoxin is a polygenic, variable trait, outcrossing between the introduced strains and native strains could produce competitive progeny with novel aflatoxin phenotypes, which could then spread to new locations.

MATERIALS AND METHODS

Strain Isolation. Isolates were taken from agricultural soils near Kingaroy, Qld, Australia, a major peanut growing region. Fourteen fields were sampled, seven adjacent to each other =3 km southeast of Kingaroy (fields represented by strain prefixes 1–7) and seven randomly selected within =20 km to the south of Kingaroy (prefixes 11–17). Soils were dilution-plated on a medium designed to select for *A. flavus* growth and incubated at 30°C for 3–5 days. Colonies producing a characteristic orange/yellow reverse were isolated and identified by standard morphological techniques (2). *A. parasiticus* isolates CA1–05 and CA3–01 were collected from tree nut orchards near Winters, CA. *A. oryzae* isolates were provided by Steven Peterson of the United States Department of Agriculture Northern Region Research Laboratory, Peoria, Ill.

Identification and Analysis of Polymorphisms. Oligonucleotide primers were designed to PCR-amplify regions of 11 different protein-encoding loci (Table 1). The primer sequences used were as follows, 5′–3′: amdS1: ccatcgtatatggaactctg, amdS2: aagttgccacaggtatc, pecA1: atctcagacactttctc, pecA2: acctcataacgcttctag, omt1: ggaattcagaggtatc, omt2: aagttgctaagttacca, niaD1: tctgtaagcgaagtg, niaD2: ggaattcagaggtatc, benA5: ctctctcgcgcagact, benA6: ggaattcagaggtatc, pgkA1: cttgaagtacagcttctg, pgkA2: tcaagcgaagttc, faeB: gaaagacgctgatatc, faeB4: atctcgtatatggaactctg, glaA1: caagctgagactaggt, glaA2: gtcgtgactaggt, trpC1: cagcggagtagatgg, trpC3: cagcggagtagatgg, cmdA7: gcctccatctccctctag, cmdA8: atctcgtcagaggtatc, gsdA1: ccaatcagacagcttctg, gsdA2: cttgaagtacagcttctg, gsdA3: ccaatcagacagcttctg. PCRs were performed by using 1 μl of diluted genomic DNA template in 50–μl reactions. Tag DNA polymerase (0.5 unit), 0.2 mM deoxynucleotide triphosphates, and 0.2 mM of each primer were used, with the following temperature profile: 2

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. §1734 solely to indicate this fact.

© 1998 by The National Academy of Sciences

Data deposition: The sequences reported in this paper have been deposited in the GenBank database [accession nos. AF036768-787 (amdS12), AF036788-807 (benA5), AF036808-830 (omt12), AF036831-850 (amdS12), and AF036851-870 (trpC13)].

To whom reprint requests should be addressed. e-mail: dgeiser@nature.berkeley.edu.
The first 31 strains are Australian A. flavus isolates (hyphenated); the last five are A. oryzae isolates.
genotype "Q," differing from genotype "C" only at the pecA12 locus.

We tested this result by sequencing five of the loci (amdS12, benA56, omt12, pecA12, and trpC13) in 17 of the A. flavus isolates, including one isolate from each of the 16 previously identified genotypes and a second isolate from genotype B. In addition, two strains of the related aflatoxin producer A. parasiticus were sequenced to be used as an outgroup, as well as one isolate of A. oryzae. The most variable locus, omt12, was sequenced from an additional three A. oryzae isolates with a range of morphologies. A total of 2,195 nucleotides was sequenced in each isolate, 144 of which were polymorphic in A. flavus, 47 of which were unique, and 97 of which were parsimony-informative. Each locus produced from 1 to 30 MP trees, with consistency indices ranging from \(~0.90\) to 1, indicating that there was little or no homoplasy within loci and no evidence for intragenic recombination. The MP trees for each of the five loci showed a topology consistent with the existence of the two groups inferred from the restriction sites and consistent with the two groups being reproductively isolated (Fig. 2). Of the 97 polymorphic and informative sites within A. flavus, only two were polymorphisms shared between the two groups (Table 3). Of the remaining 95 sites, 37 were fixed in group I, 23 were fixed in group II, and 35 were fixed for differences in both groups. Furthermore, the A. oryzae isolate grouped consistently with isolates from group I, and all four A. oryzae isolates had identical sequences for the most variable locus, omt12.

Tests for reproductive mode can be confounded if individuals are sampled from different biological species, but with the knowledge that A. flavus isolates fell into two groups, we could limit the test to the group for which we had a larger sample, group I. To address this question, we asked whether the gene genealogies for the five different loci were significantly different from each other (13, 14). Under a model of clonality, the topologies of the five gene trees should be congruent, as should those of trees made from any combination of variable sites from the five genes, because all genes and all sites are inherited.

Fig. 1. (A) Strict consensus of the 72 MP trees inferred from the restriction site data for the 17 different multilocus genotypes (A–Q). The arrow indicates the internal branch separating groups I and II. (B) UPGMA (10) distance analysis of the same data. The arrow indicates the same branch highlighted in A, which is the longest branch on the tree (branch lengths shown are proportional to those inferred from the data).* Genotype "C" represents A. flavus isolates 1–10, 5–2, and 17–4, as well as A. oryzae isolates NRRL 447, 448, 449, and 469. "Q" represents that of A. oryzae isolate NRRL 483.

Fig. 2. One MP tree from each of the five gene regions sequenced. CI, consistency index; RI, retention index; RC, rescaled consistency index. Numbers below branches represent bootstrap values based on 500 replicates. Only values \(~70\%\) are shown. The 16 different A. flavus multilocus genotypes are represented by letters. Isolates used representing each multilocus genotype: A, 1–22; B, 1–9 (B1) and 5–1 (B2); C, 17–4; D, 1–26; E, 1–29; F, 3–2; G, 4–2; H, 12–4; I, 7–2; J, 7–3; K, 7–4; L, 11–1; M, 12–3; N, 13–4; O, 14–1; and P, 14–2. GenBank accession numbers are AF036768-787 (amdS12); AF036788-807 (benA56); AF036808-830 (omt12); AF036831-850 (pecA12); and AF036851-870 (trpC13).
together. Under a model of recombination, genealogies of unlinked genes would be expected to conflict, and although sites in the same gene might all support the same genealogy, sites assembled from different genes should conflict because of recombination over larger physical distances and among different linkage groups. This conflict can be recognized by increased homoplasy and thus by increased phylogenetic tree lengths.

First, MP trees were constructed for each of the five loci, looking at group I isolates alone. Removal of group II reduced the number of informative sites from 97 to 37. The MP tree length for each locus is given in Table 4, along with the minimum number of steps possible. Two of the loci (amdS12 and benA56) produced trees of minimum length whereas omt12, pecA12, and trpC13 showed some homoplasy: they produced trees of 1, 3, and 1 excess steps, respectively, for a total of 5 extra steps. If the gene trees were perfectly concordant, the MP tree length for all five data sets combined should be the same as the minimum possible tree length: the number of informative sites (37 steps) plus the extra steps due to intralocus homoplasy (5 steps), or 42 steps. The actual MP tree length for the combined data was 57 steps, an excess of 15 steps (making both groups identical) or for another allele (making fixation occurs in the other group, either for the same allele could have arisen after the divergence of the two groups. Then, formative polymorphisms reflect this state, many of which polymorphisms could easily have arisen by independent mutation (21). In this A. flavus population, 144 variable sites were identified, 97 of which were phylogenetically informative. Only 2/97 informative sites were polymorphic in both groups, 60 were fixed in one group and not the other, and 35 were fixed for differences in both groups. Under the most common model of gradual biological speciation, early in the speciation process a reproductive barrier arises within a species. At first, the two isolated groups share polymorphisms inherited from the ancestral species (22). In A. flavus, only 2/144 polymorphisms are consistent with this early degree of divergence between groups I and II, and these polymorphisms could easily have arisen by independent mutation after their isolation. Eventually, one of the two groups becomes fixed for a particular character state, and the other group remains polymorphic. Sixty of 97 phylogenetically informative polymorphisms reflect this state, many of which could have arisen after the divergence of the two groups. Then, fixation occurs in the other group, either for the same allele (making both groups identical) or for another allele (making

Species concepts in fungi most frequently are based solely on morphology. Biological species concepts, which define species as groups of actually or potentially interbreeding individuals (17), have been applied in several cases in sexual fungi through mating tests. In each case, species barriers were discovered that were not first apparent based on morphology (18–20). However, if sex is lacking or covert, it is impossible to define biological species based on mating criteria. Genetic data provide a means for indirectly identifying reproductive isolation and can be applied to define biological species in organisms that are only cryptically outcrossing (13, 14) or clonal. The reproductive barriers that accompany biological speciation result in the conversion of genetic variability within a species to between-species differences (21). In this A. flavus population, 144 variable sites were identified, 97 of which were phylogenetically informative. Only 2/97 informative sites were polymorphic in both groups, 60 were fixed in one group and not the other, and 35 were fixed for differences in both groups. Under the most common model of gradual biological speciation, early in the speciation process a reproductive barrier arises within a species. At first, the two isolated groups share polymorphisms inherited from the ancestral species (22). In A. flavus, only 2/144 polymorphisms are consistent with this early degree of divergence between groups I and II, and these polymorphisms could easily have arisen by independent mutation after their isolation. Eventually, one of the two groups becomes fixed for a particular character state, and the other group remains polymorphic. Sixty of 97 phylogenetically informative polymorphisms reflect this state, many of which could have arisen after the divergence of the two groups. Then, fixation occurs in the other group, either for the same allele (making both groups identical) or for another allele (making

![Figure 3. Partition homogeneity test results.](image-url)
fixed differences between the two groups). Thirty-five of 97 informative polymorphisms reflect this, a number that does not reflect polymorphisms that may have existed but were fixed for the same character state in both groups. As would be expected in an advanced state of biological speciation, the very low level of shared polymorphism between groups I and II is indicative of a long history of reproductive isolation.

Although there was strong evidence for divergence between groups I and II, the A. oryzae isolates analyzed appeared to be members of group I, and almost no variation was observed among A. oryzae isolates. We also added two additional sequences available in the GenBank database from two different A. oryzae strains, an amdS12 sequence from strain RIB40 and a pecA12 sequence from strain KBN616 (23, 24). Phylogenetic analysis showed that these sequences were very similar to those of strain 469 (one nucleotide difference for both) and that they were included in group I (data not shown). Our results are consistent with DNA/DNA hybridization studies that showed 100% complementarity between an A. flavus isolate and an A. oryzae isolate, suggesting that A. oryzae is a domesticated form of A. flavus (3). Clearly, there is little molecular genetic basis for separating the two species although there are morphological differences, and A. oryzae isolates do not make aflatoxin. Based on its pecA12 and om12 sequences, the A. flavus isolate used most commonly in molecular genetic research (strain 70) also appears to belong in group I. This indicates that research on both organisms should be complementary because of the very high degree of genetic similarity between A. oryzae and A. flavus group I.

CONCLUSIONS

No fungus, including those with no known means of sexual reproduction, has passed strict population genetic tests for clonality, although isoenzyme studies of some rust populations argue against recombination (25). Clearly, there is a strong clonal component to A. flavus’ life history, evident from the simple fact that the same clone can be isolated in different places. However, the lack of concordance observed among different gene genealogies in group I can be reasonably explained only by recombination between those loci. The human pathogenic fungus Coccioidoides immitis, which also lacks a known sexual state, was found to have a recombining population structure (26) and to comprise two cryptic species (14). Candida albicans, another presumably asexual pathogen, was at first thought to be exclusively clonal (27), but a more comprehensive study showed evidence for recombination (28). Emericella nidulans, an Aspergillus-related species with a capability for sexual reproduction, also has a strongly clonal component to its life history superimposed on outcrossing (29).

Our results indicate a history of recombination in A. flavus, but we cannot say how or how often A. flavus recombines. Like many apparently asexual fungi, A. flavus has a close relative that is capable of sexual reproduction, Petromyces alliaceus (30–32). In culture, A. flavus produces hardened asexual structures called “sclerotia” that look very much like the sexual structures produced by P. alliaceus and may be homologous to them (33, 34). Perhaps under some conditions in nature the sclerotia of A. flavus harbor sexual reproduction as the homologous structures do in P. alliaceus. Nonmeiotic outcrossing (parasexuality) can be performed in A. flavus in the laboratory (6) but only between strains that are genetically very similar, and it has not been demonstrated in nature in any fungus. Furthermore, we cannot say how frequently A. flavus recombines, except that only a few individuals need to be produced sexually each generation to produce enough recombinant diversity to pass these kinds of tests (26).

The practice and utility of biological vs. phylogenetic species concepts has been debated heavily (see, e.g., ref. 35). Biological species concepts consider speciation to be inherently a process of reproductive isolation. The phylogenetic species concepts, in general, emphasize phylogenetic relationships in defining species as the smallest diagnosable taxonomic units with a clear pattern of parental ancestry (36). Thus, speciation is viewed as a process of cladogenesis and diversification. Proponents of biological species concepts argue that phylogenetic species concepts fail to consider reproductive isolation, an essential force in speciation (37). Proponents of phylogenetic species concepts argue that biological species concepts confuse information about the past with speculation about the future because they use reproductive isolation as the ranking and grouping criterion (36). We agree with Avise and Wollenberg (38), who explain that the two views are actually quite complementary, because phylogenetic and reproductive cohesiveness are strongly inter-related. The historical transmission routes of genes in sexual species are affected by reproductive barriers. Thus, comparisons of the transmission routes of different genes can be used to make inferences about reproductive isolation because of the differences in transmission predicted among interbreeding individuals (members of the same biological species) vs. reproductively isolated individuals (members of different biological species). The power of this integrated approach is particularly evident when applied to bacteria, fungi, and other organisms in which reproductive isolation cannot be ruled out by direct observation (13, 14).

It is clear from the level of fixation of polymorphisms among different loci that the two groups of A. flavus isolates have had a long history of reproductive isolation. Thus, they are best considered different biological species. The fact that the two groups appear sympatrically could easily reflect secondary contact because of the effects of agriculture. Although they are clearly divergent, we cannot say whether groups I and II retain the potential to interbreed because A. flavus isolates cannot be mated. If they do retain the potential, then their current apparent sympatry may lead to re-homogenization. Our sample of A. flavus individuals in this study is small and local and lacks representatives of morphologically distinguishable groups such as the North American large and small sclerotial forms (7). Investigation of a larger, world-wide sample of A. flavus may unravel a complex picture of different levels of allopatric speciation and secondary contact that is only hinted at by our data.

We conclude that A. oryzae is a species that evolved by domestication from A. flavus group I. Genetically, as assessed both by multilocus gene sequences, the A. oryzae isolates are more similar to each other and to some A. flavus group I isolates than they are to other group I isolates. Phenotypically, they can be distinguished from A. flavus groups I and II based on continuous characters such as more olivaceous colony color and more floccose colony texture. In addition, A. oryzae isolates do not make aflatoxin. We conclude that A. oryzae is part of a monophyletic A. flavus clade, and its phenotypic distinctiveness is the result of strong selection associated with domestication. Given that 10/11 restriction sites that were polymorphic in A. flavus were found monomorphic in A. oryzae and that almost no sequence polymorphism was found among A. oryzae gene sequences, it is impossible to apply a biological species concept to A. oryzae vs. A. flavus group I, a concept that is probably moot for a domesticated species in any event. Although we found no evolutionary justification for maintaining A. oryzae as a separate species from A. flavus, there is a regulatory justification because A. flavus isolates are known to produce important mycotoxins and A. oryzae isolates are not. Our results indicate that industrial microbiologists should continue to be vigilant in their scrutiny of the mycotoxigenic potential of A. oryzae isolates.

These results have important implications for attempts to understand and control aflatoxin contamination in crops.
First, *A. flavus* comprises at least two reproductively isolated taxa, so these efforts have multiple targets that may have distinctive ecological niches. Second, at least one of these two groups has a history of recombination, which may indicate that recombination has an important role in the diversification of aflatoxin phenotypes. Further phylogenetic research is warranted to uncover other potential cryptic species, as well as further characterization of group II. Third, recombination in group I has important implications for biological control strategies, particularly adding nontoxigenic strains to fields where crops are grown, because of the potential for outcrossing between the introduced and native strains. With this knowledge, it would be prudent to characterize the quantitative genetics of aflatoxin synthesis in natural populations of *A. flavus*.

We thank F. M. Harbinski for his assistance in the laboratory, V. Koufopanou and A. Burt for their useful insights, D. Swofford for permission to use PAUP*, S. Peterson for providing *A. oryzae* isolates, and A. Andrianopoulos for providing unpublished sequence data. This work was supported by grants from the National Science Foundation and the National Institutes of Health.