

# UC Berkeley

## UC Berkeley Previously Published Works

### Title

Microevolution in the pansecondary metabolome of *Aspergillus flavus* and its potential macroevolutionary implications for filamentous fungi

### Permalink

<https://escholarship.org/uc/item/0cb915tn>

### Journal

Proceedings of the National Academy of Sciences of the United States of America, 118(21)

### ISSN

0027-8424

### Authors

Drott, Milton T  
Rush, Tomás A  
Satterlee, Tatum R  
et al.

### Publication Date

2021-05-25

### DOI

10.1073/pnas.2021683118

Peer reviewed



# Microevolution in the pansecondary metabolome of *Aspergillus flavus* and its potential macroevolutionary implications for filamentous fungi

Milton T. Drott<sup>a,1</sup>, Tomás A. Rush<sup>b</sup>, Tatum R. Satterlee<sup>c</sup>, Richard J. Giannone<sup>b</sup>, Paul E. Abraham<sup>b</sup>, Claudio Greco<sup>d</sup>, Nandhitha Venkatesh<sup>e,f</sup>, Jeffrey M. Skerker<sup>g</sup>, N. Louise Glass<sup>g,h,i</sup>, Jesse L. Labbé<sup>b</sup>, Michael G. Milgroom<sup>j</sup>, and Nancy P. Keller<sup>a,e,1</sup>

<sup>a</sup>Department of Bacteriology, University of Wisconsin–Madison, Madison, WI 53703; <sup>b</sup>Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37830; <sup>c</sup>Department of Genetics, University of Wisconsin–Madison, Madison, WI 53703; <sup>d</sup>Department of Molecular Microbiology, John Innes Centre, Norwich, NR4 7UH, United Kingdom; <sup>e</sup>Department of Medical Microbiology and Immunology, University of Wisconsin–Madison, Madison, WI 53703; <sup>f</sup>Department of Plant Pathology, University of Wisconsin–Madison, Madison, WI 53706; <sup>g</sup>Innovative Genomics Institute, University of California, Berkeley, CA 94720; <sup>h</sup>Department of Plant and Microbial Biology, University of California, Berkeley, CA 94720; <sup>i</sup>Environmental Genomics and Systems Biology, The Lawrence Berkeley National Laboratory, Berkeley, CA 94720; and <sup>j</sup>School of Integrative Plant Science, Cornell University, Ithaca, NY 14853

Edited by Edward F. DeLong, University of Hawaii at Manoa, Honolulu, HI, and approved March 18, 2021 (received for review October 26, 2020)

Fungi produce a wealth of pharmacologically bioactive secondary metabolites (SMs) from biosynthetic gene clusters (BGCs). It is common practice for drug discovery efforts to treat species' secondary metabolomes as being well represented by a single or a small number of representative genomes. However, this approach misses the possibility that intraspecific population dynamics, such as adaptation to environmental conditions or local microbiomes, may harbor novel BGCs that contribute to the overall niche breadth of species. Using 94 isolates of *Aspergillus flavus*, a cosmopolitan model fungus, sampled from seven states in the United States, we dereplicate 7,821 BGCs into 92 unique BGCs. We find that more than 25% of pangenomic BGCs show population-specific patterns of presence/absence or protein divergence. Population-specific BGCs make up most of the accessory-genome BGCs, suggesting that different ecological forces that maintain accessory genomes may be partially mediated by population-specific differences in secondary metabolism. We use ultra-high-performance high-resolution mass spectrometry to confirm that these genetic differences in BGCs also result in chemotypic differences in SM production in different populations, which could mediate ecological interactions and be acted on by selection. Thus, our results suggest a paradigm shift that previously unrealized population-level reservoirs of SM diversity may be of significant evolutionary, ecological, and pharmacological importance. Last, we find that several population-specific BGCs from *A. flavus* are present in *Aspergillus parasiticus* and *Aspergillus minisclerotigenes* and discuss how the microevolutionary patterns we uncover inform macroevolutionary inferences and help to align fungal secondary metabolism with existing evolutionary theory.

secondary metabolism | population genomics | allopatric speciation | eukaryotic pangenome | comparative genomics

Understanding how ecologically important traits are organized among locally adapted, genetically distinct populations allows for ecological and evolutionary inferences that cannot be approached when treating species as monomorphic groups. In fungi, biosynthetic gene clusters (BGCs) that produce secondary metabolites (SMs) can define niches (1), provide selective advantages in specific ecological conditions (2–4), and affect the host range of pathogens (5, 6). BGCs are characterized by the presence of a gene encoding a core-biosynthetic enzyme (often called a backbone gene) that creates a core chemical structure and by the presence of additional genes encoding enzymes, transporters, and transcription factors that together determine the structure and localization of resulting SMs. In contrast to primary metabolites, SMs are often assumed to be retained only when specific ecological interactions select for them. Given the ecological and economic importance of SMs (e.g., penicillin and lovastatin), their distribution in fungal species has

been the subject of much research. While it has been widely held that species differ in BGC content (7), relatively few genomic studies have examined intraspecific variation. Large-scale genomic studies of intraspecific diversity have focused on strain-level differences in medically important species that are assumed to be panmictic (i.e., without population structure) (8, 9). However, many phylogenetically and ecologically diverse fungi are known to comprise genetically distinct populations (i.e., have population structure) (10–13), with recent evidence that even some species previously considered panmictic have population structure (14). Studies assuming panmixis have drawn macroevolutionary conclusions treating BGC diversity as reflecting a single ecology. While these studies are of clear value to our understanding of the evolution of secondary metabolism in fungi, alternative hypotheses that are based on macroevolutionary ideas rooted in the importance of ecological differentiation at the population level are not well represented in the current literature (8, 15). With this basis, a growing number of prominent comparative genomics

## Significance

Secondary metabolites (SMs) produced by fungi mediate ecological interactions, define fungal niches, and are of profound pharmacological importance to humans. Most work on SMs has focused on a small number of individuals from each species, not fully reflecting the importance of intraspecific diversity. We demonstrate that even in one of the best-studied model fungi, the carcinogen-producing *Aspergillus flavus*, more than 25% of SM-producing biosynthetic gene clusters (BGCs) are novel and/or show population-specific variants. These results support the finding that the organization of BGC diversity into population-specific patterns may sometimes result from ecologically important interactions and may inform evolutionary and etiological inferences of SM capacities within a species. Importantly, our work also presents a vision of sources of potential pharmaceuticals.

Author contributions: M.T.D. designed research; M.T.D., T.A.R., T.R.S., R.J.G., P.E.A., C.G., and N.V. performed research; M.T.D., J.L.L., M.G.M., and N.P.K. contributed new reagents/analytic tools; M.T.D., T.A.R., T.R.S., R.J.G., P.E.A., and C.G. analyzed data; and M.T.D., T.A.R., T.R.S., R.J.G., P.E.A., C.G., N.V., J.M.S., N.L.G., J.L.L., M.G.M., and N.P.K. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

Published under the PNAS license.

<sup>1</sup>To whom correspondence may be addressed. Email: mdrott@wisc.edu or npkeller@wisc.edu.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2021683118/-DCSupplemental>.

Published May 20, 2021.

papers have drawn ecological and evolutionary inferences between species using a small number of (or often single) genomes to represent species. Such inferences reflect a narrow view of species ecology and may miss the overall niche breadth or even the most common ecology of species that may sometimes manifest in different populations.

Population structure allows BGCs to evolve not only within specific environmental conditions encountered by a distinct population but also within the context of other genes (including other BGCs) that may be rare in a species as a whole. There is considerable evidence that different combinations of BGCs present in the same genome provide important genomic context in fungi. For example, the synthesis of several SMs is known to require two physically distinct BGCs (16, 17). Some SMs have ecologically synergistic effects with other SMs (18) or associations with other ecologically important traits (19); such SMs may be of limited ecological value when not found in the same genome, potentially resulting in fitness costs (i.e., recombination load). The evolutionary and ecological significance of such genetic interactions is particularly evident in populations that occur at the edge of species ranges (i.e., peripheral populations). Such populations often exhibit decreased recombination rates, as mixing of gene combinations that are selected for at the edge of a species' range may result in maladapted offspring when combined with genes that are selected for in the general population (i.e., recombination load) (20, 21). Importantly, most genes are highly conserved within species and are thus of little significance for population-level differentiation. When treating species as panmictic, ecologically important coevolved combinations of genes that are at high frequency within small locally adapted peripheral populations are likely to be interpreted as rare anomalies that explain little of the overall species' ecology. A better understanding of how SM diversity is distributed between populations will present opportunities to study ecological and evolutionary processes that are founded upon population-specific differentiation of some ecologically important genes.

Few SMs have received as much attention as aflatoxin, a mycotoxin produced by the fungus *Aspergillus flavus* and several closely related species in *Aspergillus* section *Flavi*. Aflatoxin is a notoriously potent hepatotoxin that causes acute toxicosis, cancer, immune suppression, and stunted growth in children (22–24), chronically impacting an estimated 4.5 billion people (25). However, only 40 to 60% of *A. flavus* isolates produce aflatoxin. While the ecological significance of BGCs are often unknown, aflatoxin production is thought to be favored in the presence of insects (2) but costly in their absence (26). Recently, Drott et al. (27) demonstrated that three distinct populations of *A. flavus* in the United States differ quantitatively in aflatoxin production. Interestingly, a small population at the northern edge of the species' range (28) that is closely related to the domesticated soy sauce-producing fungus *Aspergillus oryzae* was also the least aflatoxigenic population. Some have suggested (29, 30) that the lower prevalence of insects at higher latitudes favors nonaflatoxigenic isolates. These findings suggest that population-specific ecologies of *A. flavus* populations may favor differentiation of aflatoxin-producing ability. However, emblematic of the secondary metabolism literature, it is unclear if such population-specific patterns are common in the secondary metabolome or represent rare differentiation of specific BGCs. Recently, Uka et al. (31) demonstrated high levels of chemotypic diversity in *A. flavus* using 50 metabolites (including 11 precursors and derivatives of aflatoxin). However, their use of only a small number of molecular markers to identify population structure led them to assume that populations, as defined by geographic location, were panmictic. As mentioned above, such assumptions preclude the interpretation of intraspecific diversity as reflecting local adaptation of populations. We suggest that inferences of population-specific variation in the secondary metabolome of *A. flavus* will clarify the potential for

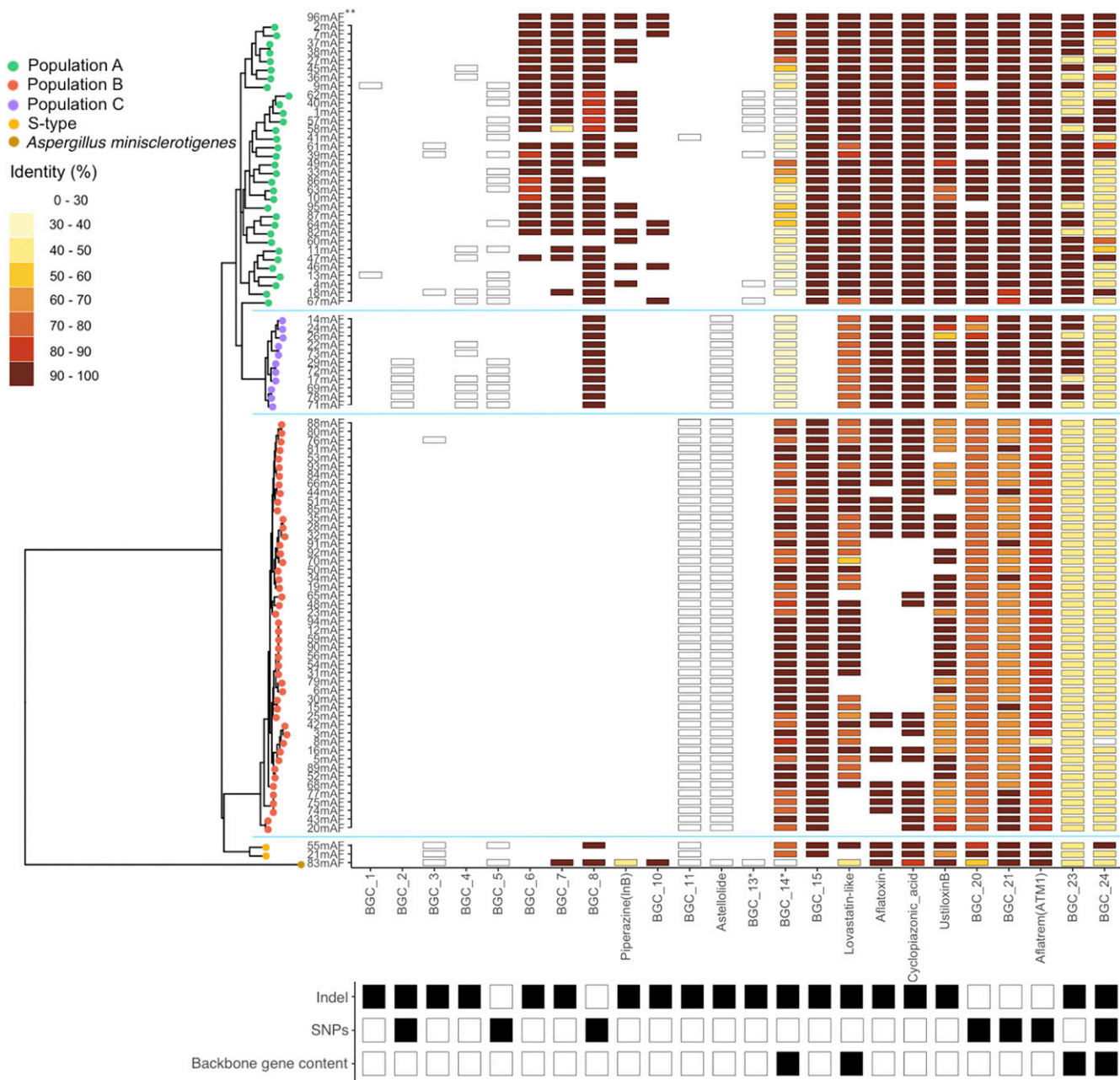
these ecologically important genes to drive population-level ecological and evolutionary processes.

Modern macroevolutionary theory infers ecological and evolutionary processes that occur at the population level. However, a dearth of information about how SM gene clusters are organized on this scale limits our understanding of SM evolution. The overall objective of this study was to determine how variation in BGCs and resulting SMs are distributed across the three *A. flavus* populations found in the United States (27). Because of the ecological relevance of many SMs, ecological differences are often inferred solely from differences in BGC content of genomes. However, it is phenotypes (chemotypes), not genotypes, that mediate ecological interactions and are subject to selection. We thus sought to understand the ecological and evolutionary significance of intraspecific variation in secondary metabolism by testing the following hypotheses: 1) variation in some BGCs show population-specific patterns in *A. flavus*; 2) population-specific BGCs that are missing from some populations are present in closely related species, illuminating ecological similarities and varied evolutionary histories of genes; and 3) population-specific differences in BGCs result in chemotypic differences that could be subject to selection.

## Results

**Pangenomic BGC Display Three Patterns of Population-Specific Conservation.** Using genome sequences with the BGCs removed, we identified the same populations (A, B, and C) as with the whole-genome data set (*SI Appendix, Fig. S1*), confirming that *A. flavus* population structure reflects differentiation across the genome, not just of BGCs. Network analysis of BGCs identified by antiSMASH (32) and by manual curation of characterized clusters dereplicated 7,821 BGCs across all isolates into 92 unique BGCs (individual isolates had between 78 and 85 BGCs with a median of 82 BGCs) that were shared by two or more individuals in the *A. flavus* pangenome (*SI Appendix, Fig. S1 and Table S1*). Of the pangenomic BGCs, 83 were found in the genome of the reference isolate obtained from Northern Regional Research Library (NRRL), NRRL3357 (BGCs and associated genes are listed in *SI Appendix, Table S2*). Three BGCs unique to single *A. flavus* isolates were removed from further analysis. While this approach determined the presence/absence of a BGC based on total gene content, given the magnitude of the dataset, we focused gene-level analyses on the 9,782 core-biosynthetic backbone genes (key genes that define the central structure of resulting SMs, e.g., polyketide synthase [PKS], dimethylallyl tryptophan synthase [DMATS], terpene synthases/cyclase, and nonribosomal peptide synthetase [NRPS]) that were identified within clusters. Consistent with previous assessments of intraspecific diversity (8, 9), we find isolate-specific variation in the form of deletions, single nucleotide polymorphisms (SNPs), and differences in gene content. For example, BGC 43 is only missing in isolate 41mAF (*SI Appendix, Fig. S2*). However, isolate-specific differences explained relatively little of the variation in BGCs when compared with population-specific variation (Fig. 1 compared with *SI Appendix, Fig. S2*).

Population-specific patterns of variation were evident in 24 BGCs (Fig. 1). Population-specific differences fell into three major categories (Fig. 1): The first two categories describe BGCs with a single backbone gene, where population-specific differences manifest as differences in protein sequence identity caused, in the first category, by the accumulation of SNPs or, in the second category, by partial or complete deletion. Differences in the third category of BGCs arose by the aforementioned mechanisms but in BGCs where there were multiple backbone genes that showed different patterns of variation from each other, for example, a backbone gene was deleted in some isolates while another backbone gene in the same BGC varied in identity. If all backbone genes within a BGC showed the same type of variation, they were categorized as one of the previous two types of variation.



**Fig. 1.** Heatmap showing BGCs with confirmed population-specific patterns among three genetically isolated L-type *A. flavus* populations, A (green), B (orange), and C (purple) identified previously (27) and confirmed in this study (SI Appendix, Fig. S1). In addition to 92 L-type *A. flavus* isolates, two S-strains (55mAF and 21mAF, Bottom, yellow) and one *Aspergillus minisclerotigenes* isolate (83mAF, Bottom, brown—formerly identified as *Aspergillus texensis* [Materials and Methods]) were included for context. The presence of a BGC, as indicated with a black-bordered box, was determined through network analysis of antiSMASH results that uses total gene content of BGCs or by previous studies (Materials and Methods). Absence of a BGC is indicated by blank spaces. When a BGC is not present in the reference genome, all corresponding boxes are white (BGCs 1 to 5 and 11 to 13). The protein sequence identity of backbone genes (core biosynthetic genes) was determined relative to the NRRL3357 reference genome (\*\*a high-quality assembly [96mAF] was used as the reference and was compared with an assembly [2mAF—which is located immediately below 96mAF in the figure] that resulted from methodology used on all other isolates [Materials and Methods]; the high-quality assembly was not included in phylogeny). Genetic relationships among isolates are depicted on the left side using a maximum likelihood tree based on genome-wide SNPs that were thinned to be at least 3 kb apart. Population structure was determined using discriminant analysis of principle components. Population-specific variation was categorized below the heatmap as being caused by insertion/deletion (indel) or differences in protein sequence identity of backbone genes (SNPs). A third type of variation was identified when the previous types of variation occurred differentially across multiple backbone genes in a single cluster (backbone gene content). The protein sequence identity and number of backbone genes, including those found by reciprocal best-hit BLAST but not incorporated into a BGC by antiSMASH, are indicated for all 92 BGCs analyzed in SI Appendix, Fig. S2 (reference SI Appendix, SI Text for validation and incorporation of reciprocal best-hit genes). \*These BGCs are physically overlapping but were distinguished by antiSMASH based on differences in gene content (see main text).

We performed extensive validations using a variety of alignment- and assembly-based methods (*SI Appendix, SI Text*) that extend substantially beyond validation typically reported in similar studies. We only interpret population-specific differences in BGCs where we could identify associated changes in protein domains (as identified using the National Center for Biotechnology Information's [NCBI's] conserved domain tool). We speculate that ecological forces may select for and maintain differentiation of population-specific patterns among BGCs. Estimates of selection derived from the site-frequency spectrum (e.g., comparing the number of nonsynonymous [dN] and synonymous [dS] SNPs) were originally developed for deeply divergent species and are insensitive to selective pressures when estimated within species, with dN/dS ratios deviating substantially from classic interpretations applied between species (reviewed by ref. 33). We find that the vast majority of population-specific BGCs have dN/dS ratios that indicate purifying selection (< 1.0) (*SI Appendix, Table S3*). This result is consistent with our interpretation that BGCs with population-specific variation are being retained by selection within respective populations. However, as ancient signatures of purifying selection can be very strong, new selective forces acting between recently diverged lineages may not be evident over ancient signals. Given problems associated with dN/dS estimates between closely related groups, we emphasize that selection acts on phenotypes and thus focus our efforts on identifying how population-specific variation is associated with differences in corresponding SM production. Furthermore, our focus on protein domains avoids issues where dN/dS ratios may not capture the outsized importance that individual SNPs can have (e.g., those that may impact the sequences associated with protein domains). Differences in proteins that result from relatively small differences in nucleotide and/or amino acid sequences are difficult to differentiate from assembly and annotation errors with complete confidence (notably, patterns of indels described in the *Pattern 1* section are easier to interpret than differences that arise from SNPs). Despite having validated population-specific differences using NCBI's conserved domain tool, the importance of these domains to the biosynthetic abilities of a protein are not always clear. We outline the details of the population-specific differences we found in the *Pattern 1*, *Pattern 2*, and *Pattern 3* sections to emphasize that a large portion of the diversity in BGCs can be explained by population-specific differences.

**Pattern 1: Indels Result in Population-Specific Differences in BGC Gene Content and Presence/Absence.** We found that ~15% of BGCs show clear population-specific patterns in presence/absence caused by deletions (BGCs 1, 3, 4, 6, 7, 10, 11, 13, 15, astellolide, aflatoxin, cyclopiazonic acid [CPA], and piperazine [InB]) (Fig. 1). Most of these BGCs, with the exception of 1 and 3, were found at relatively high frequency in at least one population but were completely missing in others. Interestingly, the backbone gene of BGC 13 is present in all isolates but was usually incorporated into BGC 14 (discussed in the *Pattern 3* section). BGC 13 was only considered a distinct cluster by antiSMASH in a subset of isolates where three additional cytochrome P450 enzymes and a transporter were also present. Production of astellolide was previously unknown in *A. flavus* but has been described in several other *Aspergillus* spp (34), as is discussed in the *Population-Specific BGCs Are Found in Isolates from Closely Related Species* section. This BGC was found in all isolates of populations B and C but in no isolates of population A, which contains the NRRL3357 *A. flavus* reference strain. Piperazine synthesis is mediated by two unlinked BGCs, InA and InB (16). While InA is present in all isolates, InB is missing from all isolates in populations B and C, raising questions about the role of this cluster without its biosynthetic counterpart. Deletions in the aflatoxin and CPA BGCs as well as the lovastatin-like (also discussed in the *Pattern 3* section) and ustiloxin B BGCs were only found in isolates of population B. Similar, though not identical,

sets of isolates with deletions observed in the aflatoxin and CPA BGCs may be explained by the physical proximity of these BGCs (*SI Appendix, Fig. S3*).

**Pattern 2: Population-Specific SNPs Result in Protein-Domain Shifts.**

In contrast to the highly conserved nature (>95% protein sequence identity) of most BGC backbone genes, we identified seven BGCs where backbone genes have diverged through the accumulation of SNPs to extents that result in protein-domain differences among populations (Fig. 1). For example, the identity of the backbone gene of the aflatrein (ATM1) BGC relative to the reference genome is lower for all isolates in population B (~89%) compared with population A (~97%), a difference that was associated with the presence of putative protein domains in population A isolates that were missing in population B isolates. Similarly, the terpene backbone gene of BGC 21 is truncated in most population B isolates, resulting in the loss of one of two putative terpene cyclase domains. Similar patterns of full domain or partial domain loss that result from the accumulation of SNPs were also observed in BGC 20. While these BGCs were still found by antiSMASH, some BGCs with similar patterns of divergence were not. A terpene backbone gene of BGC 5 was not found in any isolate of population B and was also missing in several isolates from population A despite >98% nucleotide identity in the associated region. While the domains in the backbone gene of BGC 8 appear to be retained, a less recognizable (lower e-value) DMAT domain may explain why antiSMASH did not identify these clusters in population B, despite being identified by reciprocal best-hit Basic Local Alignment Search Tool (BLAST) (*SI Appendix, Fig. S2*). Most isolates from populations A and B have the backbone gene of BGC 2 completely deleted; however, 15 isolates have variants where many of the NRPS domains, including a pp-binding and an AMP-binding domain, are missing, which we interpret as pseudogenized variants.

**Pattern 3: BGCs Where Variable Backbone Gene Content Shows Population-Specific Evolutionary Patterns.**

Because most of the BGCs we identified here are not fully characterized, it is unclear if differences in backbone gene content of BGCs with multiple backbone genes may modify or stop production of a resulting metabolite. In BGC 24, a terpene synthase backbone gene was deleted from eight isolates in population A and the *Aspergillus minisclerotigenes* isolate (Fig. 1). A full version of this backbone protein was only found in 13 population A isolates and one S-type isolate, suggesting pseudogenization of this gene in isolates with no apparent deletion. In both the lovastatin-like BGC and BGC 23, deletions of variable sizes were observed in only a subset of backbone genes found in each BGC. BGCs 13 and 14 (discussed above in the *Pattern 1* section) are physically overlapping clusters and together are part of a highly variable region where several different combinations of SM backbone genes exist. Some of the variation at this locus had previously been identified in comparisons with *A. oryzae* (35). In *Aspergillus fumigatus*, a similar set of BGC alleles was dubbed an “idiomorphic” BGC (8).

**Population-Specific BGCs Are Found in Isolates from Closely Related Species.**

The two S-type isolates of *A. flavus* have together retained all of the BGCs that were present in all isolates of the three L-type populations (*SI Appendix, Fig. S2*), as well as five BGCs that are otherwise only found in specific populations (Fig. 1 and *SI Appendix, Fig. S2*). While some S-type isolates are still considered to be part of *A. flavus* (36, 37), we also observed a similar pattern in the BGCs of the *A. minisclerotigenes* isolate. *A. minisclerotigenes* is a very closely related species to *A. flavus* that was used to determine if population-specific BGCs could be identified in other species. We found that nine population-specific BGCs were present in the *A. minisclerotigenes* isolate, for example, BGCs 23 and

24 (Fig. 1), as well as an additional 10 BGCs that were not found in any *A. flavus* isolates. Furthermore, the astellolide BGC, which has previously been described only in *A. oryzae* and *Aspergillus parasiticus* (34), is present in all population B and C isolates of *A. flavus*. Because *A. oryzae* is thought to be the domesticated form of *A. flavus*, we only interpret the presence of this BGC by *A. parasiticus*. These results emphasize the importance of population-level assessments of *Aspergillus* species, via pangenomic sequencing of populations, to better understand how population-specific BGCs may impact ecological inferences of comparative genomics and may inform varied evolutionary histories of some BGCs.

**Genetic Differences in BGCs Can Result in Population-Specific Patterns of SM Production.** For selection to act upon population-specific differences in BGCs, it is necessary that genetic differences result in associated phenotypes. We established that population-specific differences in BGC content resulted in chemotypic differences between isolates by identifying population-specific patterns in 13 characterized BGCs with known products (*SI Appendix, Table S3* and *Materials and Methods*). We found that aflatoxin, astellolide, CPA, ustiloxin B, aflatrem, and the piperazine BGCs showed population-specific genetic variation (Fig. 1). Complementarily, we identified differential production of most of these compounds, but piperazine [specifically, piperazine compound 7 identified by Forseth et al. (16)] was not detected in this analysis (Fig. 2 *B–G*). While sclerotia are produced by some isolates of *A. flavus* grown on both glucose-minimal medium (GMM) and potato dextrose agar (PDA) media, piperazine expression has previously been measured using growth media that greatly increases sclerotial production (16, 19), which we did not use, perhaps explaining the absence of these metabolites in our study. In general, populations where some isolates were missing associated BGCs had lower mean production of the associated SMs (Fig. 2).

In addition to differential production of compounds corresponding to population-specific patterns in BGCs, we also observed two SMs, aflavarin and leporin B, as differing significantly between populations despite having BGCs that were indistinguishable with our analyses. This finding may be explained by regulatory differences not within the scope of this study. Consistent with previous work on fungal SMs, we find that the magnitude of these differences varies based on media type (Fig. 2). Metabolic differentiation of populations was more evident in principle component analysis (PCA) plots stemming from growth on PDA medium than from growth on GMM (Fig. 3), likely reflecting a larger number or concentration of ions detected on PDA (Fig. 2*A*). Populations largely overlapped in metabolic space, although populations A and B both occupied some unique metabolic space (Fig. 3). Interestingly, the aflatoxin-producing ability of isolates differentiated some of the overall metabolome, with little overlap in metabolic space between nonaflatoxigenic and aflatoxigenic isolates (Fig. 3). This differentiation was particularly apparent along dimension 1, where aflatoxin explained more variation than 65% of other ions. However, aflatoxin's overall contribution to this axis was less than 1% for both PDA and GMM profiles.

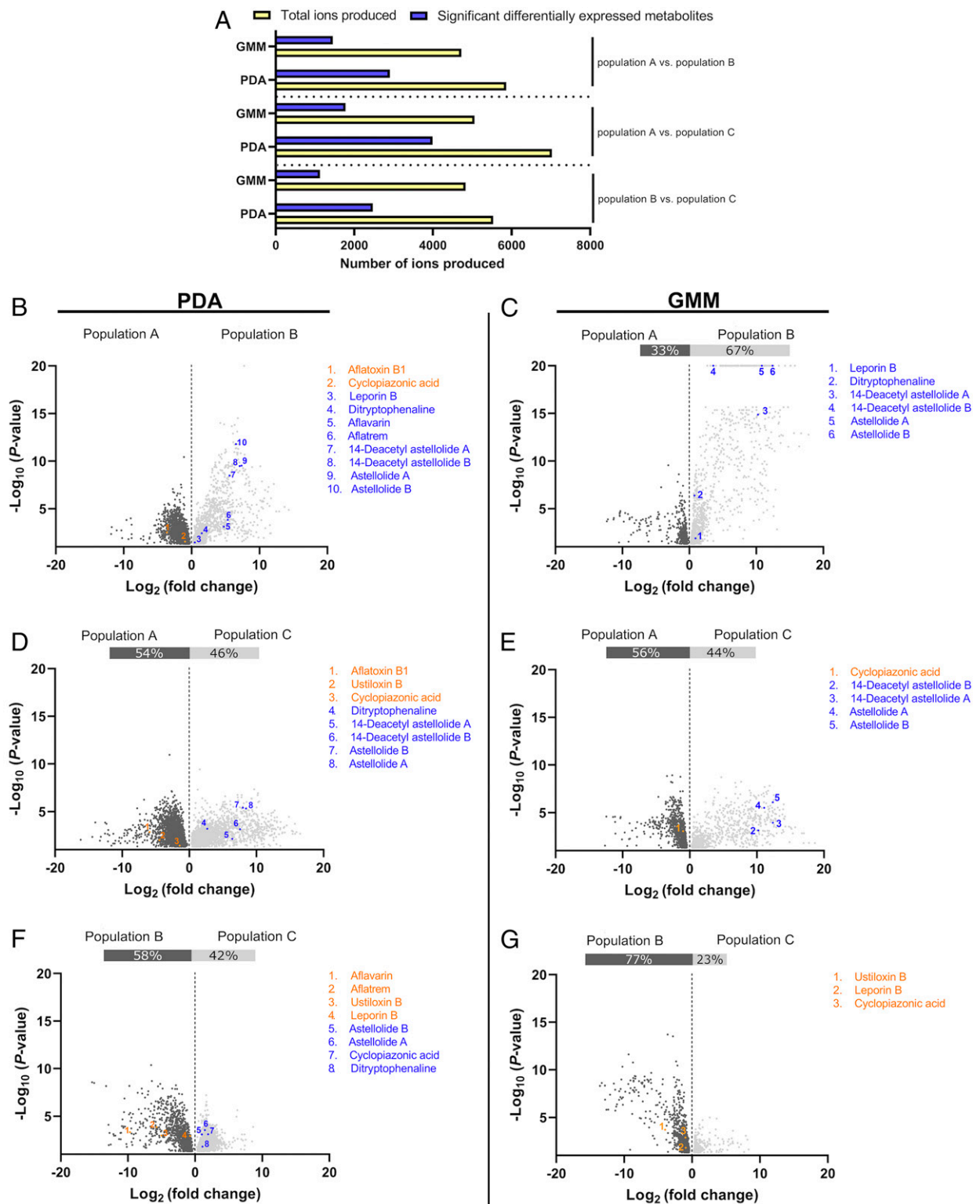
## Discussion

Our finding that ~25% of BGCs present within the *A. flavus* pangenome show population-specific differences (Fig. 1) has profound implications for inferences of ecological and macroevolutionary processes. Approximately 15% of BGCs were found in only a subset of populations, with population-specific differences in protein sequence identity accounting for another ~10% of BGCs (Fig. 1). Because selection acts on phenotypes, we emphasize the potential ecological and evolutionary importance of population differentiation of BGCs by demonstrating that, in several cases, these genomic differences result in different chemotypes (Fig. 2). Aflatoxin-producing ability differentiated a large portion of overall metabolic space (Fig. 3), suggesting that

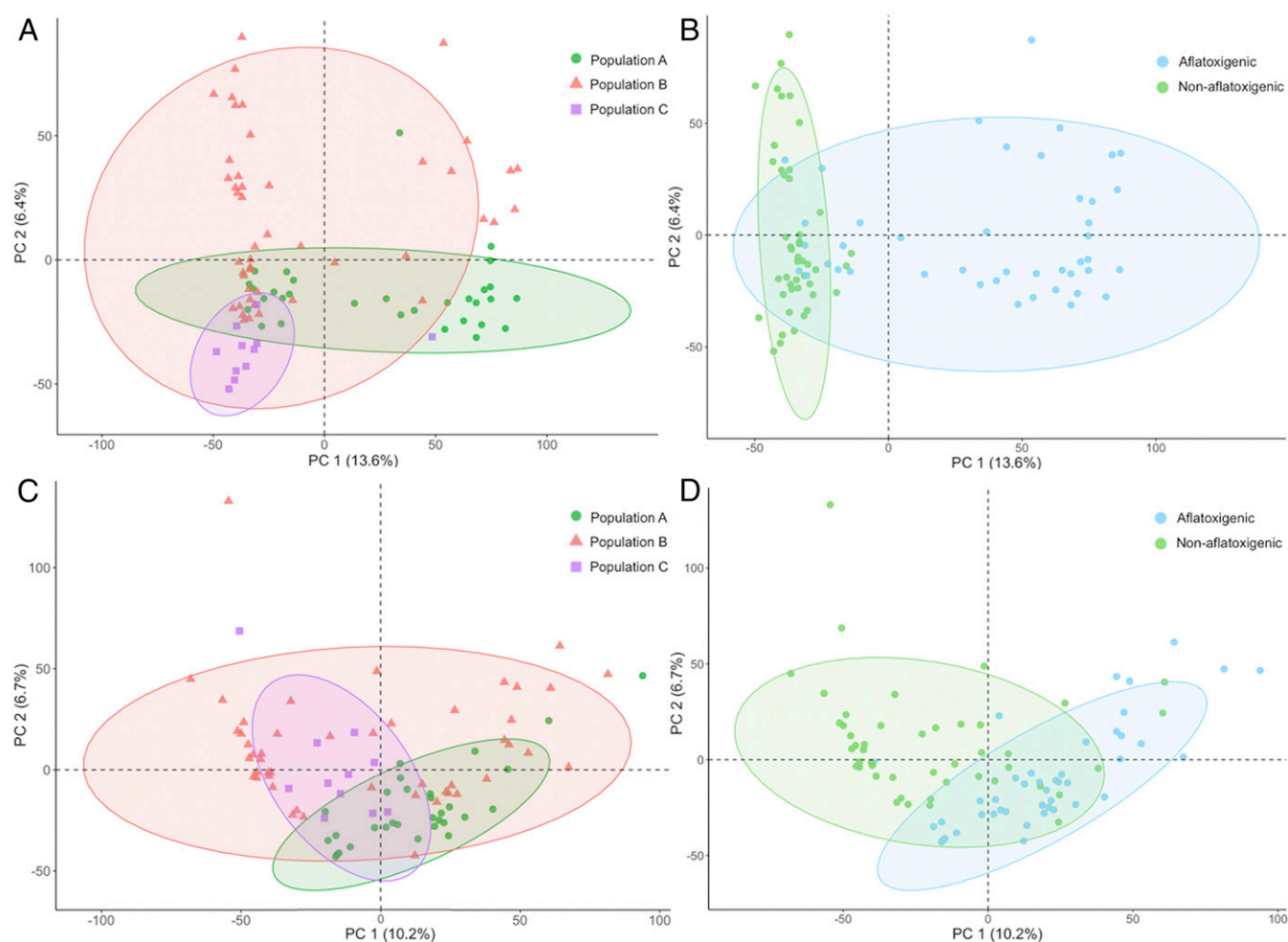
the metabolic consequences of differences in a single BGC can be wide-ranging. While differentiation of metabolic space based on aflatoxin-producing ability was independent of population structure, this finding raises the question whether similar patterns may also exist for population-specific differences in BGCs identified here. In addition to microevolutionary inferences, we show that several BGCs that are found in only a subset of populations, but not in the *A. flavus* reference genome, are present in closely related species, providing important context for macroevolutionary inferences and comparative genomic studies (as discussed below in this section). These results better align our understanding of fungal secondary metabolism with existing evolutionary theories that predict population-specific processes and have important implications for interpretation of past work and design of future studies.

The defining role of SMs on fungal niches is emphasized by a large number of recent studies that have used BGCs to infer ecological and evolutionary differences between species (38–44). However, many of these studies use single reference genomes to represent species. Without using ecological groupings, for example, pathogenic versus nonpathogenic species, whereby reference genomes become replicates within an ecotype (5, 45, 46), studies may effectively be unreplicated. Previous studies that looked for intraspecific variation in fungal secondary metabolism compared the mutations and extent of variation with that found between species but did not infer or discuss ecological distinctions within species or how such intraspecific variation might impact evolutionary inferences drawn between species (8, 40). Using population genomics, we demonstrated that a large portion of the intraspecific variation in BGCs in *A. flavus* is explained by differences between discrete populations. We speculate that population-specific variation reflects local adaptation associated with different BGCs or combinations of BGCs as mediated through chemotypes. Although we found that dN/dS ratios suggested evidence of purifying selection in most BGCs with population-specific variation (dN/dS < 1 as detailed in *SI Appendix, Table S3*), methods for inferring selection from dN/dS ratios were originally developed to describe evolution of genes in deeply diverged species and may be relatively insensitive and/or deviate from classical interpretations when applied between closely related groups as we have done here (reviewed by ref. 33). Because selection acts on phenotypes, we have focused our analysis on demonstrating that the patterns we identify are associated with chemotypic differences that could mediate ecological and evolutionary processes. Furthermore, we emphasize that even nonadaptive differentiation (e.g., as can occur through genetic drift) that occurs between populations may still be important for some ecological and evolutionary interactions if associated traits retain functionality (as we show using domain structures and chemotypes). Environmental change that is often associated with punctuated equilibrium and/or some epidemics (discussed further below in this section) can result in novel selective pressures that favor traits that were not previously under selection.

The potential ecological relevance of this diversity raises concerns about studies that do not account for population structure, as they may mischaracterize ecologically important differences between species. For example, comparing the *A. flavus* NRRL3357 reference genome with *A. minisclerotigenes* would incorrectly suggest that five clusters present in *A. minisclerotigenes* are not present in *A. flavus*. Similarly, we find the astellolide (also known as parasiticolide) BGC, which is also present in *A. parasiticus* (34), in two of the three *A. flavus* populations but not in population A, of which NRRL3357 is a member. As ecological functions are often (and increasingly) being assigned to SMs, it is important to interpret comparisons between species in the context of such intraspecific variation. Indeed, ~15% of *A. flavus* BGCs are specific to a single or subset of population(s), a number of BGCs similar to that found in some comparisons between closely related species (40).



**Fig. 2.** Production of metabolites by *A. flavus* isolates of populations A, B, and C [identified previously (27)] grown on PDA and GMM as detected by ultra-high-performance liquid chromatography–high-resolution mass spectrometry in positive mode. Ions showing significant differences ( $P < 0.05$ ) in pairwise comparison of populations were assumed to be metabolites and are referred to as significantly differentially expressed metabolites (SDEMs) (A). The magnitude and significance of SDEMs for each pairwise comparison are depicted in volcano plots for PDA (B, D, and F) and GMM (C, E, and G). Putative SMs are colored blue or orange to differentiate from unidentified metabolites, which are colored gray or dark gray. Points within each volcano plot were colored orange or dark gray if they were produced in higher abundance by the first-named population and blue or gray if produced in higher abundance by the second-named population. The identity of SMs was confirmed by fragmentation analysis (SI Appendix, SI Text).



**Fig. 3.** PCA of all metabolites produced by isolates of *A. flavus* populations A, B, and C [identified previously (27)] on PDA (A and B) and GMM (C and D). Isolates were assigned to populations in this study using discriminant analysis of principle components on variant data that does not include BGCs. Aflatoxin-producing ability was determined previously on several different media types including PDA (27, 28) but not GMM as aflatoxin production on the latter is typically low or absent. Ellipses represent 95% CI. While metabolic space of populations largely overlapped (A and B), aflatoxin-producing ability (C and D) was observed to differentiate metabolic space.

While delineations of populations from species are not always clear, the genetic distances we observe between the three *A. flavus* populations are much less than between *A. flavus* and the closely related species *A. minisclerotigenes* (as is evident in the phylogeny of Fig. 1); therefore, we interpret these as populations not species. Furthermore, we suggest that bio-prospecting efforts that have focused on intraspecific variation (47) may benefit from intentionally sampling populations, an effort that may be facilitated in some species by existing population genetic data.

Two recent analyses of the pangenomes of model fungi have concluded that differences in BGCs between species can be explained by strain-specific differences (8, 9). However, we find in *A. flavus* that by incorporating analysis of population structure, most strain-specific differences are organized into population-specific differences. While some fungi may have panmictic population structure, precluding similar patterns, we emphasize that many ecologically and taxonomically diverse fungi do have population structure (10–13), and recent large-scale population studies have even identified population structure in well-studied fungi that were previously inferred to be panmictic (14). The distinction between strain-specific and population- (or lineage-) specific differences is of great importance as the latter represents varied ecologies that arise from local adaptation while the former interprets differences between strains as representing variation around

a single mean ecology. The strain-specific interpretation lends itself to ecological comparisons between species using a small number of isolates (discussed above), as interspecific variation is assumed to be the result of a gradual accretion of mutations in a species as a whole (e.g., phyletic gradualism). However, Gould (48) referred to such generalization of microevolutionary processes to explain macroevolution as “extrapolationism.” He suggested that this approach precludes inference of important ecological and evolutionary processes by not considering alternative hypotheses that are rooted in ecological differentiation of populations within species (as discussed in the introduction). While we do not expect all fungi to show population structure, we suggest that patterns like those we find in *A. flavus* may be common across fungi and that differences in underlying population structure may allow for comparison of how secondary metabolism may evolve differently in structured versus unstructured species.

In bacteria, accessory genomes (comprising genes present in only a subset of isolates) are thought to be maintained through selection for ecological conditions that are relatively rare across the entire range of a species whereas core genomes (comprising genes present in all isolates) are thought to reflect a core ecology of species (49–51). While there have been rare suggestions that specific populations of fungi may contain unique BGCs (52) or unique alleles of BGCs (53), our findings that much of the



accessory secondary metabolome is explained by population-specific patterns in BGCs (Fig. 1) suggests that, at least in some species, SMs may be of profound importance for local adaptation and overall niche breadth of species. Furthermore, our results show how inferences of population structure raise questions about the ecology of coevolved combinations of genes. For example, the two piperazine BGCs (InA and InB) that are present in the *A. flavus* reference genome are known to have important regulatory cross talk (16). We find that while InA is part of the core genome (SI Appendix, Fig. S2), InB is part of the accessory genome, only present in a subset of isolates from population A (Fig. 1). While we were unable to detect piperazine production in this study, we speculate that cross talk between the accessory-genome BGC (InB) may have important ecological implications for the function of the core-genome BGC (InA).

The ecological and evolutionary implications of variation in BGCs is complemented by inferences of corresponding SM profiles (chemotypes)—this is an important addition to an evolutionary genomics study as selection acts upon phenotypic variation. While comparative genomic studies of fungal secondary metabolism have revealed interesting differences between species (38–44), inferences on adaptive significance in BGCs cannot be appreciated without determining SM profiles. Perhaps the best evidence of population-specific BGCs and subsequent SM production in fungi comes from *Fusarium graminearum*, where slight genetic variation in the trichothecene BGC between populations also results in different chemotypes (53–55). However, because of the very close relationship between the resulting SMs, there is controversy about the ecological significance of these differences (53), with some suggesting that inferences of population-specific patterns may sometimes result from methodological approaches (56). However, evidence that such variation in the trichothecene BGC is retained in an ancient transspecies polymorphism (57) does suggest differential ecological function of these BGC variants. In contrast to results on a single SM, our results with *A. flavus* clarify that population-specific patterns can affect a large portion of the pansecondary metabolome. We demonstrate clear differences in fungal chemotypes that result from population-specific patterns in the presence/absence of multiple BGCs (e.g., astellolide [Figs. 1 and 2]) and significant differences in metabolite expression that were not immediately explained by genetic differentiation (e.g., the aflavarin and leporin B BGCs were genetically indistinguishable using our methods [SI Appendix, Fig. S2]). Furthermore, PCA analysis of the entire metabolome differentiated aflatoxin-producing and non-aflatoxin-producing strains (Fig. 3), emphasizing that differences in a single metabolite may be associated with large-scale differences in fungal metabolism. While we have not explicitly tied these differences to ecological differentiation of populations, we suggest that differences in the geographic distribution of some populations assayed here (27) are consistent with potential variation in niche (58). In contrast to primary metabolites, SMs are often assumed to be maintained through niche-specific ecological interactions. Differences in BGC content (59) and in gene expression of some SMs may underlie differences between some biotrophic and necrotrophic life strategies (60). Similar differences in SM production are thought to be associated with niche adaptation of peripheral populations of some plant species (61). While we assume that conservation of SM production between populations reflects ecological importance, future studies are needed to elucidate the specific ecological functions of these BGCs (1, 27, 58).

In addition to evolutionary and ecological implications, population-specific patterns may also be important for applied efforts to control fungal disease outbreaks. The southern-corn-leaf-blight epidemic of 1970 was caused by a previously undetected lineage (race T) of *Cochliobolus heterostrophus* that harbored the T-toxin BGC. Corn (maize) varieties particularly susceptible to this

toxin were planted widely across the United States, resulting in race T's decimation of corn production in 1970. Since that outbreak, questions have remained about why race T had not been previously observed. Recently, Condon et al. (62) showed that the T-toxin BGC is ancient in *C. heterostrophus*, ruling out the possibility of a horizontal gene transfer near the time of the outbreak. The prevalence of population-specific SM demonstrated here raises questions of whether BGCs that are maintained in genetically isolated populations could explain past epidemics or could result in future epidemics when coupled with environmental change. The potential of locally adapted BGCs to drive novel epidemics may be of great importance as it has recently been suggested that climate change may be an important factor in favoring human pathogens (63). Consistent with ideas of punctuated equilibrium, environmental change may allow some populations to undergo range expansion. In new geographic areas, BGCs evolved in one ecological context may find novel functions (i.e., exaptation). Our results demonstrate the importance of intraspecific diversity for secondary metabolism that aligns this field with existing evolutionary theory in a way that offers opportunities to understand the ecology, evolution, and epidemiology of fungi.

## Materials and Methods

**Isolates of *A. flavus*.** We performed all analyses on existing genomic data from 92 *A. flavus* L-type isolates, two *A. flavus* S-type isolates, and a single isolate of the closely related S-type species *A. minisclerotigenes* [which was formerly identified as *Aspergillus texensis*, but whose taxonomy has since been further resolved (64) (SI Appendix, Table S1)]. In this manuscript, we refer to S-type isolates of *A. flavus* (37) as "S-type" while we refer to isolates from other species with similar morphology by their species name. Species-level identifications were previously confirmed (27) based on comparison of house-keeping gene sequences (SI Appendix, Fig. S4). These isolates comprise a stratified random sample from corn field soils in seven states spanning the eastern and central United States as described by Drott et al. (27). While isolate numbers are identical to those used by Drott et al. (27), we have amended isolate numbers with "mAF" to avoid confusion with similarly numbered BGCs. We used a high-quality assembly of NRRL3357 as a reference isolate (96mAF) but confirmed the consistency of our methods on lower-quality genome assemblies by also incorporating a genome assembly of NRRL3357 that was generated using identical methodologies to all other isolates (2mAF) (27). When we refer to the reference genome sequence, we are referring to the high-quality assembly of 96mAF from Drott et al. (27).

**Determination of Population Structure.** In order to confirm that previous inference of population structure (27) was not the result of between-subpopulation differences in BGCs, we removed SNPs from the Drott et al. (27) dataset that fell into regions containing BGCs (identified below in the *Identification of BGCs* section) using VCFtools (65). The remaining SNPs were analyzed using the non-model-based multivariate discriminant analysis of principle components from "adegetnet" v2.1.1 (66) implemented in R v3.5.2 (67) according to procedures outlined in the "adegetnet" tutorial (68).

Phylogenetic relationships between isolates was determined from a maximum likelihood analysis executed in Mega X (69) on a data set of whole-genome SNPs thinned to a minimum distance of 3,000 bp. This data set identically reflected isolate relationships evident in neighbor-net created using whole-genome data (27) but was more computationally feasible for this analysis. The resulting tree was graphed using "treeio" (70), "ggtree" (71), and "ggplot2" (72) in R.

**Identification of BGCs.** BGCs were identified from genome assemblies from Drott et al. (27) using antiSMASH v4.1.0 (32). We manually curated the borders of 13 characterized SM BGCs that have been associated with the production of specific SMs including the following: two aflatoxin BGCs (ATM1 and ATM2) (17), aflatoxin (73), aflavarin (74), asparosone A (75), aspergillilic acid (76), aspergillilicins (77), CPA (78), ditryptophenaline (79), imizoquin (80), leporin B (81), and two piperazine BGCs (InA and InB) (16). Additionally, we added the kojic acid (82) and ustiloxin B (83) BGCs, which were not found by antiSMASH. In a small number of instances where manual curation of characterized BGCs did not include nearby backbone genes (key genes that define the central structure of resulting SMs, e.g., polyketide and NRPS) that were incorporated into a BGC by antiSMASH, we separated out

the genes unique to the antiSMASH call and analyzed them separately from the curated BGC.

BGCs were dereplicated by creating networks with Biosynthetic Gene Similarity Clustering and Prospecting Engine (BiG-SCAPE) (84), which were visualized with Cytoscape v3.7.1 (85). Extensive validation of inferred BGC distributions and site-frequency estimates of selection (dN/dS) acting on these BGCs are detailed in *SI Appendix, SI Text*.

**Secondary Metabolite Extraction and Quantification.** Because selection acts upon phenotypes, we determined if differences in BGC content resulted in different chemotypes. SM production was assessed by growing all isolates on both PDA and on GMM at 30 °C for 14 d. Metabolites were extracted in ethyl acetate from three 1.2-cm plugs from each plate similar to methods described previously (86). Metabolic profiles were determined from these extracts using ultra-high-performance liquid chromatography with high-resolution mass spectrometry (UHPLC-HRMS) on a Thermo Scientific-Vanquish UHPLC system connected to a Thermo Scientific Q Exactive Orbitrap mass spectrometer in ES<sup>+</sup> mode between 200 and 1,000 *m/z* to identify metabolites as described previously (86). Acquisition and processing of UHPLC-MS data were done using the open-source software program, Maven version 2011.6.17 and the Thermo Scientific Xcalibur software version 4.3. The identification of putative SMs was further confirmed using the fragmentation patterns produced in tandem mass spectra (MS/MS) to annotate major identifiable peaks (*SI Appendix, Table S3 and Figs. S5–S29*). Additional details of metabolite identification and chemical analyses can be found in *SI Appendix, SI Text*.

- C. Schimek, "13 Evolution of special metabolism in fungi: Concepts, mechanisms, and pathways" in *Evolution of Fungi and Fungal-like Organisms*, S. Pöggler, J. Wöstmeier, Eds. (Springer, 2011), pp. 293–329.
- M. T. Drott, B. P. Lazzaro, D. L. Brown, I. Carbone, M. G. Milgroom, Balancing selection for aflatoxin in *Aspergillus flavus* is maintained through interference competition with, and fungivory by insects. *Proc. Biol. Sci.* **284**, 20172408 (2017).
- M. Khudhair *et al.*, Fusaristatin A production negatively affects the growth and aggressiveness of the wheat pathogen *Fusarium pseudograminearum*. *Fungal Genet. Biol.* **136**, 103314 (2020).
- W. P. Pfliegler, I. Pócsi, Z. Gyóri, T. Pusztahelyi, The *Aspergilli* and their mycotoxins: Metabolic interactions with plants and the soil biota. *Front. Microbiol.* **10**, 2921 (2020).
- X. Hu *et al.*, Trajectory and genomic determinants of fungal-pathogen speciation and host adaptation. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 16796–16801 (2014).
- P. Gan *et al.*, Comparative genomic and transcriptomic analyses reveal the hemibiotrophic stage shift of *Colletotrichum* fungi. *New Phytol.* **197**, 1236–1249 (2013).
- N. P. Keller, G. Turner, J. W. Bennett, Fungal secondary metabolism - from biochemistry to genomics. *Nat. Rev. Microbiol.* **3**, 937–947 (2005).
- A. L. Lind *et al.*, Drivers of genetic diversity in secondary metabolic gene clusters within a fungal species. *PLoS Biol.* **15**, e2003583 (2017).
- C. G. P. McCarthy, D. A. Fitzpatrick, Pan-genome analyses of model fungal species. *Microb. Genom.* **5**, e000243 (2019).
- H.-X. Li, M. T. Brewer, Spatial genetic structure and population dynamics of gummy stem blight fungi within and among watermelon fields in the southeastern United States. *Phytopathology* **106**, 900–908 (2016).
- A. Ahmadpour *et al.*, Population structure, genetic diversity, and sexual state of the rice brown spot pathogen *Bipolaris oryzae* from three Asian countries. *Plant Pathol.* **67**, 181–192 (2018).
- N. Nakamura, C. Tanaka, Y. Takeuchi-Kaneko, Recombination and local population structure of the root endophytic fungus *Glutonomyces brunneus* based on microsatellite analyses. *Fungal Ecol.* **41**, 56–64 (2019).
- J. L. Steenwyk, J. S. Soghigian, J. R. Perfect, J. G. Gibbons, Copy number variation contributes to cryptic genetic variation in outbreak lineages of *Cryptococcus gattii* from the North American Pacific Northwest. *BMC Genomics* **17**, 700 (2016).
- E. E. Ashu, F. Hagen, A. Chowdhary, J. F. Meis, J. Xu, Global population genetic analysis of *Aspergillus fumigatus*. *mSphere* **2**, e00019-17 (2017).
- M. T. Drott *et al.*, Diversity of secondary metabolism in *Aspergillus nidulans* clinical isolates. *mSphere* **5**, e00156-20 (2020).
- R. R. Forseth *et al.*, Homologous NRPS-like gene clusters mediate redundant small-molecule biosynthesis in *Aspergillus flavus*. *Angew. Chem. Int. Ed. Engl.* **52**, 1590–1594 (2013).
- M. J. Nicholson *et al.*, Identification of two aflatrems biosynthesis gene loci in *Aspergillus flavus* and metabolic engineering of *Penicillium paxilli* to elucidate their function. *Appl. Environ. Microbiol.* **75**, 7469–7481 (2009).
- P. F. Dowd, Synergism of aflatoxin B1 toxicity with the co-occurring fungal metabolite kojic acid to two caterpillars. *Entomol. Exp. Appl.* **47**, 69–71 (1988).
- D. R. Georgianna *et al.*, Beyond aflatoxin: Four distinct expression patterns and functional roles associated with *Aspergillus flavus* secondary metabolism gene clusters. *Mol. Plant Pathol.* **11**, 213–226 (2010).
- S. P. Otto, T. Lenormand, Resolving the paradox of sex and recombination. *Nat. Rev. Genet.* **3**, 252–261 (2002).
- A. O. B. Whitlock, R. B. R. Azevedo, C. L. Burch, Population structure promotes the evolution of costly sex in artificial gene networks. *Evolution* **73**, 1089–1100 (2019).
- J. H. Williams *et al.*, Human aflatoxicosis in developing countries: A review of toxicology, exposure, potential health consequences, and interventions. *Am. J. Clin. Nutr.* **80**, 1106–1122 (2004).
- Y. Liu, F. Wu, Global burden of aflatoxin-induced hepatocellular carcinoma: A risk assessment. *Environ. Health Perspect.* **118**, 818–824 (2010).
- C. P. Wild, Y. Y. Gong, Mycotoxins and human disease: A largely ignored global health issue. *Carcinogenesis* **31**, 71–82 (2010).
- CDC, Health Studies Branch. 2012. Understanding chemical exposures: aflatoxin. Atlanta, GA. <https://www.cdc.gov/nceh/hsb/chemicals/aflatoxin.htm>. Accessed 1 July 2020.
- M. T. Drott, T. Debenport, S. A. Higgins, D. H. Buckley, M. G. Milgroom, Fitness cost of aflatoxin production in *Aspergillus flavus* when competing with soil microbes could maintain balancing selection. *mBio* **10**, e02782–e02718 (2019).
- M. T. Drott *et al.*, The frequency of sex: Population genomics reveals differences in recombination and population structure of the aflatoxin-producing fungus *Aspergillus flavus*. *mBio* **11**, e00963-20 (2020).
- M. T. Drott, L. M. Fessler, M. G. Milgroom, Population subdivision and the frequency of aflatoxigenic isolates in *Aspergillus flavus* in the United States. *Phytopathology* **109**, 878–886 (2019).
- B. W. Horn, Biodiversity of *Aspergillus* section *Flavi* in the United States: A review. *Food Addit. Contam.* **24**, 1088–1101 (2007).
- D. T. Wicklow, P. F. Dowd, J. B. Gloer, "Antiinsect effects of *Aspergillus* metabolites" in *The Genus Aspergillus*, K. A. Powell, A. Renwick, J. F. Peberdy, Eds. (Plenum Press, New York, 1994), pp. 93–109.
- V. Uka *et al.*, Secondary metabolite dereplication and phylogenetic analysis identify various emerging mycotoxins and reveal the high intra-species diversity in *Aspergillus flavus*. *Front. Microbiol.* **10**, 667 (2019).
- K. Blin *et al.*, antiSMASH 4.0-improvements in chemistry prediction and gene cluster boundary identification. *Nucleic Acids Res.* **45**, W36–W41 (2017).
- S. Kryazhimskiy, J. B. Plotkin, The population genetics of dN/dS. *PLoS Genet.* **4**, e1000304 (2008).
- J. C. Frisvad *et al.*, Taxonomy of *Aspergillus* section *Flavi* and their production of aflatoxins, ochratoxins and other mycotoxins. *Stud. Mycol.* **93**, 1–63 (2019).
- J. G. Gibbons *et al.*, The evolutionary imprint of domestication on genome variation and function of the filamentous fungus *Aspergillus oryzae*. *Curr. Biol.* **22**, 1403–1409 (2012).
- P. Singh, M. J. Orbach, P. J. Cotty, *Aspergillus texensis*: A novel aflatoxin producer with S morphology from the United States. *Toxins (Basel)* **10**, 513 (2018).
- M. Ohkura, P. J. Cotty, M. J. Orbach, Comparative genomics of *Aspergillus flavus* S and L morphotypes yield insights into niche adaptation. *G3 (Bethesda)* **8**, 3915–3930 (2018).
- I. Kjørboelling *et al.*, Linking secondary metabolites to gene clusters through genome sequencing of six diverse *Aspergillus* species. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E753–E761 (2018).
- I. Kjørboelling *et al.*, A comparative genomics study of 23 *Aspergillus* species from section *Flavi*. *Nat. Commun.* **11**, 1106 (2020).
- T. C. Vesth *et al.*, Investigation of inter- and intraspecies variation through genome sequencing of *Aspergillus* section *Nigri*. *Nat. Genet.* **50**, 1688–1695 (2018).
- R. P. de Vries *et al.*, Comparative genomics reveals high biological diversity and specific adaptations in the industrially and medically important fungal genus *Aspergillus*. *Genome Biol.* **18**, 28 (2017).
- S. Rao, M. R. Nandinini, Genome sequencing and comparative genomics reveal a repertoire of putative pathogenicity genes in chilli anthracnose fungus *Colletotrichum truncatum*. *PLoS One* **12**, e0183567 (2017).

Drott *et al.*

Microevolution in the pansecondary metabolome of *Aspergillus flavus* and its potential macroevolutionary implications for filamentous fungi

PNAS | 9 of 10

<https://doi.org/10.1073/pnas.2021683118>

43. B. Pi *et al.*, A genomics based discovery of secondary metabolite biosynthetic gene clusters in *Aspergillus ustus*. *PLoS One* **10**, e0116089 (2015).
44. S. M. Tralamazza, L. O. Rocha, U. Oggenfuss, B. Corrêa, D. Croll, Complex evolutionary origins of specialized metabolite gene cluster diversity among the plant pathogenic fungi of the *Fusarium graminearum* species complex. *Genome Biol. Evol.* **11**, 3106–3122 (2019).
45. E. Martino *et al.*, Comparative genomics and transcriptomics depict ericoid mycorrhizal fungi as versatile saprotrophs and plant mutualists. *New Phytol.* **217**, 1213–1229 (2018).
46. T. A. Schuelke *et al.*, Comparative genomics of pathogenic and nonpathogenic beetle-vectored fungi in the genus *Geosmithia*. *Genome Biol. Evol.* **9**, 3312–3327 (2017).
47. C. Greco, N. P. Keller, A. Rokas, Unearthing fungal chemodiversity and prospects for drug discovery. *Curr. Opin. Microbiol.* **51**, 22–29 (2019).
48. S. J. Gould, *The Structure of Evolutionary Theory* (Harvard University Press, 2002).
49. L. Belbahri *et al.*, Comparative genomics of *Bacillus amyloliquefaciens* strains reveals a core genome with traits for habitat adaptation and a secondary metabolites rich accessory genome. *Front. Microbiol.* **8**, 1438 (2017).
50. P. G. Livingstone, R. M. Morphew, D. E. Whitworth, Genome sequencing and pan-genome analysis of 23 *Coralloccoccus* spp. strains reveal unexpected diversity, with particular plasticity of predatory gene sets. *Front. Microbiol.* **9**, 3187 (2018).
51. M. A. Brockhurst *et al.*, The ecology and evolution of pangenomes. *Curr. Biol.* **29**, R1094–R1103 (2019).
52. E.-M. Niehaus *et al.*, Comparative genomics of geographically distant *Fusarium fujikuroi* isolates revealed two distinct pathotypes correlating with secondary metabolite profiles. *PLoS Pathog.* **13**, e1006670 (2017).
53. A. C. Kelly, T. J. Ward, Population genomics of *Fusarium graminearum* reveals signatures of divergent evolution within a major cereal pathogen. *PLoS One* **13**, e0194616 (2018).
54. M. R. Fulcher, J. B. Winans, M. Quan, E. D. Oladipo, G. C. Bergstrom, Population genetics of *Fusarium graminearum* at the interface of wheat and wild grass communities in New York. *Phytopathology* **109**, 2124–2131 (2019).
55. A. Kelly *et al.*, The geographic distribution and complex evolutionary history of the NX-2 trichothecene chemotype from *Fusarium graminearum*. *Fungal Genet. Biol.* **95**, 39–48 (2016).
56. T. Crippin, J. B. Renaud, M. W. Sumarah, J. D. Miller, Comparing genotype and chemotype of *Fusarium graminearum* from cereals in Ontario, Canada. *PLoS One* **14**, e0216735 (2019).
57. T. J. Ward, J. P. Bielawski, H. C. Kistler, E. Sullivan, K. O'Donnell, Ancestral polymorphism and adaptive evolution in the trichothecene mycotoxin gene cluster of phytopathogenic *Fusarium*. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 9278–9283 (2002).
58. R. Levins, *Evolution in Changing Environments: Some Theoretical Explorations* (Princeton University Press, 1968).
59. J. Collemare, M. H. Lebrun, “Fungal secondary metabolites: Ancient toxins and novel effectors in plant-microbe interactions” in *Effectors in Plant-Microbe Interactions*, F. Martin, K. Sophien, Eds. (Wiley-Blackwell, 2011), pp. 377–400.
60. J. Collemare *et al.*, Secondary metabolism and biotrophic lifestyle in the tomato pathogen *Cladosporium fulvum*. *PLoS One* **9**, e85877 (2014).
61. S. Demasi *et al.*, Latitude and altitude influence secondary metabolite production in peripheral alpine populations of the Mediterranean species *Lavandula angustifolia* Mill. *Front. Plant Sci.* **9**, 983 (2018).
62. B. J. Condon *et al.*, Clues to an evolutionary mystery: The genes for T-Toxin, enabler of the devastating 1970 Southern corn leaf blight epidemic, are present in ancestral species, suggesting an ancient origin. *Mol. Plant Microbe Interact.* **31**, 1154–1165 (2018).
63. A. Casadevall, D. P. Kontoyiannis, V. Robert, On the emergence of *Candida auris*: Climate change, azoles, swamps, and birds. *mBio* **10**, e01397-19 (2019).
64. J. Houbraken *et al.*, Classification of *Aspergillus*, *Penicillium*, *Talaromyces* and related genera (*Eurotiales*): An overview of families, genera, subgenera, sections, series and species. *Stud. Mycol.* **95**, 5–169 (2020).
65. P. Danecek *et al.*; 1000 Genomes Project Analysis Group, The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
66. T. Jombart, adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics* **24**, 1403–1405 (2008).
67. R Core Team, R: A Language and Environment for Statistical Computing (Version 3.5.2, R Foundation for Statistical Computing, 2018).
68. T. Jombart (2012). An introduction to adegenet 2.0.0. <http://adegenet.r-forge.r-project.org/files/tutorial-basics.pdf>. Accessed 8 August 2020.
69. K. Tamura, J. Dudley, M. Nei, S. Kumar, MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* **24**, 1596–1599 (2007).
70. L.-G. Wang *et al.*, treeio: An R package for phylogenetic tree input and output with richly annotated and associated data. *Mol. Biol. Evol.* **37**, 599–603 (2020).
71. G. Yu, D. K. Smith, H. Zhu, Y. Guan, T. T. Y. Lam, ggtree: An R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol. Evol.* **8**, 28–36 (2017).
72. H. Wickham, *Ggplot2: Elegant Graphics for Data Analysis* (Springer-Verlag, 2016).
73. J. W. Cary, J. E. Linz, D. Bhatnagar, *Microbial Foodborne Diseases: Mechanisms of Pathogenesis and Toxin Synthesis* (CRC Press, 1999).
74. J. W. Cary *et al.*, Transcriptome analysis of *Aspergillus flavus* reveals veA-dependent regulation of secondary metabolite gene clusters, including the novel aflavarin cluster. *Eukaryot. Cell* **14**, 983–997 (2015).
75. J. W. Cary *et al.*, Functional characterization of a veA-dependent polyketide synthase gene in *Aspergillus flavus* necessary for the synthesis of asparosone, a sclerotium-specific pigment. *Fungal Genet. Biol.* **64**, 25–35 (2014).
76. M. D. Lebar *et al.*, Identification and functional analysis of the aspergillid acid gene cluster in *Aspergillus flavus*. *Fungal Genet. Biol.* **116**, 14–23 (2018).
77. C. Greco, B. T. Pfannenstiel, J. C. Liu, N. P. Keller, Depsipeptide aspergillidicins revealed by chromatin reader protein deletion. *ACS Chem. Biol.* **14**, 1121–1128 (2019).
78. P. K. Chang, B. W. Horn, J. W. Dorner, Clustered genes involved in cyclopiazonic acid production are next to the aflatoxin biosynthesis gene cluster in *Aspergillus flavus*. *Fungal Genet. Biol.* **46**, 176–182 (2009).
79. T. Saruwatari *et al.*, Cytochrome P450 as dimerization catalyst in diketopiperazine alkaloid biosynthesis. *ChemBioChem* **15**, 656–659 (2014).
80. S. Khalid *et al.*, NRPS-derived isoquinolines and lipopeptides mediate antagonism between plant pathogenic fungi and bacteria. *ACS Chem. Biol.* **13**, 171–179 (2018).
81. J. W. Cary *et al.*, An *Aspergillus flavus* secondary metabolic gene cluster containing a hybrid PKS-NRPS is necessary for synthesis of the 2-pyridones, leporins. *Fungal Genet. Biol.* **81**, 88–97 (2015).
82. Y. Terabayashi *et al.*, Identification and characterization of genes responsible for biosynthesis of kojic acid, an industrially important compound from *Aspergillus oryzae*. *Fungal Genet. Biol.* **47**, 953–961 (2010).
83. M. Umemura *et al.*, Characterization of the biosynthetic gene cluster for the ribosomally synthesized cyclic peptide ustiloxin B in *Aspergillus flavus*. *Fungal Genet. Biol.* **68**, 23–30 (2014).
84. J. C. Navarro-Muñoz *et al.*, A computational framework to explore large-scale biosynthetic diversity. *Nat. Chem. Biol.* **16**, 60–68 (2020).
85. P. Shannon *et al.*, Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
86. B. T. Pfannenstiel, C. Greco, A. T. Sukowaty, N. P. Keller, The epigenetic reader SntB regulates secondary metabolism, development and global histone modifications in *Aspergillus flavus*. *Fungal Genet. Biol.* **120**, 9–18 (2018).
87. M. Drott *et al.*, Raw genomic reads from “The frequency of sex: population genomics reveals differences in recombination and population structure of the aflatoxin-producing fungus *Aspergillus flavus*.” National Center for Biotechnology Information (NCBI). <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA639008>. Deposited 12 June 2020.
88. M. T. Drott *et al.*, Liquid chromatography with mass spectrometry data for “Microevolution in the pan-secondary metabolome of *Aspergillus flavus* and its potential macroevolutionary implications for filamentous fungi”. MassIVE. <https://doi.org/doi:10.25345/C54226>. Deposited 30 April 2021.