Variation Among Biosynthetic Gene Clusters, Secondary Metabolite Profiles, and Cards of Virulence Across Aspergillus Species

Jacob L. Steenwyk,* Matthew E. Mead,**1 Sonja L. Knowles,†1 Huzefa A. Raja,† Christopher D. Roberts, † Oliver Bader,‡ Jos Houbraken,§ Gustavo H. Goldman,** Nicholas H. Oberlies,† and Antonis Rokas*,2

*Department of Biological Sciences, Vanderbilt University, Nashville, Tennessee 37235, 1Department of Chemistry and Biochemistry, University of North Carolina at Greensboro, North Carolina 27402, 3Institute for Medical Microbiology, University Medical Center Göttingen, 37075, Germany, 4Westerdijk Fungal Biodiversity Institute, 3584 CT Utrecht, the Netherlands, and 2Faculdade de Ciências Farmacêuticas de Ribeirão Preto, Universidade de São Paulo, 14040-900, Brazil

ORCID IDs: 0000-0002-8436-595X (J.L.S.); 0000-0001-9195-7585 (M.E.M.); 0000-0002-6295-231X (S.L.K.); 0000-0002-0824-9463 (H.A.R.); 0000-0003-0949-1320 (C.D.R.); 0000-0002-1534-440X (O.B.); 0000-0003-4893-4438 (J.H.); 0000-0002-2986-350X (G.H.G.); 0000-0002-0354-8464 (N.H.O.); 0000-0002-7248-6551 (A.R.)

ABSTRACT Aspergillus fumigatus is a major human pathogen. In contrast, Aspergillus fischeri and the recently described Aspergillus oeringhausenensis, the two species most closely related to A. fumigatus, are not known to be pathogenic. Some of the genetic determinants of virulence (or “cards of virulence”) that A. fumigatus possesses are secondary metabolites that impair the host immune system, protect from host immune cell attacks, or acquire key nutrients. To examine whether secondary metabolism-associated cards of virulence vary between these species, we conducted extensive genomic and secondary metabolite profiling analyses of multiple A. fumigatus, one A. oeringhausenensis, and multiple A. fischeri strains. We identified two cards of virulence (gliotoxin and fumitremorgin) shared by all three species and three cards of virulence (trypacidin, pseurotin, and fumagillin) that are variable. For example, we found that all species and strains examined biosynthesized gliotoxin, which is known to contribute to virulence, consistent with the conservation of the gliotoxin biosynthetic gene cluster (BGC) across genomes. For other secondary metabolites, such as fumitremorgin, a modulator of host biology, we found that all species produced the metabolite but that there was strain heterogeneity in its production within species. Finally, species differed in their biosynthesis of fumagillin and pseurotin, both contributors to host tissue damage during invasive aspergillosis. A. fumigatus biosynthesized fumagillin and pseurotin, while A. oeringhausenensis biosynthesized fumagillin and A. fischeri biosynthesized neither. These biochemical differences were reflected in sequence divergence of the intertwined fumagillin/pseurotin BGCs across genomes. These results delineate the similarities and differences in secondary metabolism-associated cards of virulence between a major fungal pathogen and its nonpathogenic closest relatives, shedding light onto the genetic and phenotypic changes associated with the evolution of fungal pathogenicity.

KEYWORDS secondary metabolites; specialized metabolism; gliotoxin; chemodiversity; pathogenicity

Fungal diseases impose a clinical, economic, and social burden on humans (Drgona et al. 2014; Vallabhaneni et al. 2016; Benedict et al. 2019). Fungi from the genus Aspergillus are responsible for a considerable fraction of this burden, accounting for > 250,000 infections annually, with high mortality rates (Bongomin et al. 2017). Aspergillus infections often result in pulmonary and invasive diseases that are collectively termed aspergillosis. Among Aspergillus species, Aspergillus fumigatus is the primary etiological agent of aspergillosis (Latgé and Chamilos 2019).

Even though A. fumigatus is a major pathogen, its closest relatives are not considered pathogenic (Mead et al. 2019a; Steenwyk et al. 2019; Rokas et al. 2020b). Numerous studies have identified genetic determinants that contribute to
A. fumigatus pathogenicity, such as the organism’s ability to grow well at higher temperatures and in hypoxic conditions (Kamei and Watanabe 2005; Tekiaa and Latgé 2005; Abad et al. 2010; Grah et al. 2012). Genetic determinants that contribute to pathogenicity could be conceived as analogous to individual “cards” of a “hand” (set of cards) in a card game—that is, individual determinants are typically insufficient to cause disease but can collectively do so (Casadevall 2007).

A. fumigatus biosynthesizes a cadre of secondary metabolites and several metabolites could be conceived as “cards” of virulence because of their involvement in impairing the host immune system, protecting the fungus from host immune cell attacks, or acquiring key nutrients (Shwab et al. 2007; Losada et al. 2009; Yin et al. 2013; Wiemann et al. 2014; Bignell et al. 2016; Knox et al. 2016; Raffa and Keller 2019; Blachowicz et al. 2020). For example, the secondary metabolite gliotoxin has been shown in A. fumigatus to inhibit the host immune response (Sugi et al. 2007; Spikes et al. 2008). Other secondary metabolites implicated in virulence include fumitremorgin, which inhibits the activity of the breast cancer resistance protein (González-Lobato et al. 2010); verruculogen, which modulates the electrophysical properties of human nasal epithelial cells (Khouchafe et al. 2007); trypanadin, which is cytotoxic to lung cells and inhibits phagocytosis (Gauthier et al. 2012; Matern et al. 2015); pseudotin, which inhibits immunoglobulin E (Ishikawa et al. 2009); and fumagillin, which causes epithelial cell damage (Guruceaga et al. 2018) and impairs the function of neutrophils (Fallon et al. 2010, 2011).

By extension, the metabolic pathways responsible for the biosynthesis of secondary metabolites could also be conceived as components of these secondary metabolism-associated “cards” of virulence. Genes in these pathways are typically organized in contiguous sets termed biosynthetic gene clusters (BGCs) (Keller 2019). BGCs are known to evolve rapidly, and their composition can differ substantially across species and strains (Lind et al. 2015, 2017; de Vries et al. 2017; Kjaerbelling et al. 2018, 2020; Rokas et al. 2018, 2020a; Vesth et al. 2018). For example, even though A. fumigatus contains 33 BGCs and Aspergillus fischeri contains 48 BGCs, only 10 of those BGCs appear to be shared between the two species (Mead et al. 2019a). Interestingly, one of the BGCs that is conserved between A. fumigatus and A. fischeri is the gliotoxin BGC and both species have been shown to biosynthesize the secondary metabolite albeit at different amounts (Knowles et al. 2020). These results suggest that the gliotoxin “card” is part of a winning “hand” that facilitates virulence only in the background of the major pathogen A. fumigatus and not in that of the nonpathogen A. fischeri (Knowles et al. 2020).

To date, such comparisons of BGCs and secondary metabolite profiles among A. fumigatus and closely related nonpathogenic species have been few and restricted to single strains (Mead et al. 2019a; Knowles et al. 2020). However, genetic and phenotypic heterogeneity among strains of a single species is an important consideration when studying Aspergillus pathogenicity (Kowalski et al. 2016, 2019; Keller 2017; Ries et al. 2019; Blachowicz et al. 2020; Bastos et al. 2020; Drott et al. 2020; dos Santos et al. 2020; Steenwyk et al. 2020). Examination of multiple strains of A. fumigatus and close relatives—including the recently described closest known relative of A. fumigatus, Aspergillus oerlinghausenensis, whose virulence has yet to be examined but which is not thought to be a human pathogen (Houbraken et al. 2016) and has never been associated with human infections—will increase our understanding of the A. fumigatus secondary metabolism-associated “cards” of virulence.

To gain insight into the genomic and chemical similarities and differences in secondary metabolism among A. fumigatus and nonpathogenic close relatives, we characterized variation in BGCs and secondary metabolites produced by A. fumigatus and nonpathogenic close relatives. To do so, we first sequenced and assembled A. oerlinghausenensis CBS 139183T as well as A. fischeri strains NRRL 4585 and NRRL 4161 and analyzed them together with four A. fumigatus and three additional A. fischeri publicly available genomes. We also characterized the secondary metabolic profiles of three A. fumigatus, one A. oerlinghausenensis, and three A. fischeri strains. We observed both variation and conservation among species- and strain-level BGCs and secondary metabolites. We found that the biosynthesis of the secondary metabolites gliotoxin and fumitremorgin, which are both known to interact with mammalian cells (Yamada et al. 2000; González-Lobato et al. 2010; Li et al. 2012; Raffa and Keller 2019), as well as their BGCs, were conserved among pathogenic and nonpathogenic strains. Interestingly, we found only A. fischeri strains, but not A. fumigatus strains, biosynthesized verruculogen, which changes the electrophysical properties of human nasal epithelial cells (Khouchafe et al. 2007). Similarly, we found that both A. fumigatus and A. oerlinghausenensis biosynthesized fumagillin and trypanadin, whose effects include broad suppression of the immune response system and lung cell damage (Ishikawa et al. 2009; Fallon et al. 2010, 2011; Gauthier et al. 2012), but A. fischeri did not. Taken together, these results reveal that nonpathogenic close relatives of A. fumigatus also produce some, but not all, of the secondary metabolism-associated cards of virulence known in A. fumigatus. Further investigation of the similarities and differences among A. fumigatus and close nonpathogenic relatives may provide additional insight into the “hand of cards” that enabled A. fumigatus to evolve into a deadly pathogen.

Materials and Methods

Strain acquisition, DNA extraction, and sequencing

Two strains of A. fischeri (NRRL 4161 and NRRL 4585) were acquired from the Northern Regional Research Laboratory (NRRL) at the National Center for Agricultural Utilization Research in Peoria, Illinois, while one strain of A. oerlinghausenensis (CBS 139183T) was acquired from the Westerdijk
Fungal Biodiversity Institute, Utrecht, The Netherlands. These strains were grown in 50 ml of liquid yeast extract soy peptone dextrose (YESD) medium. After ~7 days of growth on an orbital shaker (100 rpm) at room temperature, the mycelium was harvested by filtering the liquid media through a Corning, 150 ml bottle top, 0.22 μm sterile filter and washed with autoclaved distilled water. All subsequent steps of DNA extraction from the mycelium were performed following protocols outlined previously (Mead et al. 2019b). The genomic DNA from these three strains was sequenced using a NovaSeq S4 at the Vanderbilt Technologies for Advanced Genomes facility (Nashville, Tennessee) using paired-end sequencing (150 bp) strategy with the Illumina TruSeq library kit.

**Genome assembly, quality assessment, and annotation**

To assemble and annotate the three newly sequenced genomes, we first quality-trimmed raw sequence reads using Trimmomatic, v0.36 (Bolger et al. 2014) using parameters described elsewhere (ILLUMINACLIP:TruSeq3-PE.fa:2:30:10, leading:10, trailing:10, slidingwindow:4:20, minlen:50; Steenwyk and Rokas 2017). The resulting paired and unpaired quality-trimmed reads were used as input to the SPAdes, v3.11.1 (Bankevich et al. 2012), genome assembly algorithm with the “careful” parameter and the “cov-cutoff” set to “auto”.

We evaluated the quality of our newly assembled genomes, using metrics based on continuity of assembly and gene-content completeness. To evaluate genome assemblies by scaffold size, we calculated the N50 of each assembly (or missing. Our analyses indicate the newly sequenced and assembled genomes have high gene-content completeness and assembly continuity (average percent presence of BUSCO genes: 98.80 ± 0.10%; average N50: 451,294.67 ± 9,692.11; Supplemental Material, Figure S1). These metrics suggest these genomes are suitable for comparative genomic analyses.

To predict gene boundaries in the three newly sequenced genomes, we used the MAKER, v2.31.10, pipeline (Holt and Yandell 2011), which creates consensus predictions from the collective evidence of multiple *ab initio* gene prediction software. Specifically, we created consensus predictions from SNAP, v2006-07-28 (Korf 2004), and AUGUSTUS, v3.3.2 (Stanke and Waack 2003), after training each algorithm individually on each genome. To do so, we first ran MAKER using protein evidence clues from five different publicly available annotations of *Aspergillus* fungi from section *Fumigati*. Specifically, we used protein homology clues from *A. fischeri* NRRL 181 (GenBank accession: GCA_000149645.2), *A. fumigatus* Af293 (GenBank accession: GCA_000002655.1), *Aspergillus lentulus* IFM 54703 (GenBank accession: GCA_001445615.1), *Aspergillus novofumigatus* IBT 16806 (GenBank accession: GCA_002847465.1), and *Aspergillus udaigwae* IFM 46973 (GenBank accession: GCA_001078395.1). The resulting gene predictions were used to train SNAP. MAKER was then rerun using the resulting training results. Using the SNAP trained gene predictions, we trained AUGUSTUS. A final set of gene boundary predictions were obtained by rerunning MAKER with the training results from both SNAP and AUGUSTUS.

To supplement our data set of newly sequenced genomes, we obtained publicly available ones. Specifically, we obtained genomes and annotations for *A. fumigatus* Af293 (GenBank accession: GCA_000002655.1), *A. fumigatus* CEA10 (strain synonym: CBS 144.89/ FGSC A1163; GenBank accession: GCA_001510451.1), *A. fumigatus* HMR AF 270 GenBank accession: GCA_002234955.1), *A. fumigatus* Z5 (GenBank accession: GCA_001029325.1), *A. fischeri* NRRL 181 (GenBank accession: GCA_00149645.2). We also obtained assemblies of the recently published *A. fischeri* genomes for strains IBT 3003 and IBT 3007 (Zhao et al. 2019), which lacked annotations. We annotated the genome of each strain individually using MAKER with the SNAP and AUGUSTUS training results from a close relative of both strains: *A. fischeri* NRRL 4161. Altogether, our final data set contained a total of 10 genomes from three species: four *A. fumigatus* strains, one *A. oerlinghausenensis* strain, and five *A. fischeri* strains (Table 1).

**Maximum likelihood phylogenetics and Bayesian estimation of divergence times**

To reconstruct the evolutionary history among the 10 *Aspergillus* genomes, we implemented a recently developed pipeline (Steenwyk et al. 2019) that relies on the concatenation-approach to phylogenomics (Rokas et al. 2003) and has been successfully used in reconstructing species-level relationships among *Aspergillus* and *Penicillium* fungi (Bodinaku et al. 2019; Steenwyk et al. 2019). The first step in the pipeline is to identify single-copy orthologous genes in the genomes of interest, which are ultimately concatenated into a larger phylogenomic data matrix. To identify single copy BUSCO genes across all 10 *Aspergillus* genomes, we used the BUSCO pipeline with the Pezizomycotina database as described above. We identified 3041 BUSCO genes present at a single copy in all 10 *Aspergillus* genomes, and created multi-FASTA files for each BUSCO gene that contained the protein sequences for all 10 taxa. The protein sequences of each BUSCO gene were individually aligned using Mafft, v7.4.02 (Katoh and Standley 2013), with the same parameters as described elsewhere (Steenwyk et al. 2019). Nucleotide sequences were then mapped onto the protein sequence alignments using a custom Python, v3.5.2 ([https://www.python.org/](https://www.python.org/)) script with BioPython v1.7 (Cock et al. 2009). The resulting codon-based
alignments were trimmed using trimAl, v1.2.rev59 (Capella-Gutierrez et al. 2009), with the “gappyout” parameter. The resulting trimmed nucleotide alignments were concatenated into a single matrix of 5,602,272 sites and was used as input into IQ-TREE, v1.6.11 (Ngo et al. 2015). The best-fitting model of substitutions for the entire matrix was determined using Bayesian information criterion values (Kalyaanamoorthy et al. 2017). The best-fitting model was a general time-reversible model with empirical base frequencies that allowed for a proportion of invariable sites and a discrete Gamma model with four rate categories (GTR+I+F+G4) (Tavaré 1986; Yang 1994, 1996; Vinet and Zedanov 2011). To evaluate bipartition support, we used 5000 ultrafast bootstrap approximations (Hoang et al. 2018).

To estimate divergence times among the 10 Aspergillus genomes, we used the concatenated data matrix and the resulting maximum likelihood phylogeny from the previous steps as input to Bayesian approach implemented in MCMCTree from the PAML package, v4.9d (Yang 2007). First, we estimated the substitution rate across the data matrix using a “GTR+G” model of substitutions (model = 7), a strict clock model, and the maximum likelihood phylogeny rooted on the clade of A. fischeri strains. We imposed a root age of 3.69 million years ago (MYA) according to results from recent divergence time estimates of the split between A. fischeri and A. fumigatus (Steenwyk et al. 2019). We estimated the substitution rate to be 0.005 substitutions per 1 million years. Next, the likelihood of the alignment was approximated using a gradient and Hessian matrix. To do so, we used previously established time constraints for the split between A. fischeri and A. fumigatus (1.85–6.74 MYA) (Steenwyk et al. 2019). Lastly, we used the resulting gradient and Hessian matrix, the rooted maximum likelihood phylogeny, and the concatenated data matrix to estimate divergence times using a relaxed molecular clock (model = 2). We specified the substitution rate prior based on the estimated substitution rate (rgene_gama = 1 186.63). The “sigma2_gama” and “finetune” parameters were set to ‘1 4.5’ and ‘1’, respectively. To collect a high-quality posterior probability distribution, we ran a total of 5.1 million iterations during MCMC analysis, which is 510 times greater than the minimum recommendations (Raftery and Lewis 1995). Our sampling strategy across the 5.1 million iterations was to discard the first 100,000 results followed by collecting a sample every 500th iteration until a total of 10,000 samples were collected.

Identification and characterization of secondary metabolite production

General experimental procedures: The 1H NMR data were collected using a JOEL ECS-400 spectrometer, which was equipped with a JOEL normal geometry broadband Royal probe, and a 24-slot autosampler, and operated at 400 MHz. HRESIMS experiments utilized either a Thermo LTQ Orbitrap XL mass spectrometer or a Thermo Q Exactive Plus (Thermo Fisher Scientific); both were equipped with an electrospray ionization source. A Waters Acquity ultraperformance liquid chromatography (UPLC; Waters Corp.) was utilized for both mass spectrometers, using a BEH C18 column (1.7 µm; 50 mm × 2.1 mm) set to a temperature of 40° and a flow rate of 0.3 ml/min. The mobile phase consisted of a linear gradient of CH3CN-H2O (both acidified with 0.1% formic acid), starting at 15% CH3CN and increasing linearly to 100% CH3CN over 8 min, with a 1.5-min hold before returning to the starting condition. The HPLC separations were performed with Atlantis T3 C18 semipreparative (5 µm; 10 × 250 mm) and preparative (5 µm; 19 × 250 mm) columns, at a flow rate of 4.6 ml/min and 16.9 ml/min, respectively, with a Varian Prostar HPLC system equipped with a Prostar 210 pumps and a Prostar 335 photodiode array (PDA).
detector, with the collection and analysis of data using Galaxie Chromatography Workstation software. Flash chromatography was performed on a Teledyne ISCO CombiFlash Rf 200 and monitored by both evaporative light scattering detector (ELSD) and PDA detectors.

**Chemical characterization:** To identify the secondary metabolites that were biosynthesized by *A. fumigatus*, *A. oerlinghausenensis*, and *A. fischeri*, these strains were grown as large-scale fermentations to isolate and characterize the secondary metabolites. To inoculate oatmeal cereal media (Old Fashioned Breakfast Quaker oats), agar plugs from fungal stains grown on potato dextrose agar; Difco were excised from the edge of the Petri dish culture and transferred to separate liquid seed medium that contained 10 ml YESD broth (2% soy peptone, 2% dextrose, and 1% yeast extract; 5 g of yeast extract, 10 g of soy peptone, and 10 g of d-glucose in 500 ml of deionized H$_2$O) and allowed to grow at 23° with agitation at 100 rpm for 3 days. The YESD seed cultures of the fungi were subsequently used to inoculate solid-state oatmeal fermentation cultures, which were either grown at room temperature (~23° under 12 hr light/dark cycles for 14 days), 30°, or 37°; all growth at the latter two temperatures was carried out in an incubator (VWR International) in the dark over 4 days. The oatmeal cultures were prepared in 250 ml Erlenmeyer flasks that contained 10 g of autoclaved oatmeal (10 g of oatmeal with 17 ml of deionized H$_2$O and sterilized for 15–20 min at 121°). For all fungal strains, three flasks of oatmeal cultures were grown at all three temperatures, except for *A. oerlinghausenensis* (CBS 139183T) at room temperature and *A. fumigatus* (Af293) at 37°. For CBS 139183T, the fungal cultures were grown in four flasks, while for Af293 eight flasks were grown in total. The growths of these two strains were performed differently from the rest because larger amounts of extract were required in order to perform detailed chemical characterization.

The cultures were extracted by adding 60 ml of (1:1) MeOH-CHCl$_3$ to each 250 ml flask, shaking thoroughly with a spatula, and shaking overnight (~16 hr) at ~100 rpm at room temperature. The culture was filtered in vacuo, and 90 ml CHCl$_3$ and 150 ml H$_2$O were added to the filtrate. The mixture was stirred for 30 min and then transferred to a separatory funnel. The organic layer (CHCl$_3$) was drawn off and evaporated to dryness in vacuo. The dried organic layer was reconstituted in 100 ml of (1:1) MeOH–CH$_3$CN and 100 ml of hexanes, transferred to a separatory funnel, and shaken vigorously. The defatted organic layer (MeOH–CH$_3$CN) was evaporated to dryness in vacuo.

To isolate compounds, the defatted extract was dissolved in CHCl$_3$, absorbed onto Celite 545 (Acros Organics), and fractioned by normal phase flash chromatography using a gradient of hexane-CHCl$_3$-MeOH. *A. fischeri* strain NRRL 181 was chemically characterized previously (Mead et al. 2019a; Knowles et al. 2019). *A. fumigatus* strain AF293, grown at 37°, was subjected to a 12 g column at a flow rate of 30 ml/min and 61.0 column volumes, which yielded four fractions. Fraction 2 was further purified via preparative HPLC using a gradient system of 30:70 to 100:0 of CH$_3$CN–H$_2$O with 0.1% formic acid over 40 min at a flow rate of 16.9 ml/min to yield six subfractions. Subfractions 1, 2 and 5, yielded cyclo(L-Pro-L-Leu) (Li et al. 2008) (0.89 mg), cyclo(L-Pro-L-Phe) (Campbell et al. 2009) (0.71 mg), and monomethylsulfoxochrin (Ma et al. 2004) (2.04 mg), which eluted at ~5.7, 6.3, and 10.7 min, respectively. Fraction 3 was further purified via preparative HPLC using a gradient system of 40:60 to 65:35 of CH$_3$CN–H$_2$O with 0.1% formic acid over 30 min at a flow rate of 16.9 ml/min to yield four subfractions. Subfractions 1 and 2 yielded pseurotin A (Wang et al. 2011) (12.50 mg) and bisdethiobis (methylthio) gliotoxin (Afayutullov et al. 2005) (13.99 mg), which eluted at ~7.5 and 8.0 min, respectively.

*A. fumigatus* strain CEA10, grown at 37°, was subjected to a 4 g column at a flow rate of 18 ml/min and 90.0 column volumes, which yielded five fractions. Fraction 1 was purified via preparative HPLC using a gradient system of 50:50 to 100:0 of CH$_3$CN–H$_2$O with 0.1% formic acid over 45 min at a flow rate of 16.9 ml/min to yield eight subfractions. Subfraction 1, yielded fumagillin (Halász et al. 1998) (1.69 mg), which eluted at ~18.5 min. Fraction 2 was purified via semi-preparative HPLC using a gradient system of 35:65 to 80:20 of CH$_3$CN–H$_2$O with 0.1% formic acid over 30 min at a flow rate of 4.6 ml/min to yield 10 subfractions. Subfraction 5 yielded fumitremorgin C (Kato et al. 2007) (0.30 mg), which eluted at ~15.5 min. Subfraction 3 was further purified via preparative HPLC using a gradient system of 40:60 to 100:0 of CH$_3$CN–H$_2$O with 0.1% formic acid over 30 min at a flow rate of 16.9 ml/min to yield nine subfractions. Subfraction 2 yielded pseurotin A (1.64 mg), which eluted at ~7.3 min.

*A. oerlinghausenensis* strain CBS 139183T, grown at RT, was subjected to a 4 g column at a flow rate of 18 ml/min and 90 column volumes, which yielded four fractions. Fraction 3 was further purified via preparative HPLC using a gradient system of 35:65 to 70:30 of CH$_3$CN–H$_2$O with 0.1% formic acid over 40 min at a flow rate of 16.9 ml/min to yield 11 subfractions. Subfractions 3 and 10 yielded spiro[5H,10H-dipyrrolo[1,2-a:1,2'-d]pyrazine-2-(3H),2'-(2H)indole]-3',5,10(1'H)-trione (Wang et al. 2008) (0.64 mg) and helvolic acid (Zhao et al. 2010) (1.03 mg), which eluted at ~11.5 and 39.3 min, respectively (see NMR supporting information; figshare: https://doi.org/10.6084/m9.figshare.12055503).

**Metabolite profiling by mass spectrometry:** The metabolite profiling by mass spectrometry, also known as dereplication, was performed as stated previously (El-Elimat et al. 2013). Briefly, UPLC-PDA-electrospray ionization high resolution tandem mass spectrometry (UPLC-PDA-HRMS-MS/MS) was utilized to monitor for secondary metabolites across all strains (AF293, CEA10, CEA17, CBS 139183T, NRRL 181, NRRL 4161, and NRRL 4585). Utilizing positive-ionization mode, ACD MS Manager with add-in software IntelliXtract (Advanced Chemistry Development, Inc.; Toronto, Canada) was used for the primary analysis of the UPLC-MS.
Aspergillus

Metabolomics analyses: Principal component analysis (PCA) analysis was performed on the UPLC-MS data. Untargeted UPLC-MS datasets for each sample were individually aligned, filtered, and analyzed using MZmine 2.20 software (https://sourceforge.net/projects/mzmine/) (Pluskal et al. 2010). Peak detection was achieved using the following parameters, A. fumigatus at (A293, CEA10, and CEA17): noise level (absolute value), $1 \times 10^6$; minimum peak duration, 0.05 min; $m/z$ variation tolerance, 0.05; and $m/z$ intensity variation, 20%; A. fischeri (NRRL 181, NRRL 4161, and NRRL 4585): noise level (absolute value), $1 \times 10^6$; minimum peak duration, 0.05 min; $m/z$ variation tolerance, 0.05; and $m/z$ intensity variation, 20%; and all strains (A293, CEA10, CEA17, CBS 139183T, NRRL 181, NRRL 4161, and NRRL 4585): noise level (absolute value), $7 \times 10^5$; minimum peak duration, 0.05 min; $m/z$ variation tolerance, 0.05; and $m/z$ intensity variation, 20%. Peak list filtering and retention time alignment algorithms were used to refine peak detection. The join algorithm integrated all sample profiles into a data matrix using the following parameters: $m/z$ and RT balance set at 0.5, $m/z$ tolerance set at 0.001, and RT tolerance set at 0.5 min. The resulting data matrix was exported to Excel (Microsoft) for analysis as a set of $m/z$–RT pairs with individual peak areas detected in triplicate analyses. Samples that did not possess detectable quantities of a given marker ion were assigned a peak area of zero to maintain the same number of variables for all sample sets. Ions that did not elute between 2 and 8 min and/or had an $m/z$ ratio $<200$ or $>800$ Da were removed from analysis. Relative SD was used to understand the quantity of variance between the technical replicate injections, which may differ slightly based on instrument variance. A cutoff of 1.0 was used at any given $m/z$–RT pair across the technical replicate injections of one biological replicate, and if the variance was greater than the cutoff, it was assigned a peak area of zero. Final chemometric analysis, data filtering (Caesar et al. 2018), and PCA was conducted using Sirius, v10.0 (Pattern Recognition Systems AS; Kvalheim et al. 2011), and dendrograms were created with Python. The PCA scores plots were generated using data from either the three individual biological replicates or the averaged biological replicates of the fermentations. Each biological replicate was plotted using averaged peak areas obtained across four replicate injections (technical replicates).

Data availability

The authors state that all necessary data for confirming the conclusions presented in the article are represented fully within the article. Sequence reads and associated genome assemblies generated in this project are available in NCBI’s GenBank database under the BioProject PRJNA577646. Additional descriptions of the genomes including predicted gene boundaries are available through figshare (https://doi.org/10.6084/m9.figshare.12055503). The figshare repository is also populated with other data generated from genomic and natural products analysis. Among genomic analyses, we provide information about predicted BGCs, results associated with network-based clustering of BGCs into cluster families, phylogenomic data matrices, and trees. Among natural products analysis, we provide information that supports methods and results, including NMR spectra.

Results

Conservation and diversity of biosynthetic gene clusters within and between species

We sequenced and assembled A. oerlinghausenensis CBS 139183T and A. fischeri strains NRRL 4585 and NRRL 4161. Together with publicly available genomes, we analyzed 10 Aspergillus genomes (five A. fischeri strains; four A. fumigatus strains; one A. oerlinghausenensis strain; see Materials and Methods). We found that the newly added genomes were of similar quality to other publicly available draft genomes.

Table 1 Species and strains used in the present study

| Genus and species         | Strain         | Environmental/Clinical | Genomic analysis | Secondary metabolite profiling | Reference               |
|--------------------------|----------------|------------------------|------------------|-------------------------------|-------------------------|
| Aspergillus oerlinghausenensis | CBS 139183T  | Environmental          | +                | +                             | This study              |
| Aspergillus fischeri     | NRRL 4585     | Environmental          | +                | +                             | This study              |
| A. fischeri              | NRRL 4161     | Unknown                | +                | +                             | This study              |
| A. fischeri              | NRRL 181      | Environmental          | +                | +                             | Fedorova et al. (2008)  |
| A. fischeri              | IBT 3007      | Environmental          | +                | –                             | Zhao et al. (2019)      |
| A. fischeri              | IBT 3003      | Environmental          | +                | –                             | Zhao et al. (2019)      |
| Aspergillus fumigatus    | AI293         | Clinical               | +                | +                             | Nierman et al. (2005)   |
| A. fumigatus             | CEA10/CEA17   | Clinical               | +                | +                             | Fedorova et al. (2008)  |
| A. fumigatus             | HMR AF270     | Clinical               | +                | –                             | BioSample: SAMN07177964 |
| A. fumigatus             | Z5            | Environmental          | +                | –                             | Miao et al. (2015)      |

“+” and “−” indicate if BGCs and secondary metabolite profiling was conducted on a particular strain. More specifically “+” indicates the strain was analyzed whereas “−” indicates that the strain was not analyzed.
(average percent presence of BUSCO genes: 98.80 ± 0.10%; average N50: 451,294.67 ± 9696.11; Figure S1). We predicted that *A. oerlinghausenensis* CBS 139183T, *A. fischeri* NRRL 4585, and *A. fischeri* NRRL 4161 have 10,044, 11,152 and 10,940 genes, respectively, numbers similar to publicly available genomes. Lastly, we inferred the evolutionary history of the 10 *Aspergillus* genomes using a concatenated matrix of 3041 genes (5,602,272 sites) and recapitulated species-level relationships as previously reported (Houbraken et al. 2016). Relaxed molecular clock analyses suggested *A. fumigatus, A. oerlinghausenensis*, and *A. fischeri* diverged from one another during the Neogene geologic period. Bipartition support is depicted for internodes that did not have full support. (B) *A. fumigatus* harbors the lowest number of BGCs compared to its two closest relatives. (C) Network-based clustering of BGCs into cluster families reveal extensive cluster presence and absence polymorphisms between species and strains. Cluster family identifiers are depicted on the x-axis; the number of strains represented in a cluster family are shown on the y-axis; the colors refer to a single strain from each species. Genus and species names are written using the following abbreviations: *Afum*, *A. fumigatus*; *Aoer*, *A. oerlinghausenensis*; *Afis*, *A. fischeri*. Classes of BGCs are written using the following abbreviations: NRPS, nonribosomal peptide synthetase; T1PKS, type I polyketide synthase; Hybrid, a combination of multiple BGC classes.

Examination of the total number of predicted BGCs revealed that *A. fischeri* has the largest BGC count. Among *A. fumigatus, A. oerlinghausenensis*, and *A. fischeri*, we predicted an average of 35.75 ± 2.22, 40, 50.80 ± 2.17 BGCs, respectively, and found they spanned diverse biosynthetic classes (e.g., polyketides, nonribosomal peptides, terpenes, etc.) (Figure 1B). Network-based clustering of BGCs into cluster families (or groups of homologous BGCs) resulted in qualitatively similar networks when we used moderate similarity thresholds (or edge cut-off values; Figure S3A). Using a (moderate) similarity threshold of 0.5, we inferred 88 cluster families of putatively homologous BGCs (Figure 1C).

Examination of BGCs revealed extensive presence and absence polymorphisms within and between species. We identified 17 BGCs that were present in all 10 *Aspergillus* genomes including the hexadehydroastechrome (HAS) BGC (cluster family 311 or CF311), the neosartoricin BGC (CF61), and other putative BGCs likely encoding unknown products (Figure S3B and Table S1; data available from figshare, https://doi.org/10.6084/m9.figshare.12055503). In contrast, we identified 18 BGCs found in single strains, which likely encode unknown products. Between species, similar patterns of broadly present and species-specific BGCs were observed. For example, we identified 18 BGCs that were present in at least one strain across all species; in contrast, *A. fumigatus, A. oerlinghausenensis*, and *A. fischeri* had 16, 8,
and 27 BGCs present in at least one strain but absent from the other species, respectively. These results suggest each species has a largely distinct repertoire of BGCs.

Examination of shared BGCs across species revealed A. oerlinghausenensis CBS139183T and A. fischeri shared more BGCs with each other than either did with A. fumigatus. Surprisingly, we found 10 homologous BGCs between A. oerlinghausenensis CBS 139183T and A. fischeri but only three homologous BGCs shared between A. fumigatus and A. oerlinghausenensis CBS 139183T (Figure 2A and Figure S3C) even though A. oerlinghausenensis is more closely related to A. fumigatus than to A. fischeri (Figure 1A). BGCs shared by A. oerlinghausenensis CBS 139183T and A. fischeri were uncharacterized while BGCs present in both A. fumigatus and A. oerlinghausenensis CBS 139183T included those that encode fumigaclavine and fumagillin/pseurotin. Lastly, to associate each BGC with a secondary metabolite in A. fumigatus Af293, we cross-referenced our list with a publicly available one (Table S2) (Lind et al. 2017). Importantly, all known A. fumigatus Af293 BGCs were represented in our analyses.

At the level of gene families, there were few species-specific gene families in A. oerlinghausenensis (Figure 2B). A. oerlinghausenensis CBS 139183T has only eight species-specific gene families, whereas A. fischeri and A. fumigatus have 1487 and 548 species-specific gene families, respectively. Examination of the best BLAST hits of the eight species-specific gene families suggest that most are hypothetical or uncharacterized fungal genes. To determine if the eight A. oerlinghausenensis CBS 139183T specific gene families were an artifact of using a single representative strain, we conducted and additional ortholog clustering analysis using a single strain of A. fischeri (NRRL 181), a single strain of A. fumigatus (Af293), or a single strain of each species (CBS 139183, NRRL 181, Af293). When using a single strain of A. fischeri or A. fumigatus, there were 23 or 6 gene families unique to each species, respectively. Therefore, the low number of A. oerlinghausenensis-specific gene families likely stems from our use of the genome of a single strain.

Despite a closer evolutionary relationship between A. oerlinghausenensis and A. fumigatus, we found A. oerlinghausenensis shares more gene families with A. fischeri than with A. fumigatus (685 and 109, respectively) suggestive of extensive gene loss in the A. fumigatus stem lineage. Lastly, we observed strain heterogeneity in gene family presence and absence within both A. fumigatus and A. fischeri (Figure S4). For example, the largest intersection that does not include all A. fischeri strains is 493 gene families, which were found in all but one strain, NRRL 181. For A. fumigatus, the largest intersection that does not include all strains is 233 gene families, which were shared by strains Af293 and CEA10.

Within- and between-species variation in secondary metabolite profiles of A. fumigatus and its closest relatives

To gain insight into variation in secondary metabolite profiles within and between species, we profiled A. fumigatus strains Af293, CEA10, and CEA17 (a pyrG1/URA3 derivative of CEA10), A. fischeri strains NRRL 181, NRRL 4585, and NRRL 4161, and A. oerlinghausenensis CBS 139183T for secondary metabolites. Specifically, we used three different procedures, including the isolation and structure elucidation of metabolites, where possible, followed by two different metabolite profiling procedures that use mass spectrometry techniques. Altogether, we isolated and characterized 19 secondary metabolites; 7 from A. fumigatus, 2 from A. oerlinghausenensis, and 10 from A. fischeri (Figure S5). These products encompassed a wide diversity of secondary metabolite classes, such as those derived from polyketide synthases, nonribosomal peptide-synthetases, terpene synthases, and mixed biosynthesis enzymes.

To characterize the secondary metabolites biosynthesized that were not produced in high enough quantity for structural identification through traditional isolation methods, we employed “dereplication” mass spectrometry protocols specific to natural products research on all tested strains at both 30° and 37° (see supporting information, dereplication example; figshare: https://doi.org/10.6084/m9.figshare.12055503) (El-Elmat et al. 2013; Ito and Masubuchi 2014; Gaudêncio and Pereira 2015; Hubert et al. 2017). We found that most secondary metabolites were present across strains of the same species (Table S3); for example, monomethylsulochrin was isolated from A. fumigatus Af293, but, through metabolite profiling, its spectral features were noted also in A. fumigatus strains CEA10 and CEA17. We identified metabolites that were biosynthesized...
by only one species; for example, pseurotin A was present solely in *A. fumigatus* strains. Finally, we found several secondary metabolites that were biosynthesized across species, such as fumagillin, which was biosynthesized by *A. fumigatus* and *A. oerlinghausenensis*, and fumitremorgin B, which was biosynthesized by strains of both *A. oerlinghausenensis* and *A. fischeri*.

Together, these analyses suggest that closely related *Aspergillus* species and strains exhibit variation within and between species in the secondary metabolites produced.

To further facilitate comparisons of secondary metabolite profiles within and between species, we used the 1920 features (i.e., unique m/z – retention time pairs) that were identified from all strains at all temperatures (Figure 3A), to perform hierarchical clustering (Figure 3B) and PCA (Figure S6). Hierarchical clustering at 37°C and 30°C indicated the chromatogram of *A. oerlinghausenensis* CBS 139183T is more similar to the chromatogram of *A. fischeri* than to that of *A. fumigatus*. PCA results were broadly consistent with the clustering results, but suggested that *A. oerlinghausenensis* was just as similar to *A. fischeri* strains as it was to *A. fumigatus* strains. This difference likely stems from the fact that hierarchical clustering is a total-evidence approach whereas PCA captures most but not all variance in the data (e.g., the two principal components in Figure S6B and S6C capture 84.6% of the total variance). PCA analysis revealed greater variation in secondary metabolite production at 30°C compared to 37°C (Figure S6), suggesting there is a more varied response in how BGCs are being utilized at 30°C. PCA at both 37°C and 30°C showed that variation between *A. oerlinghausenensis* CBS 139183T and *A. fischeri* strains was largely captured along the second principal component; in contrast, the differences between *A. oerlinghausenensis* CBS 139183T and *A. fumigatus* strains are captured along the first principal component (Figure S6, D and E). Taken together, these results suggest that the three *A. fischeri* strains and *A. oerlinghausenensis* were the most chemically similar to each other.

In summary, even though *A. oerlinghausenensis* is phylogenetically more closely related to *A. fumigatus* than to *A. fischeri* (Figure 1A), our chemical analyses suggest that the secondary metabolite profile of *A. oerlinghausenensis* is more similar to the profile of *A. fischeri* than it is to the profile of *A. fumigatus* (Figure 3B and Figure S6, B–E).
Gliotoxin

Inhibits host immune response

Sugui et al. (2007)

+/+ /+ /+ +/ +/ +/

Fumitremorgin

Inhibits the breast cancer resistance protein

González-Lobato et al. (2010)

+/- +/+ +/- +/+ +/+ +/+

Verruculogen

Changes electrophysical properties of human nasal epithelial cells

Khoufache et al. (2007)

+/- +/+ +/- +/+ +/+ +/+

Trypacidin

Damages lung cell tissues

Gauthier et al. (2012)

+/+ +/+ +/- +/+ +/- +/- +/- +/-

Pseurotin

Inhibits immunoglobulin E

Ishikawa et al. (2009)

+/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+ +/+

Fumitremorgin and verruculogen: Similarly, there is a high degree of conservation in the BGC that encodes fumitremorgin across all strains (Figure 5). Fumitremorgins have known antifungal activity, are lethal to brine shrimp, and are implicated in inhibiting mammalian proteins responsible for resistance to anticancer drugs in mammalian cells (Raffa and Keller 2019). We found that conservation in the fumitremorgin BGC is associated with the production of fumitremorgins in all isolates examined. The fumitremorgin BGC is also responsible for the production of verruculogen, which is implicated to aid in A. fumigatus pathogenicity by changing the electrophysical properties of human nasal epithelial cells (Khoufache et al. 2007). Interestingly, we found that only A. fischeri strains produced verruculogen under the conditions we analyzed.

Trypacidin: Examination of the trypacidin BGC, which encodes a spore-borne and cytotoxic secondary metabolite, revealed a conserved cluster found in four pathogenic and nonpathogenic strains: A. fumigatus AF293, A. fumigatus CEA10, A. oerlinghausenensis CBS 139183 T, and A. fischeri NRRL 181 (Figure S7). Furthermore, we found that three of these four isolates (except A. fischeri NRRL 181) biosynthesized a trypacidin analog, monomethylsulochrin. Examination of the microsynteny of the trypacidin BGC revealed...
that it was conserved across all four genomes with the exception of A. fischeri NRRL 181, which lacked a RING (Really Interesting New Gene) finger gene. Interestingly, RING finger proteins can mediate gene transcription (Poukka et al. 2000). We confirmed the absence of the RING finger protein by performing a sequence similarity search with the A. fumigatus Af293 RING finger protein (AFUA_4G14620; EAL89333.1) against the A. fischeri NRRL 181 genome. In the homologous locus in A. fischeri, we found no significant BLAST hit for the first 23 nucleotides of the RING finger gene suggestive of pseudogenization. Taken together, we hypothesize that presence/absence polymorphisms or a small degree of sequence divergence between otherwise homologous BGCs may be responsible for the presence or absence of a toxic secondary metabolite in A. fischeri NRRL 181. Furthermore, inter- and intra-species patterns of trypacidin presence and absence highlight the importance of strain heterogeneity when examining BGCs.

**Fumagillin/pseurotin:** Examination of the intertwined fumagillin/pseurotin BGCs revealed that fumagillin has undergone substantial sequence divergence, and that pseurotin is absent from strains of A. fischeri. The fumagillin/pseurotin BGCs are under the same regulatory control (Wiemann et al. 2013) and biosynthesize secondary metabolites that cause cellular damage during host infection (fumagillin; Guruceaga et al. 2019) and inhibit immunoglobulin E production (pseurotin; Ishikawa et al. 2009). Microsynteny of the fumagillin BGC reveals high sequence conservation between A. fumigatus and A. oerlinghausenensis; however, sequence divergence was observed between A. oerlinghausenensis and A. fischeri (Figure 5). Accordingly, fumagillin production was only observed in A. fumigatus and A. oerlinghausenensis and not in A. fischeri. Similarly, the pseurotin BGC is conserved between A. fumigatus and A. oerlinghausenensis. Rather than sequence divergence, no sequence similarity was observed in the region of the pseurotin cluster in A. fischeri, which may be due to an indel event. Accordingly, no pseurotin production was observed among A. fischeri strains. Despite sequence conservation between A. fumigatus and A. oerlinghausenensis, no evidence of pseurotin biosynthesis was observed in A. oerlinghausenensis, which suggests regulatory decoupling of the intertwined fumagillin/pseurotin BGC. Alternatively, the genes downstream of the A. fumigatus pseurotin BGC, which are absent from the A. oerlinghausenensis locus, may contribute to BGC production and could explain the lack of pseurotin production in A. oerlinghausenensis. Altogether, these results show a striking correlation between sequence divergence and the production (or absence) of secondary metabolites implicated in human disease among A. fumigatus and nonpathogenic closest relatives.

**Discussion**

Aspergillus fumigatus is a major fungal pathogen nested within a clade (known as section *Fumigati*) of at least 60 other species, the vast majority of which are nonpathogenic (Steenwyk et al. 2019; Rokas et al. 2020b). Currently, it is thought that the ability to cause human disease evolved multiple times among species in section *Fumigati* (Rokas et al. 2020b). Secondary metabolites contribute to the success of
the major human pathogen *A. fumigatus* in the host environment (Raffa and Keller 2019) and can therefore be thought of as “cards” of virulence (Casadevall 2007; Knowles et al. 2020). However, whether the closest relatives of *A. fumigatus*, *A. oerlinghausenensis*, and *A. fischeri*, both of which are nonpathogenic, biosynthesize secondary metabolites implicated in the ability of *A. fumigatus* to cause human disease remained largely unknown. By examining genomic and chemical variation between and within *A. fumigatus* and its closest nonpathogenic relatives, we identified both conservation and divergence (including within species heterogeneity) in BGCs and secondary metabolite profiles (Figures 1–5, Figures S3 and S5–S8, Table 2, and Table S1 and S3). Examples of conserved BGCs and secondary metabolites include the major virulence factor, gliotoxin (Figure 4), as well as several others (Figure 5 and Figure S7; Table 2 and Table S1 and S3); examples of BGC and secondary metabolite heterogeneity or divergence include pseurotin, fumagillin, and several others (Figure 5, Table 2, and Table S1 and S3). Lastly, we found that the fumitremorgin BGC, which biosynthesizes fumitremorgin in all three species, is also associated with verruculogen biosynthesis in *A. fischeri* strains (Figure 5).

One of the surprising findings of our study was that although *A. oerlinghausenensis* and *A. fumigatus* are evolutionarily more closely related to each other than to *A. fischeri* (Figure 1), *A. oerlinghausenensis* and *A. fischeri* appear to be more similar to each other than to *A. fumigatus* in BGC composition, gene family content, and secondary metabolite profiles. The power of pathogen–nonpathogen comparative genomics is best utilized when examining closely related species (Fedorova et al. 2008; Jackson et al. 2011; Moran et al. 2011; Mead et al. 2019a; Rokas et al. 2020b). Genomes from additional strains from the closest known nonpathogenic relatives of *A. fumigatus*, including from the closest species relative *A. oerlinghausenensis*, *A. fischeri*, and other nonpathogenic species in section *Fumigati* will be key for understanding the evolution of *A. fumigatus* pathogenicity.

Our finding that *A. oerlinghausenensis* and *A. fischeri* shares more gene families and BGCs with each other than they do with *A. fumigatus* (Figures 1C and 2 and Figures S3, S4, and S8) suggests that the evolutionary trajectory of the *A. fumigatus* ancestor was marked by gene loss. We hypothesize that there were two rounds of gene family and BGC loss in the *A. fumigatus* stem lineage: (1) gene families and BGCs were lost in the common ancestor of *A. fumigatus* and *A. oerlinghausenensis*, and (2) additional losses occurred in the *A. fumigatus* ancestor. In addition to losses, we note that S48 gene families and 16 BGCs are unique to *A. fumigatus*, which may have resulted from genetic innovation (e.g., *de novo* gene formation) or unique gene family and BGC retention (Figure 2 and Figure S8). In line with the larger number of shared BGCs between *A. oerlinghausenensis* and *A. fischeri*, we found their secondary metabolite profiles were also more similar (Figure 3 and Figure S6). Notably, the evolutionary rate of the internal branch leading to the *A. fumigatus* common ancestor is much higher than those in the rest of the branches in our genome-scale phylogeny (Figure S2B), suggesting that the observed gene loss and gene gain/retention events specific to *A. fumigatus* may be part of a wider set of evolutionary changes in the *A. fumigatus* genome. Analyses
with a greater number of strains and species will help further test the validity of this hypothesis. More broadly, these results suggest that comparisons of the pathogen *A. fumigatus* against either the nonpathogen *A. oerlinghausenensis* (this manuscript) or the nonpathogen *A. fischeri* (Mead et al. 2019a; Knowles et al. 2020; and this manuscript) will both be instructive in understanding the evolution of *A. fumigatus* pathogenicity.

When studying *Aspergillus* pathogenicity, it is important to consider any genetic and phenotypic heterogeneity between strains of a single species (Knox et al. 2016; Kowalski et al. 2016, 2019; Keller 2017; Ries et al. 2019; Blachowicz et al. 2020; Bastos et al. 2020; Drott et al. 2020; dos Santos et al. 2020; Steenwyk et al. 2020). Our finding of strain heterogeneity among gene families, BGCs, and secondary metabolites in *A. fumigatus* and *A. fischeri* (Figures 1–3 and Figures S3, S4, S6, and S8) suggests considerable strain-level diversity in each species. For example, we found secondary metabolite profile strain heterogeneity was greater in *A. fumigatus* than *A. fischeri* (Figure S6, B–E). These results suggest that strain-specific secondary metabolite profiles may play a role in variation of pathogenicity among *A. fumigatus* strains. In support of this hypothesis, differential secondary metabolite production has been associated with differences in virulence among isolates of *A. fumigatus* (Blachowicz et al. 2020). More broadly, our finding supports the hypothesis that strain-level diversity is an important parameter when studying pathogenicity (Kowalski et al. 2016, 2019; Keller 2017; Ries et al. 2019; Blachowicz et al. 2020; Bastos et al. 2020; Drott et al. 2020; dos Santos et al. 2020; Steenwyk et al. 2020).

Secondary metabolites contribute to *A. fumigatus* virulence through diverse processes including suppressing the human immune system and damaging tissues (Table 2). Interestingly, we found that the nonpathogens *A. oerlinghausenensis* and *A. fischeri* produced several secondary metabolites implicated in the ability of *A. fumigatus* human disease, such gliotoxin, tryptacin, verruculogen, and others (Figures 4 and 5, Figure S7, Table 2 and Table S3). Importantly, our work positively identified secondary metabolites for many structural classes implicated in a previous taxonomic study (Samson et al. 2007). These results suggest that several of the secondary metabolism-associated cards of virulence present in *A. fumigatus* are conserved in closely related nonpathogens (summarized in Figure 6) as well as in closely related pathogenic species, such as *A. novofumigatus* (Kjærboell et al. 2018). Interestingly, disrupting the ability of *A. fumigatus* to biosynthesize gliotoxin attenuates but does not abolish virulence (Sugui et al. 2007; Dagenais and Keller 2009; Keller 2017), whereas disruption of the ability of *A. fischeri* NRRL 181 to biosynthesize secondary metabolites, including gliotoxin, does not appear to influence virulence (Knowles et al. 2020). Our findings, together with previous studies, support the hypothesis that individual secondary metabolites are “cards” of virulence in a larger “hand” that *A. fumigatus* possesses.

**Acknowledgments**

We thank the Rokas, Oberlies, and Goldman laboratories for helpful discussion and support of this work. J.L.S. and A.R. are supported by the Howard Hughes Medical Institute through the James H. Gilliam Fellowships for Advanced
sensing by cyclic dipeptides. ACS Chem. Biol. 4: 1051–1059. https://doi.org/10.1021/cb900165y
Capella-Gutierrez, S., J. M. Silla-Martinez, and T. Gabaldon, 2009 trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25: 1972–1973. https://doi.org/10.1093/bioinformatics/btp348
Casadevall, A., 2007 Determinants of virulence in the pathogenic fungi. Fungal Biol. Rev. 21: 130–132. https://doi.org/10.1016/j.fbr.2007.02.007
Cock, P. J. A., T. Antao, J. T. Chang, B. A. Chapman, C. J. Cox et al., 2009 Biopython: freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics 25: 1422–1423. https://doi.org/10.1093/bioinformatics/btp163
Dagenais, T. R., T., and N. P. Keller, 2009 Pathogenesis of Aspergillus fumigatus in invasive aspergillosis. Clin. Microbiol. Rev. 22: 447–465. https://doi.org/10.1128/CMR.00055-08
de Vries, R. P., R. Riley, A. Wiebenga, G. Aguilar-Osorio, S. Amillis et al., 2017 Comparative genomics reveals high biological diversity and specific adaptations in the industrially and medically important fungal genus Aspergillus. Genome Biol. 18: 28. https://doi.org/10.1186/s13059-017-1151-0
Dolan, S. K., R. A. Owens, G. O‘Keeffe, S. Hammel, D. A. Fitzpatrick et al., 2014 Regulation of nonribosomal peptide synthesis: bis-thiomethylolation attenuates gliotoxin biosynthesis in Aspergillus fumigatus. Chem. Biol. 21: 999–1012. https://doi.org/10.1016/j.jchembiol.2014.07.006
dos Santos, R. A. C., J. L. Steenwyk, O. Rivero-Menendez, M. E. Mead, L. P. Silva et al., 2020 Genomic and phenotypic heterogeneity of clinical isolates of the human pathogens Aspergillus fumigatus, Aspergillus lentinus, and Aspergillus fumigatusfumigatus. Front. Genet. 11: 459. https://doi.org/10.3389/fgene.2020.00459
Drgona, L., A. Khachatryan, J. Stephens, C. Charbonneau, M. Kantecki et al., 2014 Clinical and economic burden of invasive fungal diseases in Europe: focus on pre-emptive and empirical treatment of Aspergillus and Candida species. Eur. J. Clin. Microbiol. Infect. Dis. 33: 7–21. https://doi.org/10.1007/s10096-013-1944-3
Drott, M. T., R. W. Bastos, A. Rokas, L. N. A. Ries, T. Gabaldón et al., 2020 Diversity of Secondary Metabolism in Aspergillus nidulans Clinical Isolates. mSphere 5: e00156-20. https://doi.org/10.1128/mSphere.00156-20
El-Elimat, T., M. Figueroa, B. M. Ehrmann, N. B. Cech, C. J. Pearce et al., 2013 High-resolution MS, MS/MS, and UV database of fungal secondary metabolites as a dereplication protocol for bioactive natural products. J. Nat. Prod. 76: 1709–1716. https://doi.org/10.1021/np4004307
Emms, D. M., and S. Kelly, 2019 OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. 20: 238. https://doi.org/10.1186/s13059-019-1832-y
Fallon, J. P., E. P. Reeves, and K. Kavanagh, 2010 Inhibition of neutrophil function following exposure to the Aspergillus fumigatus toxin fumagillin. J. Med. Microbiol. 59: 625–633. https://doi.org/10.1099/jmm.0.018192-0
Fallon, J. P., E. P. Reeves, and K. Kavanagh, 2011 The Aspergillus fumigatus toxin fumagillin suppresses the immune response of Galleria mellonella larvae by inhibiting the action of haemocytes. Microbiology 157: 1481–1488. https://doi.org/10.1099/mic.0.043786-0
Fedorova, N. D., N. Khalidi, V. S. Joardar, R. Maiti, P. Amedeo et al., 2008 Genomic islands in the pathogenic filamentous fungus Aspergillus fumigatus. PLoS Genet. 4: e1000046. https://doi.org/10.1371/journal.pgen.1000046
Gaudêncio, S. P., and F. Pereira, 2015 Dereplication: racing to speed up the natural products discovery process. Nat. Prod. Rep. 32: 779–810. https://doi.org/10.1039/C4NP00134F
Gauthier, T., X. Wang, J. Sifuentes Dos Santos, A. Fysikopoulos, S. Tadríst et al., 2012 Trypacinid, a spore-borne toxin from
Aspergillus fumigatus, is cytotoxic to lung cells. PLoS One 7: e29906. https://doi.org/10.1371/journal.pone.0029906

González-Lobato, L., R. G. Prieto, A. I. Álvarez, and G. Merino, 2010 Differential inhibition of murine Bcrp1/Abcg2 and human BCRP/ABCG2 by the mycotoxin fumitremorgin C. Eur. J. Pharmacol. 644: 41–48. https://doi.org/10.1016/j.ejphar.2010.07.016

Graham, N., K. M. Shepardson, D. Chung, and R. A. Cramer, 2012 Hypoxia and fungal pathogenesis: to air or not to air? Eukaryot. Cell 11: 560–570. https://doi.org/10.1128/EC.00031-12

Guruceaga, X., G. Ezepeleta, E. Mayayo, M. Sueiro-Olivares, A. Abad-Diaz-De-Cerio et al., 2018 A possible role for fumagillin in cellular damage during host infection by Aspergillus fumigatus. Virulence 9: 1548–1561. https://doi.org/10.1080/21505594.2018.1526528

Guruceaga, X., U. Perez-Cuesta, A. Abad-Diaz de Cerio, O. Gonzalez, R. M. Alonso et al., 2019 Fumagillin, a mycotoxin of Aspergillus fumigatus: biosynthesis, biological activities, detection, and applications. Toxins (Basel) 12: 7. https://doi.org/10.3390/toxins12010007

Halász, J. B., B. Podányi, L. Vasvári-Debreczy, A. Szabó, F. Hajdú and L. S. Jermiin, 2017 ModelFinder: fast model selection for accurate phylogenetic estimates. Nat. Methods 14: 587–589. https://doi.org/10.1038/nmeth.4285

Holt, C., and M. Yandell, 2011 MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. BMC Bioinformatics 12: 491. https://doi.org/10.1186/1471-2105-12-491

Houbrenken, J., M. Weig, U. Groß, M. Meijer, and O. Bader, 2020 Characterization of the orthologous gene set of Aspergillus fumigatus. mSphere.00227-16. https://doi.org/10.1128/mSphere.00227-16

Hubert, J., J.-M. Nuzillard, and J.-H. Renault, 2017 Dereplication strategies in natural product research: how many tools and methodologies behind the same concept? Phytochem. Rev. 16: 55–95. https://doi.org/10.1007/s11101-015-9448-7

Ito, T., and M. Masubuchi, 2014 Dereplication of microbial extracts and related analytical technologies. J. Antibiot. (Tokyo) 67: 353–360. https://doi.org/10.1093/ja/2014.12

Jackson, R. W., L. J. Johnson, S. R. Clarke, and D. L. Arnold, 2011 Bacterial pathogen evolution: breaking news. Trends Genet. 27: 32–40. https://doi.org/10.1016/j.tig.2010.01.001

Kalyaanamoorthy, S., B. Q. Minh, T. K. F. Wong, A. von Haeseler, and R. W. Jackson, R. W., L. J. Johnson, S. R. Clarke, and D. L. Arnold, 2010 Differential inhibition of murine Bcrp1/Abcg2 and human BCRP/ABCG2 by the mycotoxin fumitremorgin C. Eur. J. Pharmacol. 644: 41–48. https://doi.org/10.1016/j.ejphar.2010.07.016

Kato, N., H. Suzuki, H. Takagi, Y. Asami, H. Kakeya et al., 2009 Identification of cytochrome P450s required for fumitremorgin biosynthesis in Aspergillus fumigatus. ChemBioChem 10: 920–928. https://doi.org/10.1002/cbic.200800787

Katsur, S. A., K. Blin, S. Shaw, J. C. Navarro-Muñoz, B. R. Terlouw et al., 2020 MBiG 2.0: a repository for biosynthetic gene clusters of known function. Nucleic Acids Res. 8: D454–D458. https://doi.org/10.1093/nar/gkz882

Keller, N. P., 2017 Heterogeneity confounds establishment of “a” model microbial strain. MBio 8: e00135-17. https://doi.org/10.1128/mBio.00135-17

Keller, N. P., 2019 Fungal secondary metabolism: regulation, function and drug discovery. Nat. Rev. Microbiol. 17: 167–180. https://doi.org/10.1038/s41579-018-0121-1

Khoufache, K., O. Puel, N. Loiseau, M. Delaforge, D. Rivollet et al., 2007 Verruculogen associated with Aspergillus fumigatus hyphae and conidia modifies the electrophysiological properties of human nasal epithelial cells. MBio Microbiol. 7: 5. https://doi.org/10.1186/1471-2180-7-5

Kjærboell, I., T. L. Vesth, J. C. Frisvad, J. L. Nybo, S. Theobald et al., 2018 Linking secondary metabolites to gene clusters through genome sequencing of six diverse Aspergillus species. Proc. Natl. Acad. Sci. USA 115: E753–E761. https://doi.org/10.1073/pnas.1715954115

Kjærboell, I., T. Vesth, J. C. Frisvad, J. L. Nybo, S. Theobald et al., 2020 A comparative genomics study of 23 Aspergillus species from section Flav. Nat. Commun. 11: 1106. https://doi.org/10.1038/s41467-019-14051-y

Knowles, S. L., N. Vu, D. A. Todd, H. A. Raja, A. Rokas et al., 2019 Orthogonal method for double-bond placement via ozone-induced dissociation mass spectrometry (OzID-MS). J. Nat. Prod. 82: 3421–3431. https://doi.org/10.1021/acs.jnatprod.9b00778

Knowles, S. L., M. E. Mead, L. P. Silva, H. A. Raja, J. L. Steenwyk et al., 2020 Gliotoxin, a known virulence factor in the major human pathogen Aspergillus fumigatus, is also biosynthesized by its nonpathogenic relative Aspergillus fischeri. MBio. 11: e03361-19. https://doi.org/10.1128/mBio.03361-19

Knox, B. P., A. Blachowicz, J. M. Palmer, J. Romsdahl, A. Huttenlocher et al., 2016 Characterization of Aspergillus fumigatus isolates from Air and Surfaces of the International Space Station. mSphere 1: e00227-16. https://doi.org/10.1128/mSphere.00227-16

Korf, I., 2004 Gene finding in novel genomes. BMC Bioinformatics 5: 59. https://doi.org/10.1186/1471-2105-5-59

Kowalski, C. H., S. R. Beattie, K. K. Fuller, E. A. McGurk, Y.-W. Tang et al., 2019 Orthogonal method for double-bond placement via ozone-induced dissociation mass spectrometry (OzID-MS). MBio 7: e01515-16. https://doi.org/10.1128/mBio.01515-16

Kowalski, C. H., J. D. Kerkaert, K.-W. Liu, M. C. Bond, R. Hartmann et al., 2019 Fungal biofilm morphology impacts hypoxia fitness and disease progression. Nat. Microbiol. 4: 2430–2441. https://doi.org/10.1038/s41467-019-14058-7

Kralj, J.-P., and G. Chamilo, 2019 Aspergillus fumigatus and aspergillosis in 2019. Clin. Microbiol. Rev. 33: e00140-18. https://doi.org/10.1128/cmrr.00140-18

Kvalheim, O. M., H. Chan, I. F. F. Benzie, Y. Szeto, A. H. Tzang et al., 2011 Chromatographic profiling and multivariate analysis for screening and quantifying the contributions from individual components to the bioactive signature in natural products. Chemom. Intell. Lab. Syst. 107: 98–105. https://doi.org/10.1016/j.chemolab.2011.02.002

Lagó, J.-P., and G. Chamilo, 2019 Aspergillus fumigatus and aspergillosis in 2019. Clin. Microbiol. Rev. 33: e00140-18. https://doi.org/10.1128/cmrr.00140-18

Lind, A. L., J. H. Wiseacre, T. D. Smith, X. Feng, A. M. Calvo et al., 2015 Examining the evolution of the regulatory circuit controlling secondary metabolism and development in the fungal genus Aspergillus. PLoS Genet. 11: e1005096. https://doi.org/10.1371/journal.pgen.1005096

Lind, A. L., J. H. Wiseacre, C. Lameiras, P. Wiemann, J. M. Palmer et al., 2017 Drivers of genetic diversity in secondary metabolic
Tekaia, F., and J.-P. Latgé, 2005 Aspergillus fumigatus: saprophyte or pathogen? Curr. Opin. Microbiol. 8: 385–392. https://doi.org/10.1016/j.mib.2005.06.017

Vallabhaneni, S., R. K. Mody, T. Walker, and T. Chiller, 2016 The global burden of fungal diseases. Infect. Dis. Clin. North Am. 30: 1–11. https://doi.org/10.1016/j.idc.2015.10.004

van Dongen, S., 2000 Graph clustering by flow simulation. Graph Stimul. by flow Clust. PhD thesis: University of Utrecht, Netherlands. https://doi.org/10.1016/j.cosrev.2007.05.001

Vesth, T. C., J. L. Nybo, S. Theobald, J. C. Frisvad, T. O. Larsen et al., 2018 Investigation of inter- and intraspecies variation through genome sequencing of Aspergillus section Nigri. Nat. Genet. 50: 1688–1695. https://doi.org/10.1038/s41588-018-0246-1

Vinet, L., and A. Zhedanov, 2011 A ‘missing’ family of classical orthogonal polynomials. J. Phys. A Math. Theor. 44: 085201. https://doi.org/10.1088/1751-8113/44/8/085201

Wang, F., Y. Fang, T. Zhu, M. Zhang, A. Lin et al., 2008 Seven new prenylated indole diketopiperazine alkaloids from holothurian-derived fungus Aspergillus fumigatus. Tetrahedron 64: 7986–7991. https://doi.org/10.1016/j.tet.2008.06.013

Yin, W.-B., J. A. Baccile, J. W. Bok, Y. Chen, N. P. Keller et al., 2013 A nonribosomal peptide synthetase-derived iron(III) complex from the pathogenic fungus Aspergillus fumigatus. J. Am. Chem. Soc. 135: 2064–2067. https://doi.org/10.1021/ja311145n

Variation in Secondary Metabolism