bioRxiv preprint doi: https://doi.org/10.1101/2020.04.09.033902; this version posted April 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Biosynthetic gene clusters, secondary metabolite profiles, and cards of virulence in the closest nonpathogenic relatives of *Aspergillus fumigatus*

Jacob L. Steenwyk[1], Matthew E. Mead[1,+], Sonja L. Knowles[2,+], Huzefa A. Raja[2], Christopher D. Roberts[2], Oliver Bader[3], Jos Houbraken[4], Gustavo H. Goldman[5], Nicholas H. Oberlies[2], Antonis Rokas[1,*]

[1] Department of Biological Sciences, Vanderbilt University, Nashville, TN, 37235, USA

[2] Department of Chemistry and Biochemistry, University of North Carolina at Greensboro, Greensboro, NC, USA

[3] Institute for Medical Microbiology, University Medical Center Göttingen, Göttingen, Germany

[4] Westerdijk Fungal Biodiversity Institute, Uppsalalaan 8, 3584 CT Utrecht, the Netherlands

[5] Faculdade de Ciências Farmacêuticas de Ribeirão Preto, Universidade de São Paulo, São Paulo, Brazil

[+] M. E. M. and S. L. K. contributed equally to this work.

[*] Correspondence: antonis.rokas@vanderbilt.edu

Running title: Secondary metabolism in *A. fumigatus* and its relatives

Keywords: secondary metabolites, specialized metabolism, gliotoxin, chemodiversity, mycotoxin, biosynthetic gene cluster, cards of virulence, pathogenicity,

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.09.033902; this version posted April 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Abstract

Aspergillus fumigatus is a major human pathogen that causes hundreds of thousands of infections yearly with high mortality rates. In contrast, *Aspergillus fischeri* and the recently described *Aspergillus oerlinghausenensis*, the two species most closely related to *A. fumigatus*, are not known to be pathogenic. Some of the "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. Secondary metabolites and the biosynthetic gene clusters (BGCs) that typically encode them often vary within and between fungal species. To gain insight into whether secondary metabolism-associated cards of virulence vary between *A. fumigatus*, *A. oerlinghausenensis*, and *A. fischeri*, we conducted extensive genomic and secondary metabolite profiling analyses. By analyzing multiple *A. fumigatus*, one *A. oerlinghausenensis*, and multiple *A. fischeri* strains, we identified both conserved and diverged secondary metabolism-associated cards of virulence. For example, we found that all species and strains examined biosynthesized the major virulence factor gliotoxin, consistent with the conservation of the gliotoxin BGC across genomes. However, species differed in their biosynthesis of fumagillin and pseurotin, both contributors to host tissue damage during invasive aspergillosis; these 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 into the genetic and phenotypic changes associated with the evolution of fungal pathogenicity.

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.09.033902; this version posted April 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Importance

The major fungal pathogen *Aspergillus fumigatus* kills tens of thousands each year. In contrast, the two closest relatives of *A. fumigatus*, namely *Aspergillus fischeri* and *Aspergillus oerlinghausenensis*, are not considered pathogenic. *A. fumigatus* virulence stems, partly, from its ability to produce small molecules called secondary metabolites that have potent activities during infection. In this study, we examined whether *A. fumigatus* secondary metabolites and the metabolic pathways involved in their production are conserved in *A. oerlinghausenensis* and *A. fischeri*. We found that the nonpathogenic close relatives of *A. fumigatus* produce some, but not all, secondary metabolites thought to contribute to the success of *A. fumigatus* in causing human disease and that these similarities and differences were reflected in the underlying metabolic pathways involved in their biosynthesis. Compared to its nonpathogenic close relatives, *A. fumigatus* produces a distinct cocktail of secondary metabolites, which likely contributes to these organisms' vastly different potentials to cause human disease. More broadly, the study of nonpathogenic organisms that have virulence-related traits, but are not currently considered agents of human disease, may facilitate the prediction of species capable of posing future threats to human health.

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.09.033902; this version posted April 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Introduction

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 more than 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., 2020a). Numerous studies have identified factors 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; Tekaia and Latgé, 2005; Abad et al., 2010; Grahl et al., 2012). Factors 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 factors are typically insufficient to cause disease but can collectively do so (Casadevall, 2007).

Several secondary metabolites biosynthesized by *A. fumigatus* are known "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 (Raffa and Keller, 2019). For example, the secondary metabolite gliotoxin has been shown to contribute to *A. fumigatus* virulence by inhibiting the host immune response (Sugui et al., 2007). Other secondary metabolites implicated

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.09.033902; this version posted April 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

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 (Khoufache et al., 2007); trypacidin, which is cytotoxic to lung cells (Gauthier et al., 2012); pseurotin, 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) (Table 1).

By extension, the metabolic pathways responsible for the biosynthesis of secondary metabolites are also "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; Rokas et al., 2018, 2020b). For example, even though *A. fumigatus* contains 33 BGCs and *A. 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 toxic virulence factor (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 has been shown be an important factor when studying *Aspergillus* pathogenicity

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.09.033902; this version posted April 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

(Kowalski et al., 2016, 2019; Keller, 2017; Ries et al., 2019; Bastos et al., 2020; Santos et al., 2020). Examination of multiple strains of *A. fumigatus* and close relatives—including the recently described and largely uncharacterized (in the context of pathogenicity) closest known relative of *A. fumigatus*, *A. oerlinghausenensis* (Houbraken et al., 2016)—will increase our understanding of the *A. fumigatus* secondary metabolite "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 139183^{T} 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 metabolite profiles of three *A. fumigatus*, one *A. oerlinghausenensis*, and three *A. fischeri* strains. On the one hand, 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 (Khoufache et al., 2007). On the other hand, we found that both *A. fumigatus* and *A. oerlinghausenensis* biosynthesized fumagillin and trypacidin, 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. These results reveal that nonpathogenic close relatives of *A. fumigatus* also produce some,

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.09.033902; this version posted April 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

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.

Results

Conservation and diversity of biosynthetic gene clusters within and between species

We sequenced and assembled *A. oerlinghausenensis* CBS 139183^T 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 Methods). We found that the newly added genomes were of similar quality to other publicly available draft genomes (average percent presence of BUSCO genes: 98.80 ± 0.10%; average N50: 451,294.67 ± 9,696.11; Fig. S1). We predicted that *A. oerlinghausenensis* CBS 139183^T, *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 3,041 genes (5,602,272 sites) and recapitulated species-level relationships as previously reported (Houbraken et al., 2016). Relaxed molecular clock analyses suggested that *A. oerlinghausenensis* CBS 139183^T diverged from *A. fumigatus* approximately 3.9 (6.4 – 1.3) million years ago and that *A. oerlinghausenensis* and *A. fumigatus* split from *A. fischeri* approximately 4.5 (6.8 – 1.7) million years ago (Fig. 1A; Fig. S2).

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.09.033902; this version posted April 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Examination of the total number of predicted BGCs revealed that *A. fischeri* has the largest BGC repertoire. Among *A. fumigatus*, *A. oerlinghausenensis*, and *A. fischeri*, we predicted 50.80 ± 2.17, 40, 35.75 ± 2.22 BGCs, respectively, and found they spanned diverse biosynthetic classes (e.g., polyketides, non-ribosomal peptides, terpenes, etc.) (Fig. 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; Fig. S3). Using a (moderate) similarity threshold of 0.5, we inferred 88 cluster families of putatively homologous BGCs (Fig. 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 (Fig. S4A; Table S1). 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, eight, 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.

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.09.033902; this version posted April 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

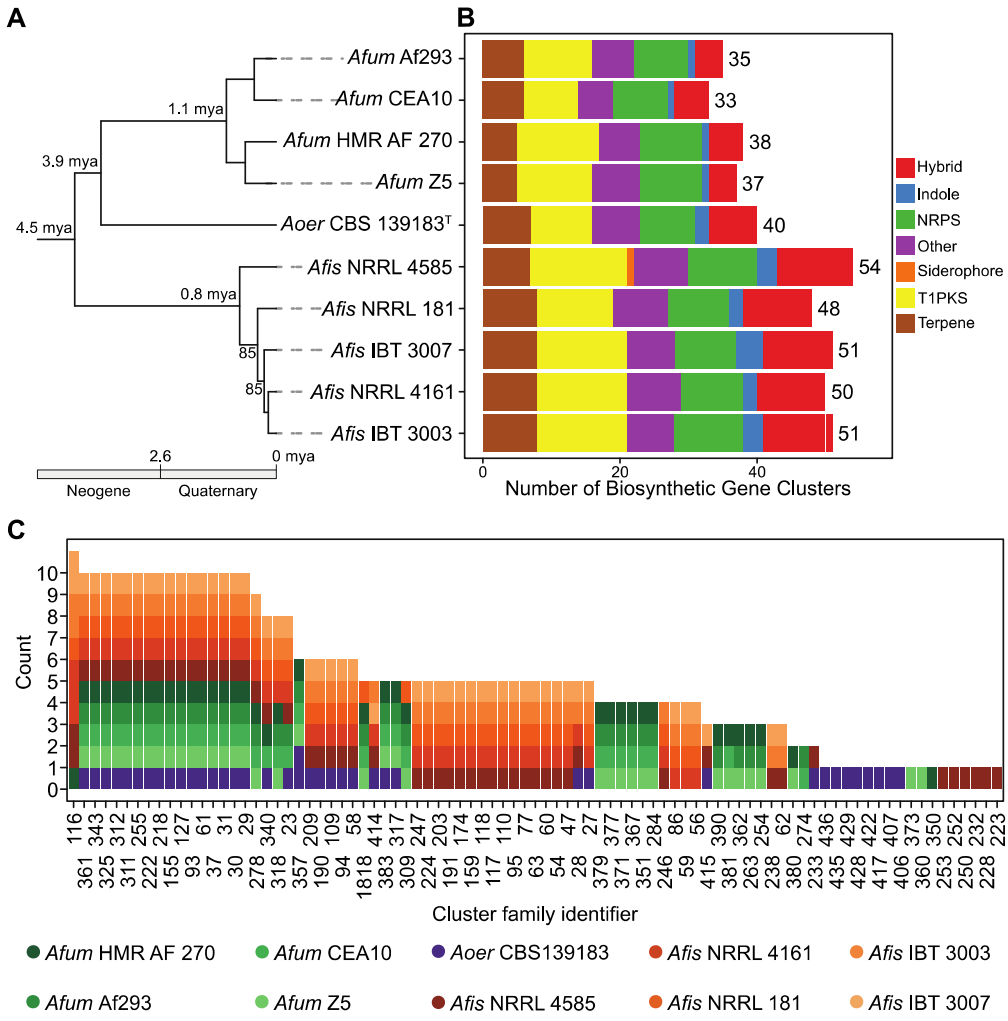


Figure 1. Diverse genetic repertoire of biosynthetic gene clusters and extensive presence and absence polymorphisms between and within species. (A) Genome-scale phylogenomic confirms *A. oerlinghausenensis* is the closest relative to *A. fumigatus*. Relaxed molecular clock analyses suggest *A. fumigatus, 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*;

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.09.033902; this version posted April 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

183 *Aoer*: *A. oerlinghausenensis*; *Afis*: *A. fischeri*.

184

185

186 Examination of shared BGCs across species revealed *A. oerlinghausenensis* CBS139183^T and *A.*
187 *fischeri* shared more BGCs with each other than either did with *A. fumigatus*. Surprisingly, we
188 found ten homologous BGCs between *A. oerlinghausenensis* CBS 139183^T and *A. fischeri* but
189 only three homologous BGCs shared between *A. fumigatus* and *A. oerlinghausenensis* CBS
190 139183^T (Fig. 2A; Fig. S4B) even though *A. oerlinghausenensis* is more closely related to *A.*
191 *fumigatus* than to *A. fischeri* (Fig. 1A). BGCs shared by *A. oerlinghausenensis* CBS 139183^T and
192 *A. fischeri* were uncharacterized while BGCs present in both *A. fumigatus* and *A.*
193 *oerlinghausenensis* CBS 139183^T included those that encode fumigaclavine and
194 fumagillin/pseurotin. Lastly, to associate each BGC with a secondary metabolite in *A. fumigatus*
195 Af293, we cross referenced our list with a publicly available one (Table S2). Importantly, all
196 known *A. fumigatus* Af293 BGCs were represented in our analyses.

197

198 More broadly, examination of shared and species-specific gene families revealed that *A.*
199 *oerlinghausenensis* does not have many species-specific gene families and shares more gene
200 families with *A. fischeri* than *A. fumigatus* (Fig. 2B). Specifically, we noted that *A.*
201 *oerlinghausenensis* CBS 139183^T has only eight species-specific gene families; in contrast, *A.*
202 *fischeri* and *A. fumigatus* have 1,487 and 548 species-specific gene families, respectively.

203

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.09.033902; this version posted April 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

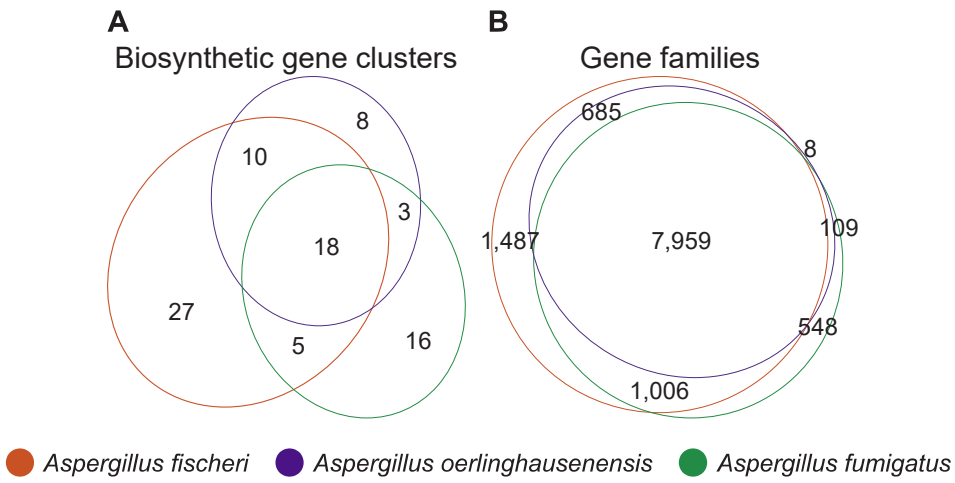


Figure 2. *Aspergillus oerlinghausenensis* shares more gene families and BGCs with *A. fischeri* than *A. fumigatus*. (A) Euler diagram showing species-level shared BGCs. (B) Euler diagram showing species-level shared gene families. In both diagrams, *A. oerlinghausenensis* shares more gene families or BGCs with *A. fischeri* than *A. fumigatus* despite a closer evolutionary relationship.

Despite a closer evolutionary relationship between *A. oerlinghausenensis* and *A. fumigatus*, we found *A. oerlinghausenensis* shares more gene families with *A. fischeri* compared to *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* (Fig. S5).

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 139183^{T} for secondary metabolites. Specifically, we used three different procedures, including

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.09.033902; this version posted April 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

222 the isolation and structure elucidation of metabolites, where possible, followed by two different
223 metabolite profiling procedures that use mass spectrometry techniques. Altogether, we isolated
224 and characterized 19 secondary metabolites; seven from *A. fumigatus*, two from *A.*
225 *oerlinghausenensis*, and ten from *A. fischeri* (Fig. S6). These products encompassed a wide
226 diversity of secondary metabolite classes, such as those derived from polyketide synthases, non-
227 ribosomal peptide-synthetases, terpene synthases and mixed biosynthesis enzymes.
228
229 To characterize the secondary metabolites biosynthesized that were not produced in high enough
230 quantity for structural identification through traditional isolation methods, we employed
231 "dereplication" mass spectrometry protocols specific to natural products research on all tested
232 strains at both 30°C and 37°C (see supporting information, dereplication example; figshare:
233 10.6084/m9.figshare.12055503) (El-Elimat et al., 2013; Ito and Masubuchi, 2014; Gaudêncio
234 and Pereira, 2015; Hubert et al., 2017). We found an overlap of secondary metabolites between
235 strains of the same species (Table S3); for example, monomethylsulochrin (**3**) was isolated from
236 *A. fumigatus* Af293, but through metabolite profiling, its spectral features were noted also in *A.*
237 *fumigatus* strains CEA10 and CEA17. We identified metabolites that were biosynthesized by
238 only one species; for example, pseurotin A (**4**) was solely present in *A. fumigatus* strains.
239 Furthermore, we also found an overlap of several secondary metabolites across species, such as
240 fumagillin (**6**), which was biosynthesized by *A. fumigatus* and *A. oerlinghausenensis*, and
241 fumitremorgin B (**16**), which was biosynthesized by strains of both *A. oerlinghausenensis* and *A.*
242 *fischeri.* Together, these analyses suggest that closely related *Aspergillus* species and strains
243 exhibit variation both within as well as between species in the secondary metabolites produced.
244

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.09.033902; this version posted April 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

To further facilitate comparisons of secondary metabolite profiles within and between species, we used the 1,920 features (i.e., unique *m/z* – retention time pairs) that were identified from all strains at all temperatures (Fig. 3A), to perform Principal Components Analysis (PCA)

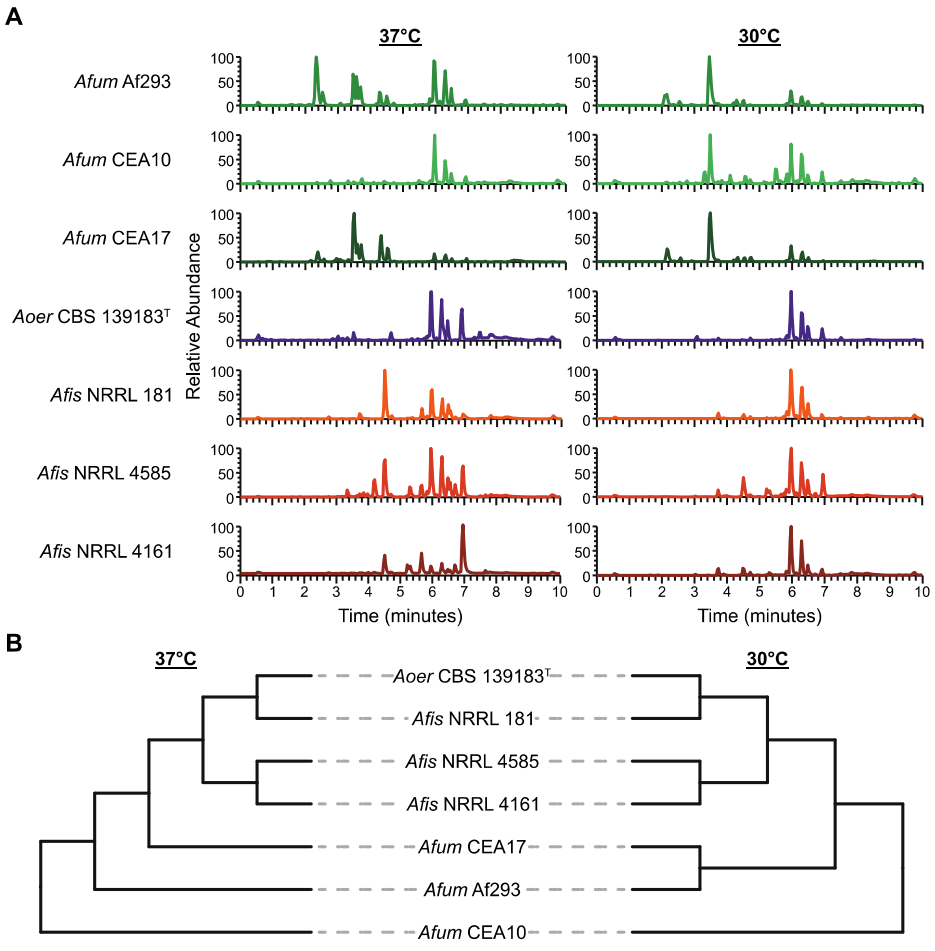


Figure 3. *A. oerlinghausenensis* and *A. fischeri* have more similar secondary metabolite profiles than *A. fumigatus*. (A) UPLC-MS chromatograms of secondary metabolite profiles of *A. fumigatus* and its closest relatives, *A. oerlinghausenensis* and *A. fischeri* at 37°C and 30°C (left and right, respectively). (B) Hierarchical clustering of chromatograms reveals *A. oerlinghausenensis* clusters with *A. fischeri* and not its closest relative, *A. fumigatus* at 37°C and 30°C (left and right, respectively).

(Fig. S7) and hierarchical clustering (Fig. 3B). The PCA plots and hierarchical clustering of the

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.09.033902; this version posted April 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

chromatograms at both 37°C and 30°C (Fig. 3B, S7B, and S7C) indicated that the three *A. fischeri* strains and *A. oerlinghausenensis* were the most chemically similar to each other. When we generated PCA plots from only the compounds that could be isolated from each culture (44 features that represent the 19 isolated compounds) at both 37°C (Fig. S7D) and 30°C (Fig. S7E), the chemical similarities between *A. oerlinghausenensis* and *A. fischeri* were even more evident, with this clustering showing similar results to the total feature dendrogram (Fig. 3B). These data suggest that there are more similarities among the secondary metabolites that are biosynthesized in higher abundance. While the clustering of *A. oerlinghausenensis* CBS 139183^{T} and *A. fischeri* NRRL 181 is conserved, these strains were also shown to be similar to *A. fumigatus* CEA10 (Fig. S7D-E). These combined results suggest that at lower temperatures, such as 30°C, there is a more varied response in how BGCs are being utilized, leading to a more diverse production of chemical compounds.

In summary, our chemical analyses suggest that the secondary metabolite profiles of *A. oerlinghausenensis* and *A. fischeri* are more similar to each other than to *A. fumigatus* (Fig. 3B and S7B-E). This finding is surprising because phylogenetic analysis indicates that *A. oerlinghausenensis* is more closely related to *A. fumigatus* than it is to *A. fischeri* (Fig. 1A). However, the similarity of secondary metabolite profiles of *A. oerlinghausenensis* and *A. fischeri* is consistent with our finding that the genome of *A. oerlinghausenensis* shares higher numbers of BGCs and gene families with *A. fischeri* than with *A. fumigatus* (Fig. 2). Similarly, clustering patterns in secondary metabolite-based plots (Fig. S7B-E) resemble those of BGC-based plots (Fig. S7A), suggesting that the observed similarities in the chemotypes of *A. oerlinghausenensis* and *A. fischeri* are broadly reflected in their metabolism-associated genotypes.

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.09.033902; this version posted April 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Conservation and divergence among biosynthetic gene clusters implicated in *A. fumigatus* pathogenicity

Secondary metabolites are known to play a role in *A. fumigatus* virulence (Raffa and Keller, 2019). We therefore conducted a focused examination of specific *A. fumigatus* BGCs and secondary metabolites that have been previously implicated in the organism's ability to cause human disease (Table 1). We found varying degrees of conservation and divergence that were associated with the absence or presence of a secondary metabolite. Among conserved BGCs that were also associated with conserved secondary metabolite production, we highlight the mycotoxins gliotoxin and fumitremorgin. Interestingly, we note that only *A. fischeri* strains synthesized verruculogen, a secondary metabolite that is implicated in human disease and is encoded by the fumitremorgin BGC (Khoufache et al., 2007; Kautsar et al., 2019). Among divergent BGCs that were associated with the absence of a secondary metabolite, we highlight the trypacidin and fumagillin/pseurotin secondary metabolites. We found that nonpathogenic close relatives of *A. fumigatus* produced some but not all mycotoxins, which provides novel insight into the unique cocktail of secondary metabolites biosynthesized by *A. fumigatus*.

Gliotoxin. Gliotoxin is a highly toxic compound and known virulence factor in *A. fumigatus* (Sugui et al., 2007). Nearly identical BGCs encoding gliotoxin are present in all pathogenic (*A. fumigatus*) and nonpathogenic (*A. oerlinghausenensis* and *A. fischeri*) strains examined (Fig. 4). Additionally, we found that all examined strains synthesized bisdethiobis(methylthio)gliotoxin a derivative from dithiogliotoxin, involved in the down-regulation of gliotoxin biosynthesis (Dolan

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.09.033902; this version posted April 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

et al., 2014), one of the main mechanisms of gliotoxin resistance in *A. fumigatus* (Kautsar et al., 2019).

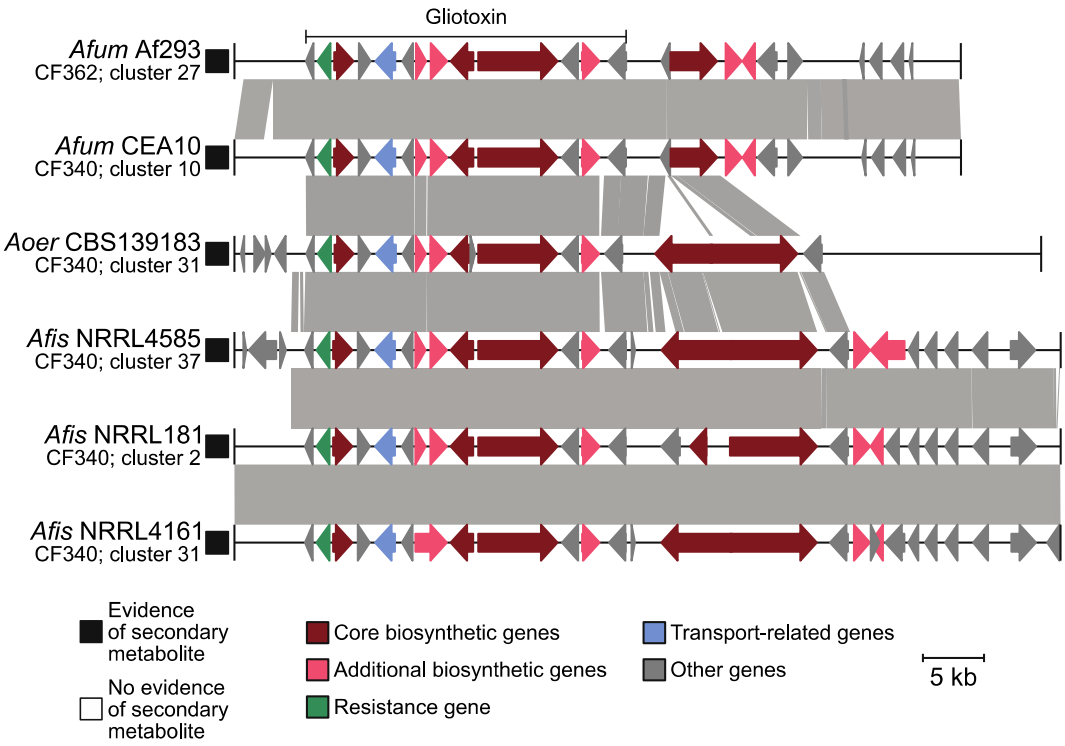
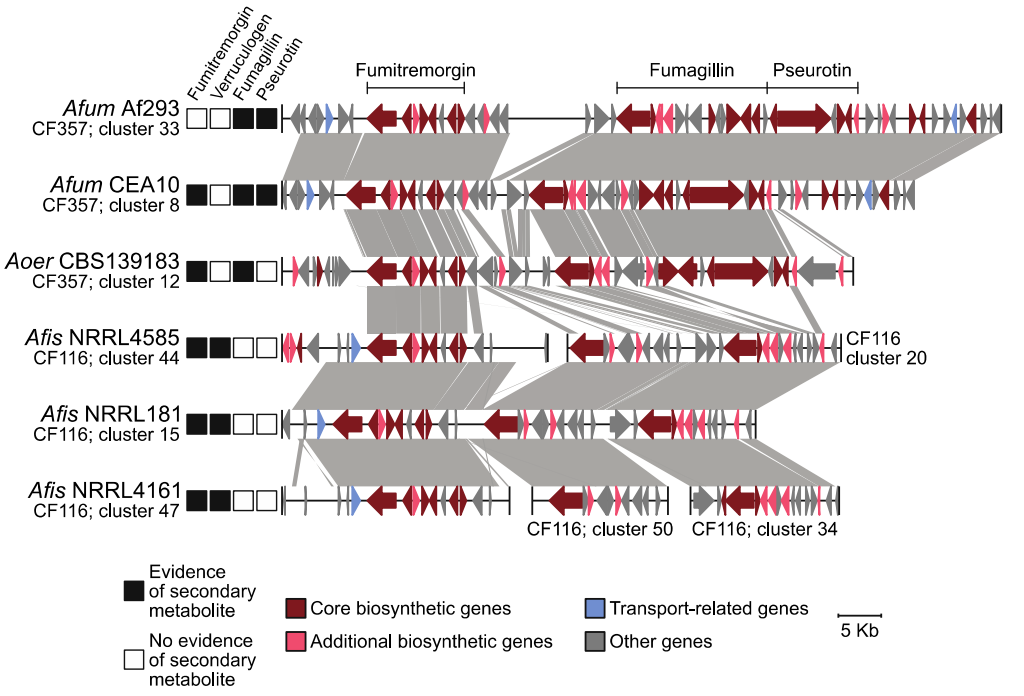


Figure 4. Conservation in the gliotoxin BGC correlates with conserved production of gliotoxin analogs in *A. fumigatus* and nonpathogenic close relatives. Microsynteny analysis reveals a high degree of conservation in the BGC encoding gliotoxin across all isolates. The known gliotoxin gene cluster boundary is indicated above the *A. fumigatus* Af293 BGC. Black and white squares correspond to the presence or absence of the associated secondary metabolite, respectively. Genes are drawn as arrows with orientation indicated by the direction of the arrow. Gene function is indicated by gene color. Genus and species names are written using the following abbreviations: *Afum*: *A. fumigatus*; *Aoer*: *A. oerlinghausenensis*; *Afis*: *A. fischeri*. Below each genus and species abbreviation is the cluster family each BGC belongs to and their cluster number.

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.09.033902; this version posted April 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

317 ***Fumitremorgin and Verruculogen.*** Similarly, there is a high degree of conservation in the
318 BGC that encodes fumitremorgin across all strains (Fig. 5). Fumitremorgins have known
319 antifungal activity, are lethal to brine shrimp, and are implicated in inhibiting mammalian
320 proteins responsible for resistance to anticancer drugs in mammalian cells (Raffa and Keller,
321 2019). We found that conservation in the fumitremorgin BGC is associated with the production
322 of fumitremorgins in all isolates examined. The fumitremorgin BGC is also responsible for the
323 production of verruculogen, which is implicated to aid in *A. fumigatus* pathogenicity by changing
324 the electrophysical properties of human nasal epithelial cells (Khoufache et al., 2007).
325 Interestingly, we found that only *A. fischeri* strains produced verruculogen under the conditions
326 we analyzed.
327
328 ***Trypacidin.*** Examination of the trypacidin BGC, which encodes a spore-borne and cytotoxic
329 secondary metabolite, revealed a conserved cluster found in four pathogenic and
330



331

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.09.033902; this version posted April 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Figure 5. Conservation and divergence in the locus encoding the fumitremorgin and intertwined fumagillin/pseurotin BGCs. Microsynteny analysis reveals conservation in the fumitremorgin BGC across all isolates. Interestingly, only *A. fischeri* strains synthesize verruculogen, a secondary metabolite also biosynthesized by the fumitremorgin BGC. In contrast, the intertwined fumagillin/pseurotin BGCs are conserved between *A. fumigatus* and *A. oerlinghausenensis* but divergent in *A. fischeri*. BGC conservation and divergence is associated with the presence and absence of a secondary metabolite, respectively. The same convention used in Fig. 4 is used to depict evidence of a secondary metabolite, represent genes and broad gene function, genus and species abbreviations, and BGC cluster families and cluster numbers.

---

nonpathogenic strains: *A. fumigatus* Af293, *A. fumigatus* CEA10, *A. oerlinghausenensis* CBS 139183^{T}, and *A. fischeri* NRRL 181 (Fig. S8). 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 *A. fischeri* NRRL l81, 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*.

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.09.033902; this version posted April 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

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* (Fig. 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. 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., 2020a). Currently, it is thought that the ability to cause human disease

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.09.033902; this version posted April 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

380 evolved multiple times among species in section *Fumigati* (Rokas et al., 2020a). Secondary
381 metabolites contribute to the success of the major human pathogen *A. fumigatus* in the host
382 environment (Raffa and Keller, 2019) and are therefore "cards" of virulence (Casadevall, 2007;
383 Knowles et al., 2020). However, whether the closest relatives of *A. fumigatus*, *A.*
384 *oerlinghausenensis* and *A. fischeri*, both of which are nonpathogenic, biosynthesize secondary
385 metabolites implicated in the ability of *A. fumigatus* to cause human disease remained largely
386 unknown. By examining genomic and chemical variation between and within *A. fumigatus* and
387 its closest nonpathogenic relatives, we identified both conservation and divergence (including
388 within species heterogeneity) in BGCs and secondary metabolite profiles (Fig. 1-5, S4, S6-9;
389 Table 1, S1, S3). Examples of conserved BGCs and secondary metabolites include the major
390 virulence factor, gliotoxin (Fig. 4), as well as several others (Fig. 5, S8; Table 1, S1, S3);
391 examples of BGC and secondary metabolite heterogeneity or divergence include pseurotin,
392 fumagillin, and several others (Fig. 5; Table 1, S1, S3). Lastly, we found that the fumitremorgin
393 BGC, which biosynthesizes fumitremorgin in all three species, is also associated with
394 verruculogen biosynthesis in *A. fischeri* strains (Fig. 5).
395
396 One of the surprising findings of our study was that although *A. oerlinghausenensis* and *A.*
397 *fumigatus* are evolutionarily more closely related to each other than to *A. fischeri* (Fig. 1), *A.*
398 *oerlinghausenensis* and *A. fischeri* appear to be more similar to each other than to *A. fumigatus* in
399 BGC composition, gene family content, and secondary metabolite profiles. The power of
400 pathogen-nonpathogen comparative genomics is best utilized when examining closely related
401 species (Fedorova et al., 2008; Jackson et al., 2011; Moran et al., 2011; Mead et al., 2019a;
402 Rokas et al., 2020a). By sequencing genomes from the closest known nonpathogenic relatives of

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.09.033902; this version posted April 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

403 *A. fumigatus*, including the genome of the closest species relative *A. oerlinghausenensis* and
404 additional strains of *A. fischeri*, we provide a powerful resource to study the evolution of *A.*
405 *fumigatus* pathogenicity.
406
407 Our finding that *A. oerlinghausenensis* and *A. fischeri* shares more gene families and BGCs with
408 each other than they do with *A. fumigatus* (Fig. 1C, 2, S4, S5, S9) suggests that the evolutionary
409 trajectory of the *A. fumigatus* ancestor was marked by gene loss. We hypothesize that there were
410 two rounds of gene family and BGC loss in the *A. fumigatus* stem lineage: (1) gene families and
411 BGCs were lost in the common ancestor of *A. fumigatus* and *A. oerlinghausenensis* and (2)
412 additional losses occurred in the *A. fumigatus* ancestor. In addition to losses, we note that 548
413 and 16 gene families and BGCs are unique to *A. fumigatus*, which may have resulted from
414 genetic innovation (e.g., *de novo* gene formation) or unique gene family and BGC retention (Fig.
415 2, S9). In line with the larger number of shared BGCs between *A. oerlinghausenensis* and *A.*
416 *fischeri*, we found their secondary metabolite profiles were also more similar (Fig. 3, S7).
417 Notably, the evolutionary rate of the internal branch leading to the *A. fumigatus* common
418 ancestor is much higher than those in the rest of the branches in our genome-scale phylogeny
419 (Fig. S2B), suggesting that the observed gene loss and gene gain / retention events specific to *A.*
420 *fumigatus* may be part of a wider set of evolutionary changes in the *A. fumigatus* genome. More
421 broadly, these results suggest that comparisons of the pathogen *A. fumigatus* against either the
422 non-pathogen *A. oerlinghausenensis* (this manuscript) or the non-pathogen *A. fischeri* ((Mead et
423 al., 2019a; Knowles et al., 2020) and this manuscript) will both be instructive in understanding
424 the evolution of *A. fumigatus* pathogenicity.
425

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.09.033902; this version posted April 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

When studying *Aspergillus* pathogenicity, it is important to consider any genetic and phenotypic heterogeneity between strains of a single species (Kowalski et al., 2016, 2019; Keller, 2017; Ries et al., 2019; Bastos et al., 2020; Santos et al., 2020). Our finding of strain heterogeneity among gene families, BGCs, and secondary metabolites in *A. fumigatus* and *A. fischeri* (Fig. 1-3, S4, S5, S7, S9) 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* (Fig. S7B-E). These results suggest that strain specific secondary metabolite profiles may play a role in variation of pathogenicity among *A. fumigatus* strains. 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; Bastos et al., 2020; Santos et al., 2020).

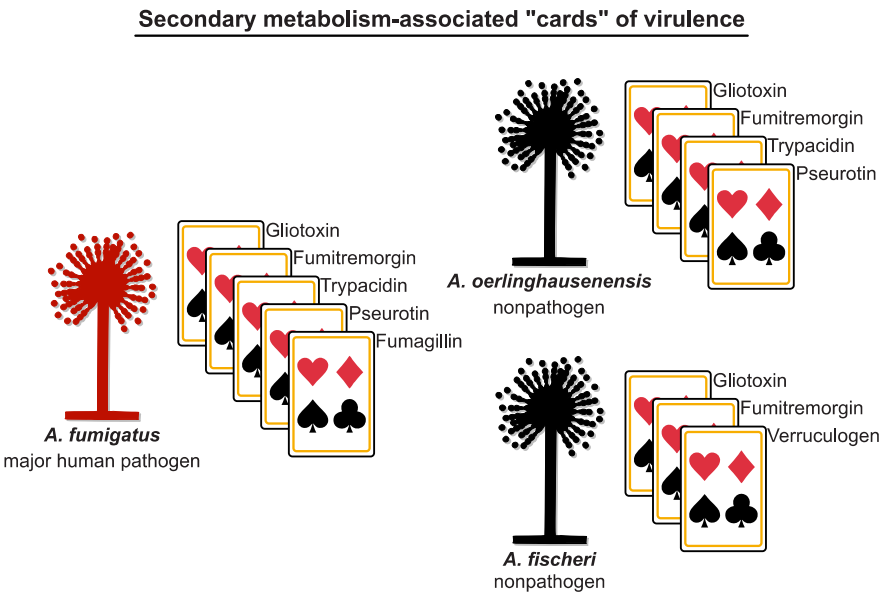


Figure 6. Secondary metabolism-associated "cards" of virulence among *A. fumigatus* and close relatives. Secondary metabolites contribute to the "hand of cards"' that enable *A. fumigatus* to cause

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.09.033902; this version posted April 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

disease. Here, we show that the nonpathogenic closest relatives of *A. fumigatus* possess a subset of the *A. fumigatus* secondary metabolism-associated cards of virulence. We hypothesize that the unique combination of cards of *A. fumigatus* contributes to its pathogenicity and that the cards in *A. oerlinghausenensis* and *A. fischeri* (perhaps in combination with other non-secondary-metabolism-associated cards, such as thermotolerance) are insufficient to cause disease. Pathogenic and nonpathogenic species are shown in red and black, respectively. Cartoons of *Aspergillus* species were obtained from WikiMedia Commons (source: M. Piepenbring) and modified in accordance with the Creative Commons Attribution-Share Alike 3.0 Unported license (https://creativecommons.org/licenses/by-sa/3.0/deed.en).

Secondary metabolites contribute to *A. fumigatus* virulence through diverse processes including suppressing the human immune system and damaging tissues (Table 1). 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, trypacidin, verruculogen, and others (Fig. 4, 5, S8; Table 1, 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 Fig. 6). 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.

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.09.033902; this version posted April 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Methods

Strain acquisition, DNA extraction, and sequencing

Two strains of *Aspergillus 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 *Aspergillus oerlinghausenensis* (CBS 139183^{T}) was acquired from Westerdijk Fungal Biodiversity Institute, Utrecht, The Netherlands. These strains were grown in 50 ml of liquid yeast extract soy peptone dextrose (YESD) medium. After approximately seven 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, 0.22µm 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, US) 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’.

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.09.033902; this version posted April 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

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 the shortest contig among the longest contigs that account for 50% of the genome assembly's length) (Yandell and Ence, 2012). To determine gene-content completeness, we implemented the BUSCO, v2.0.1 (Waterhouse et al., 2018), pipeline using the 'genome' mode. In this mode, the BUSCO pipeline examines assembly contigs for the presence of near-universally single copy orthologous genes (hereafter referred to as BUSCO genes) using a predetermined database of orthologous genes from the OrthoDB, v9 (Waterhouse et al., 2013). We used the OrthoDB database for Pezizomycotina (3,156 BUSCO genes). Each BUSCO gene is determined to be present in a single copy, as duplicate sequences, fragmented, 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 ± 9,696.11; Fig. 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

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.09.033902; this version posted April 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

accession: GCA_000002655.1), *Aspergillus lentulus* IFM 54703 (GenBank accession: GCA_001445615.1), *Aspergillus novofumigatus* IBT 16806 (GenBank accession: GCA_002847465.1), and *Aspergillus udagawae* 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_000150145.1), *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_000149645.2). 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 ten genome from three species: four *A. fumigatus* strains, one *A. oerlinghausenensis* strain, and five *A. fischeri* strains.

Maximum likelihood phylogenetics and Bayesian estimation of divergence times

To reconstruct the evolutionary history among the ten *Aspergillus* genomes, we implemented a recently developed pipeline (Steenwyk et al., 2019) which, relies on the concatenation-approach

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.09.033902; this version posted April 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

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 ten *Aspergillus* genomes, we used the BUSCO pipeline with the Pezizomycotina database as described above. We identified 3,041 BUSCO genes present at a single copy in all ten *Aspergillus* genomes and created multi-FASTA files for each BUSCO gene that contained the protein sequences for all ten 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 forced onto the protein sequence alignments using a custom Python, v3.5.2 (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 (Nguyen 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 Zhedanov, 2011). To evaluate bipartition support, we used 5,000 ultrafast bootstrap approximations (Hoang et al., 2018).

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.09.033902; this version posted April 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

To estimate divergence times among the ten *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 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 one 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 to 6.74 million years ago) (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_gamma = 1 186.63). The 'sigma2_gamma' 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 of gene families and analyses of putative biosynthetic gene clusters

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.09.033902; this version posted April 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

To identify gene families across the ten *Aspergillus* genomes, we used a Markov clustering approach. Specifically, we used OrthoFinder, v2.3.8 (Emms and Kelly, 2019). OrthoFinder first conducts a blast all-vs-all using the protein sequences of all ten *Aspergillus* genomes and NCBI's Blast+, v2.3.0 (Camacho et al., 2009), software. After normalizing blast bit scores, genes are clustered into discrete orthogroups using a Markov clustering approach (van Dongen, 2000). We clustered genes using an inflation parameter of 1.5. The resulting orthogroups were used proxies for gene families.

To identify putative biosynthetic gene clusters (BGCs), we used the gene boundaries predictions from the MAKER software as input into antiSMASH, v4.1.0 (Weber et al., 2015). To identify homologous BGCs across the ten *Aspergillus* genomes, we used the software BiG-SCAPE, v20181005 (Navarro-Muñoz et al., 2020). Based on the Jaccard Index of domain types, sequence similarity among domains, and domain adjacency, BiG-SCAPE calculates a similarity metric between pairwise combinations of clusters where smaller values indicate greater BGC similarity. BiG-SCAPE's similarity metric can then be used as an edge-length in network analyses of cluster similarity. We evaluated networks using an edge-length cutoff from 0.1-0.9 with a step of 0.1 (Fig. S4). We found networks with an edge-length cutoff of 0.4-0.6 to be similar and based further analyses on a cutoff of 0.5. For BGCs of interest, we supplemented BiG-SCAPE's approach to identifying homologous BGCs with visualize inspection of microsyteny and blast-based analyses using NCBI's BLAST+, v2.3.0 (Camacho et al., 2009). Similar sequences in microsynteny analyses were defined as at least 100 bp in length, at least 30 percent similarity, and an expectation value threshold of 0.01.

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.09.033902; this version posted April 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Identification and characterization of secondary metabolite production

General experimental procedures

The ^{1}H 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 UPLC (Waters Corp.) was utilized for both mass spectrometers, using a BEH C_{18} column (1.7 μm; 50 mm x 2.1 mm) set to a temperature of 40°C and a flow rate of 0.3 ml/min. The mobile phase consisted of a linear gradient of CH_3CN-H_2O (both acidified with 0.1% formic acid), starting at 15% CH_3CN and increasing linearly to 100% CH_3CN over 8 min, with a 1.5 min hold before returning to the starting condition. The HPLC separations were performed with Atlantis T3 C_{18} semi-preparative (5 μm; 10 x 250 mm) and preparative (5 μm; 19 x 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 detector (PDA), 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 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;

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.09.033902; this version posted April 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

626 difco (PDA) were excised from the edge of the Petri dish culture and transferred to separate
627 liquid seed media that contained 10 ml YESD broth (2% soy peptone, 2% dextrose, and 1% yeast
628 extract; 5 g of yeast extract, 10 g of soy peptone, and 10 g of D-glucose in 500 ml of deionized
629 H_2O) and allowed to grow at 23˚C with agitation at 100 rpm for three days. The YESD seed
630 cultures of the fungi were subsequently used to inoculate solid-state oatmeal fermentation
631 cultures, which were either grown at room temperature (approximately 23˚C under 12h
632 light/dark cycles for 14 days), 30˚C, or 37˚C; all growths at the latter two temperatures were
633 carried out in an incubator (VWR International) in the dark over four days. The oatmeal cultures
634 were prepared in 250 ml Erlenmeyer flasks that contained 10 g of autoclaved oatmeal (10 g of
635 oatmeal with 17 ml of deionized H_2O and sterilized for 15–20 minutes at 121°C). For all fungal
636 strains three flasks of oatmeal cultures were grown at all three temperatures, except for *A.*
637 *oerlinghausenensis* (CBS 139183^T) at room temperature and *A. fumigatus* (Af293) at 37˚C. For
638 CBS 139183^T, the fungal cultures were grown in four flasks, while for Af293 eight flasks were
639 grown in total. The growths of these two strains were performed differently from the rest because
640 larger amounts of extract were required in order to perform detailed chemical characterization.
641 The cultures were extracted by adding 60 ml of (1:1) MeOH-$CHCl_3$ to each 250 ml flask,
642 chopping thoroughly with a spatula, and shaking overnight (~ 16 h) at ~ 100 rpm at room
643 temperature. The culture was filtered *in vacuo*, and 90 ml $CHCl_3$ and 150 ml H_2O were added to
644 the filtrate. The mixture was stirred for 30 min and then transferred to a separatory funnel. The
645 organic layer ($CHCl_3$) was drawn off and evaporated to dryness *in vacuo*. The dried organic layer
646 was reconstituted in 100 ml of (1:1) MeOH–CH_3CN and 100 ml of hexanes, transferred to a
647 separatory funnel, and shaken vigorously. The defatted organic layer (MeOH–CH_3CN) was
648 evaporated to dryness *in vacuo*.

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.09.033902; this version posted April 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

649

650 To isolate compounds, the defatted extract was dissolved in $CHCl_3$, absorbed onto Celite 545
651 (Acros Organics), and fractioned by normal phase flash chromatography using a gradient of
652 hexane-$CHCl_3$-MeOH. *Aspergillus fischeri* strain NRRL 181 was chemically characterized
653 previously (Knowles et al., 2019; Mead et al., 2019a). *A. fumigatus* strain Af293, grown at 37°C,
654 was subjected to a 12g column at a flow rate of 30 ml/min and 61.0 column volumes, which
655 yielded four fractions. Fraction 2 was further purified via preparative HPLC using a gradient
656 system of 30:70 to 100:0 of CH_3CN-H_2O with 0.1% formic acid over 40 min at a flow rate of
657 16.9 ml/min to yield six subfractions. Subfractions 1, 2 and 5, yielded cyclo(L-Pro-L-Leu) (**1**)
658 (Li et al., 2008) (0.89 mg), cyclo(L-Pro-L-Phe) (**2**) (Campbell et al., 2009) (0.71 mg), and
659 monomethylsulochrin (**3**) (Ma et al., 2004) (2.04 mg), which eluted at approximately 5.7, 6.3,
660 and 10.7 min, respectively. Fraction 3 was further purified via preparative HPLC using a
661 gradient system of 40:60 to 65:35 of CH_3CN-H_2O with 0.1% formic acid over 30 min at a flow
662 rate of 16.9 ml/min to yield four subfractions. Subfractions 1 and 2 yielded pseurotin A (**4**)
663 (Wang et al., 2011) (12.50 mg) and bisdethiobis(methylthio)gliotoxin (**5**) (Afiyatullov et al.,
664 2005) (13.99 mg), which eluted at approximately 7.5 and 8.0 min, respectively.

665

666 *A. fumigatus* strain CEA10, grown at 37°C, was subjected to a 4g column at a flow rate of 18
667 ml/min and 90.0 column volumes, which yielded five fractions. Fraction 1 was purified via
668 preparative HPLC using a gradient system of 50:50 to 100:0 of CH_3CN-H_2O with 0.1% formic
669 acid over 45 min at a flow rate of 16.9 ml/min to yield eight subfractions. Subfraction 1, yielded
670 fumagillin (**6**) (Halász et al., 2000) (1.69 mg), which eluted at approximately 18.5 min. Fraction
671 2 was purified via semi-preparative HPLC using a gradient system of 35:65 to 80:20 of CH_3CN-

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.09.033902; this version posted April 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

672 H_2O with 0.1% formic acid over 30 min at a flow rate of 4.6 ml/min to yield 10 subfractions.
673 Subfraction 5 yielded fumitremorgin C (**7**) (Kato et al., 2009) (0.25 mg), which eluted at
674 approximately 15.5 min. Fraction 3 was purified via preparative HPLC using a gradient system
675 of 40:60 to 100:0 of CH_3CN-H_2O with 0.1% formic acid over 30 min at a flow rate of 16.9
676 ml/min to yield nine subfractions. Subfraction 2 yielded pseurotin A (**4**) (1.64 mg), which eluted
677 at approximately 7.3 min.
678
679 *Aspergillus oerlinghausenensis* strain CBS 139183^T, grown at RT, was subjected to a 4g column
680 at a flow rate of 18 ml/min and 90 column volumes, which yielded 4 fractions. Fraction 3 was
681 further purified via preparative HPLC using a gradient system of 35:65 to 70:30 of CH_3CN-H_2O
682 with 0.1% formic acid over 40 min at a flow rate of 16.9 ml/min to yield 11 subfractions.
683 Subfractions 3 and 10 yielded spiro [5H,10H-dipyrrolo[1,2-a:1′,2′-d]pyrazine-2-(3H),2′-
684 [2H]indole]-3′,5,10(1′H)-trione (**8**) (Wang et al., 2008) (0.64 mg) and helvolic acid (**9**) (Zhao et
685 al., 2010) (1.03 mg), which eluted at approximately 11.5 and 39.3 min, respectively. (see NMR
686 supporting information; figshare: 10.6084/m9.figshare.12055503).
687
688 **Metabolite profiling by mass spectrometry**
689 The metabolite profiling by mass spectrometry, also known as dereplication, was performed as
690 stated previously (El-Elimat et al., 2013). Briefly, ultraperformance liquid chromatography-
691 photodiode array-electrospray ionization high resolution tandem mass spectrometry (UPLC-
692 PDA-HRMS-MS/MS) was utilized to monitor for secondary metabolites across all strains
693 (Af293, CEA10, CEA17, CBS 139183^T, NRRL 181, NRRL 4161, and NRRL 4585). Utilizing
694 positive-ionization mode, ACD MS Manager with add-in software IntelliXtract (Advanced

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.09.033902; this version posted April 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Chemistry Development, Inc.; Toronto, Canada) was used for the primary analysis of the UPLC-MS chromatograms. The data from 19 secondary metabolites are provided in the Supporting Information (see Dereplication table; figshare: 10.6084/m9.figshare.12055503), which for each secondary metabolite lists: molecular formula, retention time, UV-absorption maxima, high-resolution full-scan mass spectra, and MS-MS data (top 10 most intense peaks).

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 (Af293, CEA10, and CEA17): noise level (absolute value), 1×10^6; 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), 1×10^6; minimum peak duration, 0.05 min; *m/z* variation tolerance, 0.05; and *m/z* intensity variation, 20%; and all strains (Af293, CEA10, CEA17, CBS 139183^T, NRRL 181, NRRL 4161, and NRRL 4585): noise level (absolute value), 7×10^5; 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 retention time balance set at 10.0 each, *m/z* tolerance set at 0.001, and RT tolerance set at 0.5 mins. The resulting data matrix was exported to Excel (Microsoft) for analysis as a set of *m/z* – retention time pairs with individual peak areas detected in triplicate analyses. Samples that did not possess detectable quantities of a given marker ion were assigned

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.09.033902; this version posted April 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

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 minutes and/or had an *m/z* ratio less than 200 or greater than 800 Da were removed from analysis. Relative standard deviation 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* – retention time 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

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 will become available through the Figshare repository 10.6084/m9.figshare.12055503 upon publication. 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 others. Among natural products analysis, we provide information that supports methods, and results, including NMR spectra.

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.09.033902; this version posted April 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

741

742 **Funding**

743 JLS and AR are supported by the Howard Hughes Medical Institute through the James H.
744 Gilliam Fellowships for Advanced Study program. AR has additional support from a Discovery
745 Grant from Vanderbilt University. GHG is supported by the Brazilian funding agencies
746 Fundacão de Amparo a Pesquisa do Estado de São Paulo (FAPESP 2016/07870-9) and Conselho
747 Nacional de Desenvolvimento Cientıfico e Tecnologico (CNPq). NHO is supported by the
748 National Cancer Institute (P01 CA125066). SLK and CDR were supported in part by the
749 National Institutes of Health via the National Center for Complementary and Integrative Health
750 (F31 AT010558) and the National Institute of General Medical Sciences (T34 GM113860),
751 respectively.

752

753 **Acknowledgements**

754 We thank the labs of Rokas, Oberlies, and Goldman for helpful discussion and support of this
755 work.

Table 1. Select *A. fumigatus* secondary metabolites implicated in human disease

	Function	Reference(s)	Evidence of secondary metabolite						
			A. fumigatus			*A. oerlinghausenensis*	*A. fischeri*		
			Af293	CEA10	CEA17	CBS 139183^T	NRRL 181	NRRL 4585	NRRL 4161
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)	+	+	+	+	-	-	-
Fumagillin	Inhibits neutrophil function	(Fallon et al., 2010, 2011)	+	+	+	-	-	-	-

A list of select secondary metabolites implicated in human disease and their functional role are described here. All secondary metabolites listed or analogs thereof were identified during secondary metabolite profiling. '+' (orange) and '-' (blue) indicate if the secondary metabolite was or was not produced by a strain of *A. fumigatus* or its two closest relatives.

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.09.033902; this version posted April 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.09.033902; this version posted April 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

760 **Supplementary legends**

761 **Fig. S1. Metrics of genomes assembly quality and number of predicted gene.** Genome size,
762 N50, number of genes, number of scaffolds and percent single copy BUSCO genes (scBUSCO)
763 are depicted here. Examination of metrics reveal genomes are of sufficient quality for
764 comparative genomics purposes. Of concern, we noted A. oerlinghausenensis CBS 139183^{T} was
765 assembled into 5,300 contigs; however, we evaluated the assembly's N50 value and gene content
766 completeness (461,327 base pairs and 98.90% BUSCO genes present in single copy,
767 respectively) and found the *A. oerlinghausenensis* genome is suitable for comparative genomic
768 analyses.

769

770 **Fig. S2. A reconstructed evolutionary history and timetree of *A. fumigatus* and its closest**
771 **relatives.** (top) Divergence times were estimated using a concatenated matrix of 3,041 genes
772 (5,602,272 sites). Blue bars at each node correspond to the 95% divergence time confidence
773 interval. Divergence times and confidence intervals for each internode are shown on the right
774 side of the figure. (bottom) A phylogeny where branch lengths represent substitutions per site
775 rather than geologic time. A cladogram is drawn to the right of the phylogeny to clarify
776 divergences where branch lengths are short (e.g., among strains of *A. fischeri*).

777

778 **Fig. S3. Moderate edge cut-off lengths resulted in qualitatively similar networks.** Networks
779 using edge cut-offs ranging from 0.1-0.9 with a step of 0.1 were evaluated. Networks from 0.4 to
780 0.6 were qualitatively similar. Thus, we used cluster families inferred from the network using an
781 edge cut-off of 0.5.

782

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.09.033902; this version posted April 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

783 **Fig. S4. Species and strain heterogeneity among BGC cluster presence and absence.** (A)
784 Examination of the number of strains in each cluster family reveal a wide variation. For example,
785 all 10 genomes are represented in 17 cluster families; in contrast, 18 cluster families have only
786 one BGC. (B) Species occupancy among cluster families reveal *A. fischeri* has the largest
787 number of unique BGCs followed by *A. fumigatus* and *A. oerlinghausenensis*. (C) Strain-level
788 presence and absence patterns among cluster families reveal substantial species and strain
789 heterogeneity. Cluster family identifiers are shown along the x-axis. Genus and species names
790 are written using the following abbreviations: *Afum*: *A. fumigatus*; *Aoer*: *A. oerlinghausenensis*;
791 *Afis*: *A. fischeri*.
792
793 **Fig. S5. Gene family presence and absence follows a similar pattern to BGCs.** Orthogroups
794 were used as proxies for gene families. A strain-level UpSet plot for gene family presence and
795 absence patterns reveals heterogeneity among strains.
796
797 **Fig. S6. Structures of isolated fungal metabolites.** Secondary metabolites produced in
798 sufficient quantity were isolated for structural determination. The structures of compounds are
799 correlated to each strain from where they were isolated.
800
801 **Fig. S7. Principal component analysis of BGC presence and absence and secondary**
802 **metabolite profiles mirror one another.** (A) Principal component analysis of BGC presence
803 and absence reveal that each species is distinct from the other. Furthermore, *A.*
804 *oerlinghausenensis* is between *A. fischeri* and *A. fumigatus*. (B, C) Broadly, similar patterns of

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.09.033902; this version posted April 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

805 species relationships in principal component space are observed among all metabolites produced
806 at 37°C and 30°C. (D, E) Similar results were observed for isolatable metabolites.
807
808 **Fig. S8. Small sequence divergences in the trypacidin BGC are associated with the**
809 **production of trypacidin or the lack thereof.** The trypacidin BGC is found in *A. fumigatus*
810 strains Af293 and CEA10, *A. oerlinghausenensis* CBS 139183^{T}, and *A. fischeri* NRRL 181.
811 Evidence of trypacidin biosynthesis is found in all isolates with the exception of *A. fischeri*
812 NRRL 181. The absence of trypacidin biosynthesis is associated with the absence of a RING
813 finger gene in *A. fischeri* NRRL 181. Black and white squares correspond to the presence or
814 absence of the associated secondary metabolite, respectively. Genus and species names are
815 written using the following abbreviations: *Afum*: *A. fumigatus*; *Aoer*: *A. oerlinghausenensis*; *Afis*:
816 *A. fischeri*.

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.09.033902; this version posted April 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

References

Abad, A., Victoria Fernández-Molina, J., Bikandi, J., Ramírez, A., Margareto, J., Sendino, J., et al. (2010). What makes Aspergillus fumigatus a successful pathogen? Genes and molecules involved in invasive aspergillosis. *Rev. Iberoam. Micol.* 27, 155–182. doi:10.1016/j.riam.2010.10.003.

Afiyatullov, S. S., Kalinovskii, A. I., Pivkin, M. V., Dmitrenok, P. S., and Kuznetsova, T. A. (2005). Alkaloids from the Marine Isolate of the Fungus Aspergillus fumigatus. *Chem. Nat. Compd.* 41, 236–238. doi:10.1007/s10600-005-0122-y.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., et al. (2012). SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *J. Comput. Biol.* 19, 455–477. doi:10.1089/cmb.2012.0021.

Bastos, R. W., Valero, C., Silva, L. P., Schoen, T., Drott, M., Brauer, V., et al. (2020). Functional characterization of clinical isolates of the opportunistic fungal pathogen Aspergillus nidulans. *bioRxiv*, 2020.01.28.917278. doi:10.1101/2020.01.28.917278.

Benedict, K., Jackson, B. R., Chiller, T., and Beer, K. D. (2019). Estimation of Direct Healthcare Costs of Fungal Diseases in the United States. *Clin. Infect. Dis.* 68, 1791–1797. doi:10.1093/cid/ciy776.

Bodinaku, I., Shaffer, J., Connors, A. B., Steenwyk, J. L., Biango-Daniels, M. N., Kastman, E. K., et al. (2019). Rapid Phenotypic and Metabolomic Domestication of Wild Penicillium Molds on Cheese. *MBio* 10. doi:10.1128/mBio.02445-19.

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi:10.1093/bioinformatics/btu170.

Bongomin, F., Gago, S., Oladele, R., and Denning, D. (2017). Global and Multi-National

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.09.033902; this version posted April 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

840 Prevalence of Fungal Diseases—Estimate Precision. *J. Fungi* 3, 57.
841 doi:10.3390/jof3040057.
842 Caesar, L. K., Kvalheim, O. M., and Cech, N. B. (2018). Hierarchical cluster analysis of
843 technical replicates to identify interferents in untargeted mass spectrometry metabolomics.
844 *Anal. Chim. Acta* 1021, 69–77. doi:10.1016/j.aca.2018.03.013.
845 Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009).
846 BLAST+: architecture and applications. *BMC Bioinformatics* 10, 421. doi:10.1186/1471-
847 2105-10-421.
848 Campbell, J., Lin, Q., Geske, G. D., and Blackwell, H. E. (2009). New and Unexpected Insights
849 into the Modulation of LuxR-Type Quorum Sensing by Cyclic Dipeptides. *ACS Chem. Biol.*
850 4, 1051–1059. doi:10.1021/cb900165y.
851 Capella-Gutierrez, S., Silla-Martinez, J. M., and Gabaldon, T. (2009). trimAl: a tool for
852 automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25,
853 1972–1973. doi:10.1093/bioinformatics/btp348.
854 Casadevall, A. (2007). Determinants of virulence in the pathogenic fungi. *Fungal Biol. Rev.* 21,
855 130–132. doi:10.1016/j.fbr.2007.02.007.
856 Cock, P. J. A., Antao, T., Chang, J. T., Chapman, B. A., Cox, C. J., Dalke, A., et al. (2009).
857 Biopython: freely available Python tools for computational molecular biology and
858 bioinformatics. *Bioinformatics* 25, 1422–1423. doi:10.1093/bioinformatics/btp163.
859 Dagenais, T. R. T., and Keller, N. P. (2009). Pathogenesis of Aspergillus fumigatus in Invasive
860 Aspergillosis. *Clin. Microbiol. Rev.* 22, 447–465. doi:10.1128/CMR.00055-08.
861 Dolan, S. K., Owens, R. A., O'Keeffe, G., Hammel, S., Fitzpatrick, D. A., Jones, G. W., et al.
862 (2014). Regulation of Nonribosomal Peptide Synthesis: bis-Thiomethylation Attenuates

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.09.033902; this version posted April 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Gliotoxin Biosynthesis in Aspergillus fumigatus. *Chem. Biol.* 21, 999–1012. doi:10.1016/j.chembiol.2014.07.006.

Drgona, L., Khachatryan, A., Stephens, J., Charbonneau, C., Kantecki, M., Haider, S., 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. doi:10.1007/s10096-013-1944-3.

El-Elimat, T., Figueroa, M., Ehrmann, B. M., Cech, N. B., Pearce, C. J., and Oberlies, N. H. (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. doi:10.1021/np4004307.

Emms, D. M., and Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* 20, 238. doi:10.1186/s13059-019-1832-y.

Fallon, J. P., Reeves, E. P., and Kavanagh, K. (2010). Inhibition of neutrophil function following exposure to the Aspergillus fumigatus toxin fumagillin. *J. Med. Microbiol.* 59, 625–633. doi:10.1099/jmm.0.018192-0.

Fallon, J. P., Reeves, E. P., and Kavanagh, K. (2011). The Aspergillus fumigatus toxin fumagillin suppresses the immune response of Galleria mellonella larvae by inhibiting the action of haemocytes. *Microbiology* 157, 1481–1488. doi:10.1099/mic.0.043786-0.

Fedorova, N. D., Khaldi, N., Joardar, V. S., Maiti, R., Amedeo, P., Anderson, M. J., et al. (2008). Genomic islands in the pathogenic filamentous fungus Aspergillus fumigatus. *PLoS Genet.* 4. doi:10.1371/journal.pgen.1000046.

Gaudêncio, S. P., and Pereira, F. (2015). Dereplication: racing to speed up the natural products discovery process. *Nat. Prod. Rep.* 32, 779–810. doi:10.1039/C4NP00134F.

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.09.033902; this version posted April 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

886 Gauthier, T., Wang, X., Sifuentes Dos Santos, J., Fysikopoulos, A., Tadrist, S., Canlet, C., et al.
887 (2012). Trypacidin, a Spore-Borne Toxin from Aspergillus fumigatus, Is Cytotoxic to Lung
888 Cells. *PLoS One* 7, e29906. doi:10.1371/journal.pone.0029906.
889 González-Lobato, L., Real, R., Prieto, J. G., Álvarez, A. I., and Merino, G. (2010). Differential
890 inhibition of murine Bcrp1/Abcg2 and human BCRP/ABCG2 by the mycotoxin
891 fumitremorgin C. *Eur. J. Pharmacol.* 644, 41–48. doi:10.1016/j.ejphar.2010.07.016.
892 Grahl, N., Shepardson, K. M., Chung, D., and Cramer, R. A. (2012). Hypoxia and Fungal
893 Pathogenesis: To Air or Not To Air? *Eukaryot. Cell* 11, 560–570. doi:10.1128/EC.00031-
894 12.
895 Guruceaga, X., Ezpeleta, G., Mayayo, E., Sueiro-Olivares, M., Abad-Diaz-De-Cerio, A., Aguirre
896 Urízar, J. M., et al. (2018). A possible role for fumagillin in cellular damage during host
897 infection by Aspergillus fumigatus. *Virulence* 9, 1548–1561.
898 doi:10.1080/21505594.2018.1526528.
899 Guruceaga, X., Perez-Cuesta, U., Abad-Diaz de Cerio, A., Gonzalez, O., Alonso, R. M.,
900 Hernando, F. L., et al. (2019). Fumagillin, a Mycotoxin of Aspergillus fumigatus:
901 Biosynthesis, Biological Activities, Detection, and Applications. *Toxins (Basel).* 12, 7.
902 doi:10.3390/toxins12010007.
903 Halász, J., Podányi, B., Vasvári-Debreczy, L., Szabó, A., Hajdú, F., Böcskei, Z., et al. (2000).
904 Structure Elucidation of Fumagillin-Related Natural Products. *Tetrahedron* 56, 10081–
905 10085. doi:10.1016/S0040-4020(00)00979-0.
906 Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q., and Vinh, L. S. (2018). UFBoot2:
907 Improving the Ultrafast Bootstrap Approximation. *Mol. Biol. Evol.* 35, 518–522.
908 doi:10.1093/molbev/msx281.

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.09.033902; this version posted April 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

909 Holt, C., and Yandell, M. (2011). MAKER2: an annotation pipeline and genome-database
910 management tool for second-generation genome projects. *BMC Bioinformatics* 12, 491.
911 doi:10.1186/1471-2105-12-491.
912 Houbraken, J., Weig, M., Groß, U., Meijer, M., and Bader, O. (2016). Aspergillus
913 oerlinghausenensis , a new mould species closely related to A. fumigatus. *FEMS Microbiol.*
914 *Lett.* 363, fnv236. doi:10.1093/femsle/fnv236.
915 Hubert, J., Nuzillard, J.-M., and Renault, J.-H. (2017). Dereplication strategies in natural product
916 research: How many tools and methodologies behind the same concept? *Phytochem. Rev.*
917 16, 55–95. doi:10.1007/s11101-015-9448-7.
918 Ishikawa, M., Ninomiya, T., Akabane, H., Kushida, N., Tsujiuchi, G., Ohyama, M., et al. (2009).
919 Pseurotin A and its analogues as inhibitors of immunoglobuline E production. *Bioorg. Med.*
920 *Chem. Lett.* 19, 1457–1460. doi:10.1016/j.bmcl.2009.01.029.
921 Ito, T., and Masubuchi, M. (2014). Dereplication of microbial extracts and related analytical
922 technologies. *J. Antibiot. (Tokyo).* 67, 353–360. doi:10.1038/ja.2014.12.
923 Jackson, R. W., Johnson, L. J., Clarke, S. R., and Arnold, D. L. (2011). Bacterial pathogen
924 evolution: breaking news. *Trends Genet.* 27, 32–40. doi:10.1016/j.tig.2010.10.001.
925 Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., and Jermiin, L. S. (2017).
926 ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* 14,
927 587–589. doi:10.1038/nmeth.4285.
928 Kamei, K., and Watanabe, A. (2005). Aspergillus mycotoxins and their effect on the host. *Med.*
929 *Mycol.* 43, 95–99. doi:10.1080/13693780500051547.
930 Kato, N., Suzuki, H., Takagi, H., Asami, Y., Kakeya, H., Uramoto, M., et al. (2009).
931 Identification of Cytochrome P450s Required for Fumitremorgin Biosynthesis in

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.09.033902; this version posted April 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

932 Aspergillus fumigatus. *ChemBioChem* 10, 920–928. doi:10.1002/cbic.200800787.

933 Katoh, K., and Standley, D. M. (2013). MAFFT Multiple Sequence Alignment Software Version
934 7: Improvements in Performance and Usability. *Mol. Biol. Evol.* 30, 772–780.
935 doi:10.1093/molbev/mst010.

936 Kautsar, S. A., Blin, K., Shaw, S., Navarro-Muñoz, J. C., Terlouw, B. R., van der Hooft, J. J. J.,
937 et al. (2019). MIBiG 2.0: a repository for biosynthetic gene clusters of known function.
938 *Nucleic Acids Res.* doi:10.1093/nar/gkz882.

939 Keller, N. P. (2017). Heterogeneity confounds establishment of "a" model microbial strain. *MBio*
940 8. doi:10.1128/mBio.00135-17.

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

943 Khoufache, K., Puel, O., Loiseau, N., Delaforge, M., Rivollet, D., Coste, A., et al. (2007).
944 Verruculogen associated with Aspergillus fumigatus hyphae and conidia modifies the
945 electrophysiological properties of human nasal epithelial cells. *BMC Microbiol.* 7, 5.
946 doi:10.1186/1471-2180-7-5.

947 Knowles, S. L., Mead, M. E., Silva, L. P., Raja, H. A., Steenwyk, J. L., Goldman, G. H., et al.
948 (2020). Gliotoxin, a Known Virulence Factor in the Major Human Pathogen Aspergillus
949 fumigatus , Is Also Biosynthesized by Its Nonpathogenic Relative Aspergillus fischeri.
950 *MBio* 11. doi:10.1128/mBio.03361-19.

951 Knowles, S. L., Vu, N., Todd, D. A., Raja, H. A., Rokas, A., Zhang, Q., et al. (2019). Orthogonal
952 Method for Double-Bond Placement via Ozone-Induced Dissociation Mass Spectrometry
953 (OzID-MS). *J. Nat. Prod.* 82, 3421–3431. doi:10.1021/acs.jnatprod.9b00787.

954 Korf, I. (2004). Gene finding in novel genomes. *BMC Bioinformatics* 5, 59. doi:10.1186/1471-

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.09.033902; this version posted April 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

2105-5-59.

Kowalski, C. H., Beattie, S. R., Fuller, K. K., McGurk, E. A., Tang, Y.-W., Hohl, T. M., et al. (2016). Heterogeneity among Isolates Reveals that Fitness in Low Oxygen Correlates with Aspergillus fumigatus Virulence. *MBio* 7. doi:10.1128/mBio.01515-16.

Kowalski, C. H., Kerkaert, J. D., Liu, K.-W., Bond, M. C., Hartmann, R., Nadell, C. D., et al. (2019). Fungal biofilm morphology impacts hypoxia fitness and disease progression. *Nat. Microbiol.* 4, 2430–2441. doi:10.1038/s41564-019-0558-7.

Kvalheim, O. M., Chan, H., Benzie, I. F. F., Szeto, Y., Tzang, A. H., Mok, D. K., 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. doi:10.1016/j.chemolab.2011.02.002.

Latgé, J.-P., and Chamilos, G. (2019). Aspergillus fumigatus and Aspergillosis in 2019. *Clin. Microbiol. Rev.* 33. doi:10.1128/CMR.00140-18.

Li, X.-J., Zhang, Q., Zhang, A.-L., and Gao, J.-M. (2012). Metabolites from Aspergillus fumigatus, an endophytic fungus associated with Melia azedarach, and their antifungal, antifeedant, and toxic activities. *J. Agric. Food Chem.* 60, 3424–31. doi:10.1021/jf300146n.

Li, Z., Peng, C., Shen, Y., Miao, X., Zhang, H., and Lin, H. (2008). l,l-Diketopiperazines from Alcaligenes faecalis A72 associated with South China Sea sponge Stelletta tenuis. *Biochem. Syst. Ecol.* 36, 230–234. doi:10.1016/j.bse.2007.08.007.

Lind, A. L., Wisecaver, J. H., Lameiras, C., Wiemann, P., Palmer, J. M., Keller, N. P., et al. (2017). Drivers of genetic diversity in secondary metabolic gene clusters within a fungal species. *PLoS Biol.* 15. doi:10.1371/journal.pbio.2003583.

Lind, A. L., Wisecaver, J. H., Smith, T. D., Feng, X., Calvo, A. M., and Rokas, A. (2015).

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.09.033902; this version posted April 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

978 Examining the evolution of the regulatory circuit controlling secondary metabolism and
979 development in the fungal genus Aspergillus. *PLoS Genet.* 11, e1005096.
980 doi:10.1371/journal.pgen.1005096.
981 Ma, Y. ., Li, Y., Liu, J. ., Song, Y. ., and Tan, R. . (2004). Anti-Helicobacter pylori metabolites
982 from Rhizoctonia sp. Cy064, an endophytic fungus in Cynodon dactylon. *Fitoterapia* 75,
983 451–456. doi:10.1016/j.fitote.2004.03.007.
984 Mead, M. E., Knowles, S. L., Raja, H. A., Beattie, S. R., Kowalski, C. H., Steenwyk, J. L., et al.
985 (2019a). Characterizing the Pathogenic, Genomic, and Chemical Traits of Aspergillus
986 fischeri , a Close Relative of the Major Human Fungal Pathogen Aspergillus fumigatus.
987 *mSphere* 4. doi:10.1128/mSphere.00018-19.
988 Mead, M. E., Raja, H. A., Steenwyk, J. L., Knowles, S. L., Oberlies, N. H., and Rokas, A.
989 (2019b). Draft Genome Sequence of the Griseofulvin-Producing Fungus Xylaria
990 flabelliformis Strain G536. *Microbiol. Resour. Announc.* 8. doi:10.1128/MRA.00890-19.
991 Moran, G. P., Coleman, D. C., and Sullivan, D. J. (2011). Comparative Genomics and the
992 Evolution of Pathogenicity in Human Pathogenic Fungi. *Eukaryot. Cell* 10, 34–42.
993 doi:10.1128/EC.00242-10.
994 Navarro-Muñoz, J. C., Selem-Mojica, N., Mullowney, M. W., Kautsar, S. A., Tryon, J. H.,
995 Parkinson, E. I., et al. (2020). A computational framework to explore large-scale
996 biosynthetic diversity. *Nat. Chem. Biol.* 16, 60–68. doi:10.1038/s41589-019-0400-9.
997 Nguyen, L.-T., Schmidt, H. A., von Haeseler, A., and Minh, B. Q. (2015). IQ-TREE: A Fast and
998 Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Mol.*
999 *Biol. Evol.* 32, 268–274. doi:10.1093/molbev/msu300.
1000 Pluskal, T., Castillo, S., Villar-Briones, A., and Orešič, M. (2010). MZmine 2: Modular

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.09.033902; this version posted April 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics* 11, 395. doi:10.1186/1471-2105-11-395.

Poukka, H., Aarnisalo, P., Santti, H., Jänne, O. A., and Palvimo, J. J. (2000). Coregulator Small Nuclear RING Finger Protein (SNURF) Enhances Sp1- and Steroid Receptor-mediated Transcription by Different Mechanisms. *J. Biol. Chem.* 275, 571–579. doi:10.1074/jbc.275.1.571.

Raffa, N., and Keller, N. P. (2019). A call to arms: Mustering secondary metabolites for success and survival of an opportunistic pathogen. *PLOS Pathog.* 15, e1007606. doi:10.1371/journal.ppat.1007606.

Raftery, A. E., and Lewis, S. M. (1995). The number of iterations, convergence diagnostics and generic Metropolis algorithms. *Pract. Markov Chain Monte Carlo* 7, 763–773. doi:10.1.1.41.6352.

Ries, L. N. A., Steenwyk, J. L., de Castro, P. A., de Lima, P. B. A., Almeida, F., de Assis, L. J., et al. (2019). Nutritional Heterogeneity Among Aspergillus fumigatus Strains Has Consequences for Virulence in a Strain- and Host-Dependent Manner. *Front. Microbiol.* 10. doi:10.3389/fmicb.2019.00854.

Rokas, A., Mead, M. E., Steenwyk, J. L., Oberlies, N. H., and Goldman, G. H. (2020a). Evolving moldy murderers: Aspergillus section Fumigati as a model for studying the repeated evolution of fungal pathogenicity. *PLOS Pathog.* 16, e1008315. doi:10.1371/journal.ppat.1008315.

Rokas, A., Mead, M. E., Steenwyk, J. L., Raja, H. A., and Oberlies, N. H. (2020b). Biosynthetic gene clusters and the evolution of fungal chemodiversity. *Nat. Prod. Rep.* doi:10.1039/C9NP00045C.

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.09.033902; this version posted April 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

1024 Rokas, A., Williams, B. L., King, N., and Carroll, S. B. (2003). Genome-scale approaches to
1025 resolving incongruence in molecular phylogenies. *Nature* 425, 798–804.
1026 doi:10.1038/nature02053.
1027 Rokas, A., Wisecaver, J. H., and Lind, A. L. (2018). The birth, evolution and death of metabolic
1028 gene clusters in fungi. *Nat. Rev. Microbiol.* doi:10.1038/s41579-018-0075-3.
1029 Samson, R. A., Hong, S., Peterson, S. W., Frisvad, J. C., and Varga, J. (2007). Polyphasic
1030 taxonomy of Aspergillus section Fumigati and its teleomorph Neosartorya. *Stud. Mycol.* 59,
1031 147–203. doi:10.3114/sim.2007.59.14.
1032 Santos, R. A. C. dos, Steenwyk, J. L., Rivero-Menendez, O., Mead, M. E., Silva, L. P., Bastos,
1033 R. W., et al. (2020). Genomic and phenotypic heterogeneity of clinical isolates of the
1034 human pathogens Aspergillus fumigatus, Aspergillus lentulus and Aspergillus
1035 fumigatiaffinis. *bioRxiv*, 2020.02.28.970384. doi:10.1101/2020.02.28.970384.
1036 Stanke, M., and Waack, S. (2003). Gene prediction with a hidden Markov model and a new
1037 intron submodel. *Bioinformatics* 19, ii215–ii225. doi:10.1093/bioinformatics/btg1080.
1038 Steenwyk, J. L., Shen, X.-X., Lind, A. L., Goldman, G. H., and Rokas, A. (2019). A Robust
1039 Phylogenomic Time Tree for Biotechnologically and Medically Important Fungi in the
1040 Genera Aspergillus and Penicillium. *MBio* 10. doi:10.1128/mBio.00925-19.
1041 Steenwyk, J., and Rokas, A. (2017). Extensive Copy Number Variation in Fermentation-Related
1042 Genes Among Saccharomyces cerevisiae Wine Strains. *G3 Genes, Genomes, Genet.* 7.
1043 Available at: http://www.g3journal.org/content/7/5/1475#ref-25 [Accessed July 3, 2017].
1044 Sugui, J. A., Pardo, J., Chang, Y. C., Zarember, K. A., Nardone, G., Galvez, E. M., et al. (2007).
1045 Gliotoxin Is a Virulence Factor of Aspergillus fumigatus : gliP Deletion Attenuates
1046 Virulence in Mice Immunosuppressed with Hydrocortisone. *Eukaryot. Cell* 6, 1562–1569.

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.09.033902; this version posted April 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

1047 doi:10.1128/EC.00141-07.

1048 Tavaré, S. (1986). Some probabilistic and statistical problems in the analysis of DNA sequences.

1049 *Lect. Math. life Sci.* 17, 57–86.

1050 Tekaia, F., and Latgé, J.-P. (2005). Aspergillus fumigatus: saprophyte or pathogen? *Curr. Opin.*

1051 *Microbiol.* 8, 385–392. doi:10.1016/j.mib.2005.06.017.

1052 Vallabhaneni, S., Mody, R. K., Walker, T., and Chiller, T. (2016). The Global Burden of Fungal

1053 Diseases. *Infect. Dis. Clin. North Am.* 30, 1–11. doi:10.1016/j.idc.2015.10.004.

1054 van Dongen, S. (2000). Graph clustering by flow simulation. *Graph Stimul. by flow Clust.* PhD

1055 thesis, University of Utrecht. doi:10.1016/j.cosrev.2007.05.001.

1056 Vinet, L., and Zhedanov, A. (2011). A 'missing' family of classical orthogonal polynomials. *J.*

1057 *Phys. A Math. Theor.* 44, 085201. doi:10.1088/1751-8113/44/8/085201.

1058 Wang, F.-Z., Li, D.-H., Zhu, T.-J., Zhang, M., and Gu, Q.-Q. (2011). Pseurotin A 1 and A 2 , two

1059 new 1-oxa-7-azaspiro[4.4]non-2-ene-4,6-diones from the holothurian-derived fungus

1060 Aspergillus fumigatus WFZ-25. *Can. J. Chem.* 89, 72–76. doi:10.1139/V10-157.

1061 Wang, F., Fang, Y., Zhu, T., Zhang, M., Lin, A., Gu, Q., et al. (2008). Seven new prenylated

1062 indole diketopiperazine alkaloids from holothurian-derived fungus Aspergillus fumigatus.

1063 *Tetrahedron* 64, 7986–7991. doi:10.1016/j.tet.2008.06.013.

1064 Waterhouse, R. M., Seppey, M., Simão, F. A., Manni, M., Ioannidis, P., Klioutchnikov, G., et al.

1065 (2018). BUSCO Applications from Quality Assessments to Gene Prediction and

1066 Phylogenomics. *Mol. Biol. Evol.* 35, 543–548. doi:10.1093/molbev/msx319.

1067 Waterhouse, R. M., Tegenfeldt, F., Li, J., Zdobnov, E. M., and Kriventseva, E. V. (2013).

1068 OrthoDB: a hierarchical catalog of animal, fungal and bacterial orthologs. *Nucleic Acids*

1069 *Res.* 41, D358–D365. doi:10.1093/nar/gks1116.

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.09.033902; this version posted April 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

1070 Weber, T., Blin, K., Duddela, S., Krug, D., Kim, H. U., Bruccoleri, R., et al. (2015). antiSMASH
1071 3.0—a comprehensive resource for the genome mining of biosynthetic gene clusters.
1072 *Nucleic Acids Res.* 43, W237–W243. doi:10.1093/nar/gkv437.
1073 Wiemann, P., Guo, C.-J., Palmer, J. M., Sekonyela, R., Wang, C. C. C., and Keller, N. P. (2013).
1074 Prototype of an intertwined secondary-metabolite supercluster. *Proc. Natl. Acad. Sci.* 110,
1075 17065–17070. doi:10.1073/pnas.1313258110.
1076 Yamada, A., Kataoka, T., and Nagai, K. (2000). The fungal metabolite gliotoxin:
1077 immunosuppressive activity on CTL-mediated cytotoxicity. *Immunol. Lett.* 71, 27–32.
1078 doi:10.1016/s0165-2478(99)00155-8.
1079 Yandell, M., and Ence, D. (2012). A beginner's guide to eukaryotic genome annotation. *Nat.*
1080 *Rev. Genet.* 13, 329–42. doi:10.1038/nrg3174.
1081 Yang, Z. (1994). Maximum likelihood phylogenetic estimation from DNA sequences with
1082 variable rates over sites: Approximate methods. *J. Mol. Evol.* 39, 306–314.
1083 doi:10.1007/BF00160154.
1084 Yang, Z. (1996). Among-site rate variation and its impact on phylogenetic analyses. *Trends Ecol.*
1085 *Evol.* 11, 367–372. doi:10.1016/0169-5347(96)10041-0.
1086 Yang, Z. (2007). PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Mol. Biol. Evol.* 24,
1087 1586–1591. doi:10.1093/molbev/msm088.
1088 Zhao, J., Mou, Y., Shan, T., Li, Y., Zhou, L., Wang, M., et al. (2010). Antimicrobial Metabolites
1089 from the Endophytic Fungus Pichia guilliermondii Isolated from Paris polyphylla var.
1090 yunnanensis. *Molecules* 15, 7961–7970. doi:10.3390/molecules15117961.
1091 Zhao, S., Latgé, J.-P., and Gibbons, J. G. (2019). Genome Sequences of Two Strains of the Food
1092 Spoilage Mold Aspergillus fischeri. *Microbiol. Resour. Announc.* 8.

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.09.033902; this version posted April 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

1093 doi:10.1128/MRA.01328-19.

1094