

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25

Drivers of genetic diversity in secondary metabolic gene clusters in a fungal population

Abigail L. Lind¹, Jennifer H. Wisecaver², Catarina Lameiras³, Fernando Rodrigues^{4,5}, Gustavo H. Goldman⁶, Antonis Rokas^{1,2}

1. Department of Biomedical Informatics, Vanderbilt University School of Medicine, Nashville, Tennessee, USA.

2. Department of Biology, Vanderbilt University, Nashville, Tennessee, USA.

3. Department of Microbiology, Portuguese Oncology Institute of Porto, Porto, Portugal

4. Life and Health Sciences Research Institute (ICVS), School of Medicine, University of Minho, Braga, Portugal

5. ICVS/3B's - PT Government Associate Laboratory, Braga/Guimarães, Portugal.

6. Faculdade de Ciências Farmacêuticas de Ribeirão Preto, Universidade de São Paulo, São Paulo, Brazil

†Corresponding author and lead contact: antonis.rokas@vanderbilt.edu

26 **Summary**

27
28 Filamentous fungi produce a diverse array of secondary metabolites (SMs) critical for defense,
29 virulence, and communication. The metabolic pathways that produce SMs are found in
30 contiguous gene clusters in fungal genomes, an atypical arrangement for metabolic pathways in
31 other eukaryotes. Comparative studies of filamentous fungal species have shown that SM gene
32 clusters are often either highly divergent or uniquely present in one or a handful of species,
33 hampering efforts to determine the genetic basis and evolutionary drivers of SM gene cluster
34 divergence. Here we examined SM variation in 66 cosmopolitan strains of a single species, the
35 opportunistic human pathogen *Aspergillus fumigatus*. Investigation of genome-wide
36 population-level variation showed that *A. fumigatus* strains contained five general types of
37 variation in SM gene clusters: non-functional gene polymorphisms, gene gain and loss
38 polymorphisms, whole cluster gain and loss polymorphisms, allelic polymorphisms where
39 different alleles corresponded to distinct, non-homologous clusters, and location
40 polymorphisms in which a cluster was found to differ in its genomic location across strains.
41 These polymorphisms affect the function of representative *A. fumigatus* SM gene clusters, such
42 as those involved in the production of gliotoxin, fumigaclavine, and helvolic acid, as well as the
43 function of clusters with undefined products. In addition to enabling the identification of
44 polymorphisms whose detection requires extensive genome-wide synteny conservation (e.g.,
45 mobile gene clusters and non-homologous cluster alleles), our population genomics approach
46 also implicated multiple underlying genetic drivers, including point mutations, recombination,
47 genomic deletion and insertion events, as well as horizontal gene transfer from distant fungi.
48 Finally, most of the population variants that we uncover have been previously hypothesized to
49 contribute to SM gene cluster diversity across entire fungal classes and phyla. We suggest that
50 the drivers of genetic diversity operating within a fungal population shown here are sufficient
51 to explain SM cluster macroevolutionary patterns.

52

53 **Keywords:** chemodiversity, specialized metabolism, genome evolution, genome architecture,
54 gene loss, genomic rearrangement

55 Introduction

56 Filamentous fungi produce a diverse array of small molecules that function as toxins,
57 antibiotics, and pigments [1]. Though by definition secondary metabolites (SMs) are not strictly
58 necessary for growth and development, they are critical to the lifestyle of filamentous fungi [2].
59 For example, antibiotic SMs gain give their fungal producers a competitive edge in
60 environments crowded with other microbes [3]. SMs can additionally mediate communication
61 between and within species, as well as contribute to virulence on animal and plant hosts in
62 pathogenic fungi [4,5].

63
64 A genomic hallmark of SMs in filamentous fungi is that the biosynthetic pathways that
65 produce them are typically organized into contiguous gene clusters in the genome [6]. These
66 gene clusters contain the chemical backbone synthesis genes whose enzymatic products
67 produce a core metabolite, such as non-ribosomal peptide synthases (NRPS) and polyketide
68 synthases (PKS), tailoring enzymes that chemically modify the metabolite, transporters involved
69 in product export, and transcription factors that control the expression of the clustered genes
70 [6]. Filamentous fungal genomes, particularly those in the phylum Ascomycota [6], typically
71 contain dozens of SM gene clusters. However, most individual SM gene clusters appear to be
72 either species-specific or narrowly taxonomically distributed in only a handful of species [6,7].
73 SM gene clusters that are more broadly distributed show discontinuous taxonomic distributions
74 and are often highly divergent between species. Consequently, the identity and total number of
75 SM gene clusters can vary widely even between very closely related species whose genomes
76 exhibit very high sequence and synteny conservation [8,9].

77
78 In the last decade, several comparative studies have described macroevolutionary
79 patterns of SM gene cluster diversity. For example, studies centered on genomic comparisons
80 of closely related species, such as members of the same genus, have identified several different
81 types of inter-species divergence, from single nucleotide substitutions (e.g., differences in
82 fumonisins produced by *Fusarium* species are caused by variants in one gene [10]), to gene gain
83 / loss events (e.g., the trichothecene gene clusters in *Fusarium* species and the aflatoxin family

84 SM gene clusters in *Aspergillus* species) [11–16], and genomic rearrangements (e.g., the
85 trichothecene gene clusters in *Fusarium*) [11]. Additionally, genetic and genomic comparisons
86 across fungal orders and classes have identified several instances of gene gain or loss [17–19]
87 and horizontal gene transfer [13,20–23] acting on individual genes or on entire gene clusters,
88 providing explanations for the diversity and discontinuity of the taxonomic distribution of
89 certain SM gene clusters across fungal species.

90
91 Although inter-species comparative studies have substantially contributed to our
92 understanding of SM diversity, the high levels of evolutionary divergence of SM clusters make
93 inference of the genetic drivers of SM gene cluster evolution challenging; simply put, it has
94 been difficult to “catch” the mechanisms that generate SM gene cluster variation “in the act” .
95 Several previous studies have examined intra-species or population-level differences in
96 individual SM gene clusters, typically focusing on the presence and frequency of non-functional
97 alleles of clusters involved in production of mycotoxins. Examples of clusters exhibiting such
98 polymorphisms include the gibberellin gene cluster in *Fusarium oxysporum* [24], the fumonisin
99 gene cluster in *Fusarium fujikuroi* [25], the aflatoxin and cyclopiazonic gene clusters in
100 *Aspergillus flavus* [26], and the bikaverin gene cluster in *Botrytis cinerea* [27]. While these
101 studies have greatly advanced our understanding of SM gene cluster genetic variation and
102 highlighted the importance of population-level analyses, studies examining the entirety of SM
103 gene cluster polymorphisms in fungal populations are so far lacking. We currently do not
104 understand the types and frequency of SM gene cluster polymorphisms in populations,
105 whether these polymorphisms affect all types of SM gene clusters, as well as the genetic drivers
106 of SM gene cluster evolution.

107
108 To address these questions, we investigated the genetic diversity of all 36 known and
109 predicted SM gene clusters in whole genome sequence data from 66 strains of the
110 opportunistic human pathogen *Aspergillus fumigatus*, 8 of which were sequenced in this study.
111 We found that 13 SM gene clusters were generally conserved and harbored low amounts of
112 variation. In contrast, the remaining 23 SM gene clusters were highly variable and contained

113 one or more of five different types of genetic variation: single-nucleotide polymorphisms
114 including nonsense and frameshift variants, individual gene gain and loss polymorphisms, entire
115 cluster gain and loss polymorphisms, polymorphisms associated with changes in cluster
116 genomic location, and clusters with non-homologous alleles resembling the idiomorphs of
117 fungal mating loci. Many clusters contained interesting combinations of these different
118 polymorphisms, such as pseudogenization in some strains and entire cluster loss in others. The
119 types of variants we find are likely generated by a combination of DNA replication and repair
120 errors, recombination, genomic insertions and deletions, and horizontal transfer. We
121 additionally find an enrichment for transposable elements (TEs) around horizontally transferred
122 clusters, clusters that change in genomic locations, and idiomorphic clusters. Taken together,
123 our results provide a guide to both the types of polymorphisms and the genetic drivers of SM
124 gene cluster diversification in filamentous fungi. As most of the genetic variants that we
125 observe have been previously associated with SM gene cluster diversity across much larger
126 evolutionary distances and timescales, we argue that population-level processes influencing SM
127 gene cluster diversity are sufficient to explain SM cluster macroevolutionary patterns.
128
129

130 **Results**

131 We analyzed the genomes of 66 globally distributed strains of *Aspergillus fumigatus* for
132 polymorphisms in SM gene clusters. We performed whole-genome sequencing on 8 strains, and
133 collected the remaining 58 strains from publicly available databases including NCBI Genome
134 and the NCBI Short Read Archive (Figure 1, Table S1) [28–32]. We analyzed all strains for
135 polymorphisms in 33 curated SM gene clusters present in the reference Af293 genome and
136 additionally searched for novel SM gene clusters (see Methods). These examinations revealed
137 five distinct types of polymorphisms which influence SM gene cluster variation (Table 1):

- 138 a) Single nucleotide and short indel polymorphisms. 33 / 33 SM gene clusters (present in
139 the reference Af293 strain) contained multiple genes with missense SNPs and short
140 indel variants in at least one strain in the population. 23 / 33 SM gene clusters contained
141 one or more genes with frameshift or nonsense variants in the population.
- 142 b) Gene content polymorphisms involving loss or gain of one or more genes. 6 / 33 SM
143 gene clusters contained a gene content polymorphism in the population.
- 144 c) Whole SM gene cluster gain and loss polymorphisms. 3 / 33 SM gene clusters present in
145 the genome of the reference Af293 strain were absent in at least one strain in the
146 population and an additional 3 previously unknown SM gene clusters were present in
147 the population.
- 148 d) Idiomorphic polymorphisms. One locus contained multiple non-homologous SM gene
149 cluster alleles in different strains of the population.
- 150 e) Genomic location polymorphisms. 2 / 33 SM gene clusters were found in different
151 genomic locations (e.g., different chromosomes) between strains.

152

153 **Single-nucleotide and indel polymorphisms**

154 It is well established that single nucleotide polymorphisms (SNPs) and short indel
155 polymorphisms are caused by errors in DNA replication and repair, and are a major source of
156 genomic variation [33]. Non-synonymous SNPs and indels with missense, frameshift, and
157 nonsense effects were widespread across the 33 SM reference gene clusters (Table S2). Every
158 strain contained numerous missense mutations and at least one nonsense or frameshift

159 mutation in its SM gene clusters. Although missense mutations are likely to influence SM
160 production, the functional effects of nonsense and frameshift mutations are comparatively
161 easier to infer from genomic sequence data because they often lead to truncated proteins
162 lacking a significant portion of their amino acid sequence. For example, a frameshift mutation in
163 the polyketide synthase (PKS) of the tryptacidin gene cluster in the A1163 strain results in loss of
164 tryptacidin production [34]. Interestingly, we identified a premature stop codon (Gln273*) in a
165 transcription factor required for tryptacidin production, *tpcD*, in a strain sequenced in this study
166 (MO79587EXP). These data suggest that function of this SM gene cluster has been lost through
167 at least two independent genetic events in *A. fumigatus*.

168

169 Individual nonsense or frameshift variants ranged from very common in the population
170 to rarer variants present in one or a handful of strains. For example, the non-ribosomal peptide
171 synthase (NRPS) *pes3* gene (Afu5g12730) in SM gene cluster 21 harbors 16 nonsense or
172 frameshift polymorphisms in 55 strains. Seven of these polymorphisms are common and
173 present in 10 or more strains, while seven are rarer and found in 5 or fewer strains. Strains with
174 lab-mutated null alleles of the *pes3* gene are more virulent than strains with functional copies
175 [35], which may explain the widespread occurrence of null *pes3* alleles in the *Aspergillus*
176 population.

177

178 **Gene content polymorphisms**

179 We additionally identified several SM gene clusters that gained or lost genes in some
180 strains. These gene content polymorphisms were most likely generated through genomic
181 deletion or insertion events and were often present in high frequencies in the population (Table
182 1). In three cases, these polymorphisms impact backbone synthesis genes, rendering the SM
183 gene cluster non-functional. One example involves SM gene cluster 14, whose standard
184 composition includes a pyoverdine synthase gene, an NRPS-like gene, an NRPS backbone gene,
185 and several additional modification genes (Figure 2). We discovered that 4 / 66 strains lack an
186 11-kb region on the 3' end of the cluster which normally contains an NRPS gene and two
187 additional cluster genes, and the first non-SM genes on the 3' end flanking the cluster. All *A.*

188 *fumigatus* strains contain a *copia* family transposable element [36] at the 3' end of the cluster,
189 suggesting that transposable elements may have been involved in the generation of this
190 polymorphism. While this polymorphism could have arisen through a deletion event, a
191 homologous cluster lacking the 11-kb region is also present in the reference genomes of
192 *Aspergillus lentulus* and *Aspergillus fischerianus*, close relatives of *A. fumigatus* (Figure 2). The
193 most parsimonious explanation is that the genome of the *A. fumigatus* ancestor contained an
194 SM gene cluster that lacked the 11-kb region, and that this genomic region was subsequently
195 gained and increased in frequency in the *A. fumigatus* population.

196

197 Two additional gene content polymorphisms affecting SM backbone genes were
198 restricted to one strain each and appear to have arisen through genomic deletion events.
199 Specifically, strain IF1SWF4 lacks an 8-Kb region near the helvolic acid SM gene cluster,
200 resulting in the loss of the backbone oxidosqualene cyclase gene as well an upstream region
201 containing two non-SM genes (Figure S1A). Strain LMB35Aa lacks a 54-kb region on the end of
202 chromosome 2, which includes five genes from the telomere-proximal fumigaclavine C cluster
203 (Figure S1B).

204

205 In three other cases, gene content polymorphisms involved gene loss or truncation
206 events of non-backbone structural genes. We found that the second half of the ORF of the *gliM*
207 *O*-methyltransferase gene in the gliotoxin gene cluster has been lost in 2 / 66 strains (Figure
208 S1C), that the first half of the permease *fmqE* in the fumiquinazoline gene cluster has been lost
209 in 4 / 66 strains (Figure S1D), and that an ABC transporter gene in SM cluster 21 has been
210 almost entirely lost in 21 / 66 strains (Figure S1E-F).

211

212 **Whole gene cluster gain and loss polymorphisms**

213 Several SM gene clusters were gained or lost entirely in 13 / 66 strains. We observed
214 instances where a cluster present in the genome of the reference Af293 strain was absent or
215 pseudogenized in other strains as well as cases in which SM clusters present in other strains
216 were absent from the reference Af293 strain.

217 The most notable example of an SM gene cluster that was present in the Af293
218 reference genome but absent or pseudogenized in others was SM cluster 4. This cluster
219 contains 5 genes on the tip of the Af293 chromosome 1 and contains orthologs to five of the six
220 genes in the fusarielin gene cluster in *Fusarium graminearum* [37]. This cluster is also present in
221 several other *Aspergillus* species, including *A. clavatus* and *A. niger* [37]. Phylogenetic analysis
222 of the genes in this SM gene cluster is consistent with horizontal gene transfer between fungi in
223 the class Sordariomycetes and fungi in the class Eurotiomycetes, or alternatively with extensive
224 gene loss in both Sordariomycetes and Eurotiomycetes (Figure S2). This gene cluster is entirely
225 absent in 4 / 66 strains, and its genes are undergoing pseudogenization in an additional 44
226 strains via multiple independent mutational events (Figure 3A). Specifically, 19 strains shared a
227 single frameshift variant in the polyketide synthase gene (4380_4381insAATGGGCT; frameshift
228 at Glu1461 in Afu1g17740) and an additional 13 strains shared a single frameshift variant
229 (242delG; frameshift at Gly81) in an aldose 1-epimerase gene (Afu1g17723). Twelve other
230 strains each contained one to several frameshift or nonsense polymorphisms involving nine
231 unique mutational sites, suggesting that this pathway is undergoing multiple independent
232 pseudogenization events. Five of these strains contained multiple distinct frameshifts and
233 premature stop codons in more than one gene in the cluster, indicating that the entire pathway
234 is pseudogenized in these strains.

235
236 By searching for novel SM gene clusters in the genomes of the other 65 *A. fumigatus*
237 strains, we found three SM gene clusters that were absent from the genome of the Af293
238 reference strain. As SM gene clusters are often present in repeat-rich and subtelomeric regions
239 that are challenging to assemble [38,39], these strains might harbor additional novel SM gene
240 clusters.

241
242 One of the novel SM gene clusters that we identified, cluster 34, was present in all but
243 two of the strains (Af293 and F7763). This cluster contains a PKS backbone gene, one PKS-like
244 gene with a single PKS associated domain, nine genes with putative biosynthetic functions
245 involved in secondary metabolism, and six hypothetical proteins (Figure 3B). The two strains

246 that lack this cluster contain a likely non-functional cluster fragment that includes the PKS-like
247 gene, two biosynthetic genes, and three hypothetical proteins. Interestingly, the 3' region
248 flanking this cluster is syntenic across all 66 strains but the 5' region is not, suggesting that a
249 recombination or deletion event may have resulted in the loss of this cluster in the Af293 and
250 F7763 strains.

251
252 The other SM gene clusters that were absent from the Af293 genome are present at
253 lower frequencies in the population; cluster 35 is present in 2 / 66 strains and cluster 36 in 4 /
254 66 strains. Cluster 35 is located in a region syntenic with an Af293 chromosome 4 region and is
255 flanked on both sides by transposable elements (Figure S3). Eight of the 14 genes in this SM
256 gene cluster are homologous to genes in an SM gene cluster in the genome of the insect
257 pathogenic fungus *Metarhizium anisopliae* (Figure S3). Phylogenetic analysis of these 8 genes is
258 consistent with a horizontal transfer event (Figure S4). Cluster 36 is an NRPS containing cluster
259 located on genomic scaffolds that lack homology to either the Af293 or A1163 genomes,
260 making it impossible to determine on which chromosome this cluster is located (Figure S3). The
261 evolutionary histories of the genes in the cluster are consistent with vertical inheritance and are
262 present in multiple *Aspergillus* species.

263

264 **Idiomorph polymorphisms**

265 One of the most peculiar types of polymorphisms that we identified is a locus containing
266 different unrelated alleles of SM gene clusters, reminiscent of the idiomorph alleles at the
267 fungal mating loci [40]. This locus, which resides on chromosome 3 and corresponds to cluster
268 10 in the Af293 genome (Figure 4), was previously described as being strain-specific in a
269 comparison between Af293 and A1163 [29] and is thought to reside in a recombination hot
270 spot [30]. Our analysis showed that this locus contained at least 6 different alleles present in
271 two or more of the 66 strains as well as 2 additional alleles that were each only present in one
272 strain (Figure S5).

273

274 In the Af293 reference genome, the cluster present at this locus contains one full-length
275 PKS gene along with multiple genes that contain NRPS- or PKS-associated domains (Allele C). In
276 the A1163 reference genome and 17 other strains, there is a full-length NRPS and a full-length
277 PKS (Allele B). These alleles show an almost complete lack of sequence similarity except for a
278 conserved hypothetical protein and a fragment of the full-length A1163 PKS in the Af293 allele;
279 in contrast, the upstream and downstream flanking regions of the two alleles, which do not
280 contain any backbone genes, are syntenic. Remarkably, another allele, present in 12 strains,
281 contains all of the genes from both the Af293 and A1163 clusters (Allele D). The remaining
282 three alleles contain various combinations of these genes. One allele found in 22 strains
283 contains some A1163-specific genes and no Af293-specific genes (Allele A), while another allele
284 found in 3 strains contains some Af293-specific genes but no A1163 genes (Allele F). The final
285 allele, present in 8 strains, contains the entire Af293 allele as well as part of the A1163 allele
286 (Allele E). Every allele is littered with long terminal repeat sequence fragments from *gypsy* and
287 *copia* TE families as well as with sequence fragments from DNA transposons from the *mariner*
288 family [36]. In some cases, these TEs correspond with breakpoints in synteny between alleles,
289 suggesting that the diverse alleles of this SM gene cluster may arise via TE-driven
290 recombination. Further, both of the alleles that are restricted to a single strain had an insertion
291 event of several genes near a TE, though the rest of the locus is highly similar to one of the
292 more common alleles (Figure S5). The evolutionary history of this highly diverse locus is unclear.
293 While it is tempting to speculate that the largest allele containing all observed genes represents
294 the ancestral state, it does not explain the presence of a shared hypothetical protein and PKS
295 gene fragment between the Af293 locus (Allele C) and the A1163 locus (Allele B).

296

297 **Genomic location polymorphisms**

298 The final type of polymorphism that we observed is associated with SM gene clusters
299 that are located in different genomic locations in different strains, suggesting that these SM
300 gene clusters are behaving like mobile elements. This type of polymorphism was observed in
301 SM gene clusters 1 and 33, both of which produce as yet identified products, and are present at
302 low frequencies in the population.

303

304 SM gene cluster 1, which is present in six strains at three different genomic locations
305 (Figure 5A), consists of a PKS and four other modification genes that are always flanked by a 15
306 Kb region (upstream) and a 43 Kb region (downstream) containing TEs. In the reference Af293
307 strain and in strain F7763, this SM gene cluster and its flanking regions are located on
308 chromosome 1, while in strains dutch3, F13619, and Z5 they are located between Afu4g07320
309 and Afu4g07340 on chromosome 4. In contrast, in strain JCM_10253, the cluster and flanking
310 regions are located on chromosome 8 immediately adjacent to the 3' end of the intertwined
311 fumagillin and pseurotin SM gene supercluster [41].

312

313 In 5 / 6 strains, the cluster appears to be functional and does not contain nonsense SNPs
314 or indels. However, the cluster found on chromosome 1 in strain F7763 contains two stop
315 codons in the oxidoreductase gene (Gln121* and Gln220*) and two premature stop codons in
316 the polyketide synthase (Gln1156* and Gln1542*), suggesting this strain contains a null allele.

317

318 This “jumping” gene cluster is not present in any other sequenced *Aspergillus* genus,
319 and phylogenetic analysis of its constituent genes is consistent with horizontal gene transfer
320 between fungi (Figure S6). Specifically, this gene cluster is also present in *Phaeosphaeria*
321 *nodorum*, a plant pathogen from the class Dothideomycetes, *Pseudogymnoascus pannorum*, a
322 fungus isolated from permafrost from the Leotiomycetes, *Escovopsis weberi*, and a fungal
323 parasite of fungus-growing ants from the Sordariomycetes (Figure 5B). One additional species,
324 the endophyte *Hypoxylon* sp. CI4A from the class Sordariomycetes, contains four of the five
325 cluster genes but is missing Afu1g00970, an MFS drug transporter. However, this species
326 contains an unrelated gene annotated as an MFS drug transporter immediately adjacent to this
327 cluster, so this species may be using a different transporter (Figure 5B). None of these fungi
328 contain the upstream or downstream TE-rich flanking regions present in *A. fumigatus*, and each
329 fungus contains additional unique genes with putative biosynthetic functions adjacent to the
330 transferred cluster. The most likely explanation for this change in flanking regions is that this

331 SM gene cluster was transferred into *A. fumigatus* once and has subsequently moved its
332 location in the genome.

333

334 The second SM gene cluster that shows variation in its genomic location across strains,
335 cluster 33, contains a terpene synthase. This cluster is present in only 5 strains at 3 distinct
336 locations (Figure S7). Similar to cluster 1, this cluster is also flanked by TEs and in one strain the
337 clusters is located 58 Kb from another SM gene cluster. In contrast to cluster 1, this cluster does
338 not appear to have been horizontally transferred between fungi and its genes are present in
339 other sequenced *Aspergillus* species. As this mobile cluster was not horizontally transferred, it
340 is possible that the horizontal transfer of cluster 1 and its mobile nature throughout the *A.*
341 *fumigatus* genome are driven by different mechanisms.

342

343 Discussion

344 Our examination of 66 genomes from strains of *Aspergillus fumigatus* revealed that five
345 general types of polymorphisms describe variation in SM gene clusters in a fungal population.
346 These polymorphisms include variation in single nucleotides, gene and gene cluster gains and
347 losses, non-homologous clusters at the same genomic position, and changes in genomic
348 locations of clusters (Figure 6). In several cases, the genetic mechanisms that gave rise to these
349 polymorphisms and their functional consequences are clear. Polymorphisms in single-
350 nucleotides most likely arose from errors in DNA replication or repair, while gene loss
351 polymorphisms likely arose from genomic deletion or recombination events. Others, such as SM
352 gene clusters that exist as non-homologous alleles or that entered the population through
353 horizontal transfer, could arise through multiple genetic mechanisms. Transposable elements
354 are common in polymorphic SM gene clusters and are found flanking mobile and horizontally
355 transferred clusters, as well as in regions adjacent to non-homologous alleles and where gene
356 gain has occurred, suggesting they contribute to SM gene cluster diversity. Using a population
357 genomics approach to identify SM gene cluster variants allowed us to also capture and describe
358 novel polymorphisms, including mobile gene clusters and idiomorphic clusters, which are
359 difficult to identify in comparative genomic studies between species whose conservation of
360 genome-wide synteny is low. Strikingly, the variants and genetic drivers we observe at the
361 population level are also implicated as driving SM gene cluster variation between fungal
362 species, suggesting that the observed microevolutionary processes are sufficient to explain
363 macroevolutionary patterns of SM gene cluster evolution. Below, we discuss our key results and
364 place them in the broader context of SM gene cluster evolution.

365
366 The first novel type of polymorphism was observed in SM gene clusters 1 and 33, which
367 by occupying different genomic locations in different strains, appear to behave in a manner
368 similar to mobile genetic elements. Interestingly, both clusters are located near or immediately
369 adjacent to other SM gene clusters in some strains. For example, cluster 1 is located
370 immediately adjacent to the intertwined fumagillin and pseurotin supercluster [41] in one
371 strain. This supercluster is regulated by the transcriptional factor *fapR* and is located in a

372 chromosomal region controlled by the master SM regulators *laeA* and *veA* [41,42], raising the
373 hypothesis that these mobile gene clusters might be co-opting the regulatory machinery
374 already in place. Previous work has hypothesized that the fumagillin and pseurotin supercluster
375 formed through genomic rearrangement events placing the once-independent gene clusters in
376 close proximity to each other [41]. Our observation that this mobile gene cluster is located in
377 this same region not only supports this hypothesis but also implicates transposable elements as
378 one of the mechanisms by which such genomic rearrangements are formed. These
379 superclusters may also represent an intermediate stage in the formation of new SM gene
380 clusters. Supercluster formation, potentially mediated by mobile gene clusters, and followed
381 by gene loss, could explain macroevolutionary patterns of SM gene clusters have shown that
382 clustered genes in one species can be dispersed over multiple gene clusters in other species
383 [9,11].

384

385 The second novel type of polymorphism was the presence of multiple non-homologous
386 alleles at the cluster 10 locus, echoing the structure of the idiomorphic mating type locus [40].
387 This region has previously been reported to be a recombination hotspot in *A. fumigatus* [30]
388 and all alleles contain numerous transposable elements. Thus, it is possible that polymorphism
389 at this locus originated via SM gene cluster fusion or splitting events driven by transposable
390 elements. Interestingly, two other previously described instances of SM gene cluster variation
391 bear close resemblance to the *A. fumigatus* idiomorphic SM gene cluster 10 locus. The first is
392 the presence of two non-homologous *Aspergillus flavus* alleles, where some strains contain a 9-
393 gene sesquiterpene-like SM gene cluster and others contain a non-homologous 6-gene SM gene
394 cluster at the same genomic location [43]. The second is the presence of two non-homologous
395 SM gene clusters at the same, well-conserved, locus in a comparison of six species of
396 dermatophyte fungi [44]. Based on these results, we hypothesize that idiomorphic clusters may
397 be common in fungal populations and contribute to the broad diversity of SM gene clusters
398 across filamentous fungi.

399

400 The remaining types of SM polymorphism in the *A. fumigatus* population have
401 previously been described at the species level. For example, null alleles have been reported for
402 many individual SM gene clusters across diverse fungal species [10,17,26,27] and our results
403 show that it is a widespread phenomenon, affecting two-thirds of the *A. fumigatus* SM gene
404 clusters. Consistent with previous literature reporting the presence of additional SM gene
405 clusters not present in the reference strain [24,25], we also identify four low-frequency SM
406 gene clusters in the *A. fumigatus* population, two of which are horizontally transferred from
407 distantly related fungi and two that appear vertically inherited. We find numerous cases of
408 gene loss and relatively fewer cases of gene gain; previous studies have implicated these
409 processes across numerous types of SM gene clusters and reported their effects on metabolite
410 production [45]. Previous work has indicated that gene and gene cluster gain and loss tends to
411 occur near telomeres, suggesting that higher rates of genetic events like recombination lead to
412 loss [46]. While we do find several cases of gene and gene cluster loss in subtelomeric clusters,
413 including the fumigaclavine gene cluster and the helvolic acid gene cluster, we also find gene
414 loss in clusters that are not located near telomeres, such as the gliotoxin gene cluster and the
415 fumiquinazoline gene cluster. These findings suggest that while gene and gene cluster loss can
416 occur in telomeric regions, clusters in other genomic regions also experience high rates of loss.

417
418 Previous work has demonstrated that SM gene clusters, like the metabolites that they
419 produce, are highly divergent between fungal species. Our examination of genome-wide
420 variation shows that these SM gene clusters are correspondingly diverse within individual
421 strains of a single fungal species. Furthermore, the observed genetic changes in SM gene
422 clusters are widespread across different types of gene clusters and are caused by many
423 underlying genetic drivers, including gene and gene cluster gain, loss, non-homologous cluster
424 alleles, and mobile gene clusters. The net effect of these substitutions, gains, losses, and
425 rearrangements raises the hypothesis that fungal SM gene clusters are likely in a state of
426 evolutionary flux, constantly altering their SM gene cluster repertoire, and consequently
427 modifying and diversifying their chemodiversity.

428

429 **Methods**

430 **Strains analyzed**

431 Eight strains of *A. fumigatus* were isolated from four patients with recurrent cases of
432 aspergillosis in the Portuguese Oncology Institute in Porto, Portugal. Each strain was
433 determined to be *A. fumigatus* using macroscopic features of the culture and microscopic
434 morphology observed in the slide preparation from the colonies with lactophenol solution [47].
435 Based on the morphological characterization, all clinical strains were classified as *A. fumigatus*
436 *complex*-Fumigati. The genomes of all eight strains were sequenced using 150bp Illumina
437 paired-end sequence reads at the Genomic Services Lab of Hudson Alpha (Huntsville, Alabama,
438 USA). Genomic libraries were constructed with the Illumina TruSeq library kit and sequenced on
439 an Illumina HiSeq 2500 sequencer. Samples of all eight strains were sequenced at greater than
440 180X coverage or depth (Table S1). Short read sequences for these 8 strains are available in the
441 Short Read Archive under accession SRP109032.

442

443 In addition to the 8 strains sequenced in this study, we retrieved 58 *A. fumigatus* strains
444 with publicly available whole genome sequencing data, resulting in a population genomics
445 dataset of 66 strains (Table S1). The strains used included both environmental and clinical
446 strains and were isolated from multiple continents. Genome assemblies for 10 of these strains,
447 including the Af293 and A1163 reference strains, were available for download from GenBank
448 [28–32,48]. For 6 of these strains, short read sequences were also available from the NCBI Short
449 Read Archive (SRA), which were used for variant discovery only (see Single nucleotide variant
450 (SNV) and indel discovery) and not for genome assembly. Short read sequences were not
451 available for the remaining 4 strains. Short read sequences were downloaded for an additional
452 48 strains from the Short Read Archive if they were sequenced with paired-end reads and at
453 greater than 30x coverage.

454

455 **Single nucleotide variant (SNV) and indel discovery**

456 All strains with available short read data (62 of 66 strains) were aligned to both the
457 Af293 and A1163 reference genomes using BWA mem version 0.7.12-r1044 [49]. Coverage of
458 genes present in the reference genome was calculated using bedtools v2.25.0 [50]. SNV and
459 indel discovery and genotyping was performed relative to the Af293 reference genome and was
460 conducted across all samples simultaneously using the Genome Analysis Toolkit version 3.5-0-
461 g36282e4 with recommended hard filtering parameters [51–53] and annotated using snpEff
462 version 4.2 [54].

463

464 ***De novo* genome assembly and gene annotation**

465 All 56 strains without publicly available genome assemblies were *de novo* assembled
466 using the iWGS pipeline [55]. Specifically, all strains were assembled using SPAdes v3.6.2 and
467 MaSuRCA v3.1.3 and resulting assemblies were evaluated using QUAST v3.2 [56–58]. The
468 average N50 of assemblies constructed with this strategy was 463 KB (Table S1). Genes were
469 annotated in these assemblies as well as in five GenBank assemblies with no predicted genes
470 using augustus v3.2.2 trained on *A. fumigatus* gene models [59]. Repetitive elements were
471 annotated in all assemblies using RepeatMasker version open-4.0.6 [60].

472

473 **Secondary metabolic gene cluster annotation and discovery**

474 Secondary metabolic gene clusters in the Af293 reference genome were taken from two
475 recent reviews, both of which considered computational and experimental data to delineate
476 cluster boundaries [61,62] (Table S3). The genomes of the other 65 strains were scanned for
477 novel SM gene clusters using identified using antiSMASH v3.0.5.1 [63]. To prevent potential
478 assembly errors from confounding the analysis, any inference about changes in genomic
479 locations of genes or gene clusters was additionally verified by manually inspecting alignments
480 and ensuring that paired end reads supported an alternative genomic location (see SNV and
481 indel discovery). Cases where paired end reads did not support the change in genomic location
482 or where mapping was ambiguous or low quality were discarded.

483

484 **Phylogenetic analysis**

485 To construct a SNP-based strain phylogeny, biallelic SNPs with no missing data were
486 pruned using SNPRelate v1.8.0 with a linkage disequilibrium threshold of 0.8 [64]. A phylogeny
487 was constructed using RAxML v8.0.25 using the ASC_BINGAMMA substitution model [65]. The
488 tree was midpoint rooted and all branches with bootstrap support less than 80% were
489 collapsed.

490
491 To understand the evolutionary histories of specific SM gene clusters showing unusual
492 taxonomic distributions, we reconstructed the phylogenetic trees of their SM genes.
493 Specifically, SM cluster protein sequences were queried against a local copy of the NCBI non-
494 redundant protein database (downloaded May 30, 2017) using phmmer, a member of the
495 HMMER3 software suite [66] using acceleration parameters --F1 1e-5 --F2 1e-7 --F3 1e-10. A
496 custom perl script sorted the phmmer results based on the normalized bitscore (nbs), where
497 nbs was calculated as the bitscore of the single best-scoring domain in the hit sequence divided
498 by the best bitscore possible for the query sequence (i.e., the bitscore of the query aligned to
499 itself). No more than five hits were retained for each unique NCBI Taxonomy ID. Full-length
500 proteins corresponding to the top 100 hits ($E\text{-value} < 1 \times 10^{-10}$) to each query sequence were
501 extracted from the local database using esl-sfetch [66]. Sequences were aligned with MAFFT
502 v7.310 using the E-INS-i strategy and the BLOSUM30 amino acid scoring matrix [67] and
503 trimmed with trimAL v1.4.rev15 using its gappyout strategy [68]. The topologies were inferred
504 using maximum likelihood as implemented in RAxML v8.2.9 [65] using empirically determined
505 substitution models and rapid bootstrapping (1000 replications). The phylogenies were
506 midpoint rooted and branches with less than 80% bootstrap support were collapsed using the
507 ape and phangorn R packages [69,70]. Phylogenies were visualized using ITOL version 3.0 [71].
508

509

510 **Acknowledgements**

511 This work was supported in part by the National Science Foundation (DEB-1442113 to A.R.) and
512 by the Northern Portugal Regional Operational Programme (NORTE 2020), under the Portugal
513 2020 Partnership Agreement, through the European Regional Development Fund (FEDER)
514 (NORTE-01-0145-FEDER-000013 to F.R.). G.H.G. thanks the Fundação de Amparo à Pesquisa do
515 Estado de São Paulo (FAPESP) and Conselho Nacional de Desenvolvimento Científico e
516 Tecnológico (CNPq), both from Brazil, for funding and support. This work was conducted in part
517 using the resources of the Advanced Computing Center for Research and Education at
518 Vanderbilt University (<http://www.accre.vanderbilt.edu/>).

519

520 **Author contributions**

521 Conceptualization, A.R., G.H.G., A.L.L.; Methodology, A.L.L., J.H.W, C.L.; Investigation, A.L.L.;
522 Visualization, A.L.L., J.H.W; Resources, G.H.G., F.R., C.L.; Writing, A.L.L, A.R.

523

524 **Table 1. Types and rates of SM gene cluster variants in *A. fumigatus* strains.** See also Figure S3
525 and S4.

526

| Description | Phenotype | Drivers | Frequency at cluster level | Frequency at strain level | Previous reports |
|--|---|--|--|-----------------------------|--|
| Single-nucleotide polymorphisms and indels | Potential for protein function change (missense); abrogation of protein function (nonsense and frameshift) | DNA replication errors; relaxation of purifying selection | 100% (33/33 clusters; missense); 70% (23/33 clusters; nonsense and frameshift) | Every strain affected | Bikaverin in <i>Botrytis</i> [17,27], aflatoxin in <i>Aspergillus oryzae</i> and <i>Aspergillus flavus</i> [26], fumonisins in <i>Fusarium</i> [10], many others |
| Gene content polymorphisms | Loss of gene cluster function; structural changes in the metabolite; change in cluster expression or metabolite transport | Deletion and insertion events; recombination; transposable elements | 6 clusters | 27 / 66 strains | Trichothecene in <i>Fusarium</i> , aflatoxin and sterigmatocystin in <i>Aspergillus</i> [11–15] |
| Whole gene cluster polymorphisms | Loss or gain of novel metabolites | Deletion and insertion events; horizontal gene transfer; transposable elements | 6 clusters | 13 / 66 strains | Gibberellin and fumonisin in <i>Fusarium</i> [24,25] |
| fCluster idiomorphs | Changes in metabolites produced or structure of metabolites | Transposable elements; recombination; other mechanisms? | 1 gene cluster | 8 unique identified alleles | Putative SM gene clusters in dermatophytes; putative SM gene cluster in <i>Aspergillus flavus</i> and <i>Aspergillus oryzae</i> [43,44] |
| Mobile gene clusters | Potential for change in gene regulation | Transposable elements; horizontal gene transfer; other mechanisms? | 2 gene clusters | 8 / 66 strains | None |

527

528

529

530 **Table S1. Summary of strains, sequence data, and assemblies used.**

531

532 **Table S2. All nonsynonymous variants in SM gene cluster genes.**

533

534 **Table S3. Description of reference Af293 SM gene clusters.**

535

536 **References**

- 537 1. Vining LC. Functions of secondary metabolites. *Annu Rev Microbiol. Annual Reviews* 4139
538 El Camino Way, P.O. Box 10139, Palo Alto, CA 94303-0139, USA; 1990;44: 395–427.
539 doi:10.1146/annurev.mi.44.100190.002143
- 540 2. Schimek C. Evolution of Special Metabolism in Fungi : Concepts , Mechanisms , and
541 Pathways. In: Pöggler S, Wöstmeyer J, editors. *Evolution of Fungi and Fungal-Like*
542 *Organisms, The Mycota. XIV.* Berlin, Heidelberg: Springer-Verlag; 2011. pp. 293–328.
- 543 3. Fox EM, Howlett BJ. Secondary metabolism: regulation and role in fungal biology. *Curr*
544 *Opin Microbiol.* 2008;11: 481–7. doi:10.1016/j.mib.2008.10.007
- 545 4. Scharf DH, Heinekamp T, Brakhage AA. Human and Plant Fungal Pathogens: The Role of
546 Secondary Metabolites. Heitman J, editor. *PLoS Pathog.* Public Library of Science;
547 2014;10: e1003859. doi:10.1371/journal.ppat.1003859
- 548 5. Yim G, Wang HH, Davies J. Antibiotics as signalling molecules. *Philos Trans R Soc Lond B*
549 *Biol Sci.* 2007;362: 1195–200. doi:10.1098/rstb.2007.2044
- 550 6. Keller NP, Turner G, Bennett JW. Fungal secondary metabolism - from biochemistry to
551 genomics. *Nat Rev Microbiol.* 2005;3: 937–47. doi:10.1038/nrmicro1286
- 552 7. Bennett J, Bentley R. What’s in a name?—Microbial secondary metabolism. *Adv Appl*
553 *Microbiol.* 1989;34.
- 554 8. Khaldi N, Seifuddin FT, Turner G, Haft D, Nierman WC, Wolfe KH, et al. SMURF: Genomic
555 mapping of fungal secondary metabolite clusters. *Fungal Genet Biol.* 2010;47: 736–41.
556 doi:10.1016/j.fgb.2010.06.003
- 557 9. Lind AL, Wisecaver JH, Smith TD, Feng X, Calvo AM, Rokas A. Examining the evolution of
558 the regulatory circuit controlling secondary metabolism and development in the fungal
559 genus *Aspergillus*. *PLoS Genet.* Public Library of Science; 2015;11: e1005096.
560 doi:10.1371/journal.pgen.1005096
- 561 10. Proctor RH, Busman M, Seo J-A, Lee YW, Plattner RD. A fumonisin biosynthetic gene
562 cluster in *Fusarium oxysporum* strain O-1890 and the genetic basis for B versus C
563 fumonisin production. *Fungal Genet Biol.* 2008;45: 1016–1026.
564 doi:10.1016/j.fgb.2008.02.004
- 565 11. Proctor RH, McCormick SP, Alexander NJ, Desjardins AE. Evidence that a secondary
566 metabolic biosynthetic gene cluster has grown by gene relocation during evolution of the
567 filamentous fungus *Fusarium*. *Mol Microbiol.* 2009;74: 1128–42. doi:10.1111/j.1365-
568 2958.2009.06927.x
- 569 12. Berry DB, Guan Q, Hose J, Haroon S, Gebbia M, Heisler LE, et al. Multiple Means to the
570 Same End: The Genetic Basis of Acquired Stress Resistance in Yeast. *Copenhaver GP,*

- 571 editor. PLoS Genet. Public Library of Science; 2011;7: e1002353.
572 doi:10.1371/journal.pgen.1002353
- 573 13. Slot JC, Rokas A. Horizontal transfer of a large and highly toxic secondary metabolic gene
574 cluster between fungi. *Curr Biol*. Elsevier Ltd; 2011;21: 134–9.
575 doi:10.1016/j.cub.2010.12.020
- 576 14. Ehrlich KC, Chang P-K, Yu J, Cotty PJ. Aflatoxin Biosynthesis Cluster Gene *cypA* Is Required
577 for G Aflatoxin Formation. *Appl Environ Microbiol*. 2004;70: 6518–6524.
578 doi:10.1128/AEM.70.11.6518-6524.2004
- 579 15. Carbone I, Ramirez-Prado JH, Jakobek JL, Horn BW. Gene duplication, modularity and
580 adaptation in the evolution of the aflatoxin gene cluster. *BMC Evol Biol*. 2007;7: 111.
581 doi:10.1186/1471-2148-7-111
- 582 16. Yu J, Chang P-K, Ehrlich KC, Cary JW, Bhatnagar D, Cleveland TE, et al. Clustered pathway
583 genes in aflatoxin biosynthesis. *Appl Environ Microbiol*. American Society for
584 Microbiology; 2004;70: 1253–62. doi:10.1128/AEM.70.3.1253-1262.2004
- 585 17. Campbell M a, Rokas A, Slot JC. Horizontal transfer and death of a fungal secondary
586 metabolic gene cluster. *Genome Biol Evol*. 2012;4: 289–93. doi:10.1093/gbe/evs011
- 587 18. Kroken S, Glass NL, Taylor JW, Yoder OC, Turgeon BG. Phylogenomic analysis of type I
588 polyketide synthase genes in pathogenic and saprobic ascomycetes. *Proc Natl Acad Sci U*
589 *S A*. 2003;100: 15670–5. doi:10.1073/pnas.2532165100
- 590 19. Bushley KE, Turgeon BG. Phylogenomics reveals subfamilies of fungal nonribosomal
591 peptide synthetases and their evolutionary relationships. *BMC Evol Biol*. 2010;10: 26.
592 doi:10.1186/1471-2148-10-26
- 593 20. Patron NJ, Waller RF, Cozijnsen AJ, Straney DC, Gardiner DM, Nierman WC, et al. Origin
594 and distribution of epipolythiodioxopiperazine (ETP) gene clusters in filamentous
595 ascomycetes. *BMC Evol Biol*. 2007;7: 174. doi:10.1186/1471-2148-7-174
- 596 21. Khaldi N, Collemare J, Lebrun M-H, Wolfe KH. Evidence for horizontal transfer of a
597 secondary metabolite gene cluster between fungi. *Genome Biol*. BioMed Central; 2008;9:
598 R18. doi:10.1186/gb-2008-9-1-r18
- 599 22. Khaldi N, Wolfe KH. Evolutionary Origins of the Fumonisin Secondary Metabolite Gene
600 Cluster in *Fusarium verticillioides* and *Aspergillus niger*. *Int J Evol Biol*. Hindawi Publishing
601 Corporation; 2011;2011: 423821. doi:10.4061/2011/423821
- 602 23. Reynolds HT, Slot JC, Divon HH, Lysøe E, Proctor RH, Brown DW. Differential Retention of
603 Gene Functions in a Secondary Metabolite Cluster. *Mol Biol Evol*. Oxford University Press
604 (OUP); 10: e1004816. doi:10.1093/molbev/msx145
- 605 24. Wiemann P, Sieber CMK, von Bargen KW, Studt L, Niehaus E-M, Espino JJ, et al.
606 Deciphering the Cryptic Genome: Genome-wide Analyses of the Rice Pathogen *Fusarium*

- 607 fujikuroi Reveal Complex Regulation of Secondary Metabolism and Novel Metabolites.
608 Mitchell AP, editor. PLoS Pathog. Public Library of Science; 2013;9: e1003475.
609 doi:10.1371/journal.ppat.1003475
- 610 25. Chiara M, Fanelli F, Mulè G, Logrieco AF, Pesole G, Leslie JF, et al. Genome Sequencing of
611 Multiple Isolates Highlights Subtelomeric Genomic Diversity within *Fusarium fujikuroi*.
612 Genome Biol Evol. 2015;7: 3062–3069. doi:10.1093/gbe/evv198
- 613 26. Chang P-K, Horn BW, Dorner JW. Sequence breakpoints in the aflatoxin biosynthesis gene
614 cluster and flanking regions in nonaflatoxigenic *Aspergillus flavus* isolates. Fungal Genet
615 Biol. 2005;42: 914–923. doi:10.1016/j.fgb.2005.07.004
- 616 27. Schumacher J, Gautier A, Morgant G, Studt L, Ducrot P-H, Le Pêcheur P, et al. A
617 Functional Bikaverin Biosynthesis Gene Cluster in Rare Strains of *Botrytis cinerea* Is
618 Positively Controlled by VELVET. Yun S-H, editor. PLoS One. Public Library of Science;
619 2013;8: e53729. doi:10.1371/journal.pone.0053729
- 620 28. Nierman WC, Pain A, Anderson MJ, Wortman JR, Kim HS, Arroyo J, et al. Genomic
621 sequence of the pathogenic and allergenic filamentous fungus *Aspergillus fumigatus*.
622 Nature. 2005;438: 1151–6. doi:10.1038/nature04332
- 623 29. Fedorova ND, Khaldi N, Joardar VS, Maiti R, Amedeo P, Anderson MJ, et al. Genomic
624 islands in the pathogenic filamentous fungus *Aspergillus fumigatus*. Richardson PM,
625 editor. PLoS Genet. Public Library of Science; 2008;4: e1000046.
626 doi:10.1371/journal.pgen.1000046
- 627 30. Abdolrasouli A, Rhodes J, Beale MA, Hagen F, Rogers TR, Chowdhary A, et al. Genomic
628 Context of Azole Resistance Mutations in *Aspergillus fumigatus* Determined Using
629 Whole-Genome Sequencing. MBio. 2015;6. doi:10.1128/mBio.00536-15
- 630 31. Knox BP, Blachowicz A, Palmer JM, Romsdahl J, Huttenlocher A, Wang CCC, et al.
631 Characterization of *Aspergillus fumigatus* Isolates from Air and Surfaces of the
632 International Space Station. mSphere. 2016;1.
- 633 32. Paul S, Zhang A, Ludeña Y, Villena GK, Yu F, Sherman DH, et al. Insights from the genome
634 of a high alkaline cellulase producing *Aspergillus fumigatus* strain obtained from Peruvian
635 Amazon rainforest. J Biotechnol. 2017;251: 53–58. doi:10.1016/j.jbiotec.2017.04.010
- 636 33. Roberts JD, Kunkel TA. Fidelity of DNA replication. In: DePamphilis ML, editor. DNA
637 replication in eukaryotic cells. Cold Spring Harbor, New York: Cold Spring Harbor
638 Laboratory Press; 1996. pp. 217–247.
- 639 34. Throckmorton K, Lim FY, Kontoyiannis DP, Zheng W, Keller NP. Redundant synthesis of a
640 conidial polyketide by two distinct secondary metabolite clusters in *Aspergillus*
641 *fumigatus*. Environ Microbiol. 2015; doi:10.1111/1462-2920.13007
- 642 35. O’Hanlon KA, Cairns T, Stack D, Schrettl M, Bignell EM, Kavanagh K, et al. Targeted

- 643 Disruption of Nonribosomal Peptide Synthetase *pes3* Augments the Virulence of
644 *Aspergillus fumigatus*. *Infect Immun*. 2011;79: 3978–3992. doi:10.1128/IAI.00192-11
- 645 36. Kapitonov V V., Jurka J. A universal classification of eukaryotic transposable elements
646 implemented in Repbase. *Nat Rev Genet*. Nature Publishing Group; 2008;9: 411–412.
647 doi:10.1038/nrg2165-c1
- 648 37. Sørensen JL, Hansen FT, Sondergaard TE, Staerk D, Lee TV, Wimmer R, et al. Production
649 of novel fusarielins by ectopic activation of the polyketide synthase 9 cluster in *Fusarium*
650 *graminearum*. *Environ Microbiol*. Blackwell Publishing Ltd; 2012;14: 1159–1170.
651 doi:10.1111/j.1462-2920.2011.02696.x
- 652 38. Treangen TJ, Salzberg SL. Repetitive DNA and next-generation sequencing: computational
653 challenges and solutions. *Nat Rev Genet*. NIH Public Access; 2011;13: 36–46.
654 doi:10.1038/nrg3117
- 655 39. Palmer JM, Keller NP. Secondary metabolism in fungi: does chromosomal location
656 matter? *Curr Opin Microbiol*. 2010;13: 431–6. doi:10.1016/j.mib.2010.04.008
- 657 40. Metzzenberg RL, Glass NL. Mating type and mating strategies in *Neurospora*. *BioEssays*.
658 Wiley Subscription Services, Inc., A Wiley Company; 1990;12: 53–59.
659 doi:10.1002/bies.950120202
- 660 41. Wiemann P, Guo C-J, Palmer JM, Sekonyela R, Wang CCC, Keller NP. Prototype of an
661 intertwined secondary-metabolite supercluster. *Proc Natl Acad Sci U S A*. 2013;110:
662 17065–70. doi:10.1073/pnas.1313258110
- 663 42. Lin H-C, Chooi Y-H, Dhingra S, Xu W, Calvo AM, Tang Y. The Fumagillin Biosynthetic Gene
664 Cluster in *Aspergillus fumigatus* Encodes a Cryptic Terpene Cyclase Involved in the
665 Formation of β -trans-Bergamotene. *J Am Chem Soc*. 2013;135: 4616–9.
666 doi:10.1021/ja312503y
- 667 43. Gibbons JG, Salichos L, Slot JC, Rinker DC, McGary KL, King JG, et al. The evolutionary
668 imprint of domestication on genome variation and function of the filamentous fungus
669 *Aspergillus oryzae*. *Curr Biol*. 2012;22: 1403–9. doi:10.1016/j.cub.2012.05.033
- 670 44. Zhang H, Rokas A, Slot JCI. Two different secondary metabolism gene clusters occupied
671 the same ancestral locus in fungal dermatophytes of the Arthrodermataceae. *PLoS One*.
672 2012;7: e41903. doi:10.1371/journal.pone.0041903
- 673 45. Schardl CL, Young CA, Hesse U, Amyotte SG, Andreeva K, Calie PJ, et al. Plant-Symbiotic
674 Fungi as Chemical Engineers: Multi-Genome Analysis of the Clavicipitaceae Reveals
675 Dynamics of Alkaloid Loci. Heitman J, editor. *PLoS Genet*. Public Library of Science;
676 2013;9: e1003323. doi:10.1371/journal.pgen.1003323
- 677 46. Young CA, Schardl CL, Panaccione DG, Florea S, Takach JE, Charlton ND, et al. Genetics,
678 genomics and evolution of ergot alkaloid diversity. *Toxins*. 2015. pp. 1273–1302.

- 679 doi:10.3390/toxins7041273
- 680 47. de Hoog G, Guarro J, Gené J, Figueras M. Atlas of Clinical Fungi. Washington, DC: ASM
681 Press; 2001.
- 682 48. Liu D, Zhang R, Yang X, Wu H, Xu D, Tang Z, et al. Thermostable cellulase production of
683 *Aspergillus fumigatus* Z5 under solid-state fermentation and its application in
684 degradation of agricultural wastes. *Int Biodeterior Biodegradation*. 2011;65: 717–725.
685 doi:10.1016/j.ibiod.2011.04.005
- 686 49. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform.
687 *Bioinformatics*. 2009;25: 1754–1760. doi:10.1093/bioinformatics/btp324
- 688 50. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic
689 features. *Bioinformatics*. Oxford University Press; 2010;26: 841–2.
690 doi:10.1093/bioinformatics/btq033
- 691 51. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome
692 Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing
693 data. *Genome Res*. Cold Spring Harbor Laboratory Press; 2010;20: 1297–303.
694 doi:10.1101/gr.107524.110
- 695 52. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, del Angel G, Levy-Moonshine A, et
696 al. From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best
697 Practices Pipeline. *Current Protocols in Bioinformatics*. Hoboken, NJ, USA: John Wiley &
698 Sons, Inc.; 2013. p. 11.10.1-11.10.33. doi:10.1002/0471250953.bi1110s43
- 699 53. DePristo MA, Banks E, Poplin R, Garimella K V, Maguire JR, Hartl C, et al. A framework for
700 variation discovery and genotyping using next-generation DNA sequencing data. *Nat*
701 *Genet*. 2011;43: 491–498. doi:10.1038/ng.806
- 702 54. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. A program for
703 annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in
704 the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)*. Taylor &
705 Francis; 2012;6: 80–92. doi:10.4161/fly.19695
- 706 55. Zhou X, Peris D, Kominek J, Kurtzman CP, Hittinger CT, Rokas A. in silico Whole Genome
707 Sequencer & Analyzer (iWGS): A Computational Pipeline to Guide the Design and
708 Analysis of de novo Genome Sequencing Studies. *G3 & #58; Genes|Genomes|Genetics*. 2016; doi:10.1534/g3.116.034249
- 710 56. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new
711 genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol*.
712 Mary Ann Liebert, Inc.; 2012;19: 455–77. doi:10.1089/cmb.2012.0021
- 713 57. Zimin A V., Marcais G, Puiu D, Roberts M, Salzberg SL, Yorke JA. The MaSuRCA genome
714 assembler. *Bioinformatics*. 2013;29: 2669–2677. doi:10.1093/bioinformatics/btt476

- 715 58. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUASt: quality assessment tool for genome
716 assemblies. *Bioinformatics*. 2013;29: 1072–1075. doi:10.1093/bioinformatics/btt086
- 717 59. Stanke M, Morgenstern B. AUGUSTUS: a web server for gene prediction in eukaryotes
718 that allows user-defined constraints. *Nucleic Acids Res. Oxford University Press*; 2005;33:
719 W465-7. doi:10.1093/nar/gki458
- 720 60. Smit A, Hubley R, Green P. Repeatmasker Open-4.0 [Internet]. Available:
721 <http://www.repeatmasker.org>
- 722 61. Inglis DO, Binkley J, Skrzypek MS, Arnaud MB, Cerqueira GC, Shah P, et al.
723 Comprehensive annotation of secondary metabolite biosynthetic genes and gene
724 clusters of *Aspergillus nidulans*, *A. fumigatus*, *A. niger* and *A. oryzae*. *BMC Microbiol.*
725 2013;13: 91. doi:10.1186/1471-2180-13-91
- 726 62. Bignell E, Cairns TC, Throckmorton K, Nierman WC, Keller NP. Secondary metabolite
727 arsenal of an opportunistic pathogenic fungus. *Philos Trans R Soc B Biol Sci*. 2016;371.
- 728 63. Medema MH, Blin K, Cimermancic P, de Jager V, Zakrzewski P, Fischbach M a, et al.
729 antiSMASH: rapid identification, annotation and analysis of secondary metabolite
730 biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Res.*
731 2011;39: W339-46. doi:10.1093/nar/gkr466
- 732 64. Zheng X, Levine D, Shen J, Gogarten SM, Laurie C, Weir BS. A high-performance
733 computing toolset for relatedness and principal component analysis of SNP data.
734 *Bioinformatics. Oxford University Press*; 2012;28: 3326–3328.
735 doi:10.1093/bioinformatics/bts606
- 736 65. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large
737 phylogenies. *Bioinformatics*. 2014;30: 1312–1313. doi:10.1093/bioinformatics/btu033
- 738 66. Eddy SR. A new generation of homology search tools based on probabilistic inference.
739 *Genome Inform*. 2009;23: 205–11.
- 740 67. Katoh K, Standley DM. MAFFT Multiple Sequence Alignment Software Version 7:
741 Improvements in Performance and Usability. *Mol Biol Evol*. 2013;30: 772–780.
742 doi:10.1093/molbev/mst010
- 743 68. Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T. trimAl: a tool for automated
744 alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*. 2009;25: 1972–
745 1973. doi:10.1093/bioinformatics/btp348
- 746 69. Paradis E, Claude J, Strimmer K. APE: Analyses of Phylogenetics and Evolution in R
747 language. *Bioinformatics. Oxford University Press*; 2004;20: 289–290.
748 doi:10.1093/bioinformatics/btg412
- 749 70. Schliep KP. phangorn: phylogenetic analysis in R. *Bioinformatics. Oxford University Press*;
750 2011;27: 592–3. doi:10.1093/bioinformatics/btq706

751 71. Letunic I, Bork P. Interactive tree of life (iTOL) v3: an online tool for the display and
752 annotation of phylogenetic and other trees. *Nucleic Acids Res.* 2016;44: W242–W245.
753 doi:10.1093/nar/gkw290

754

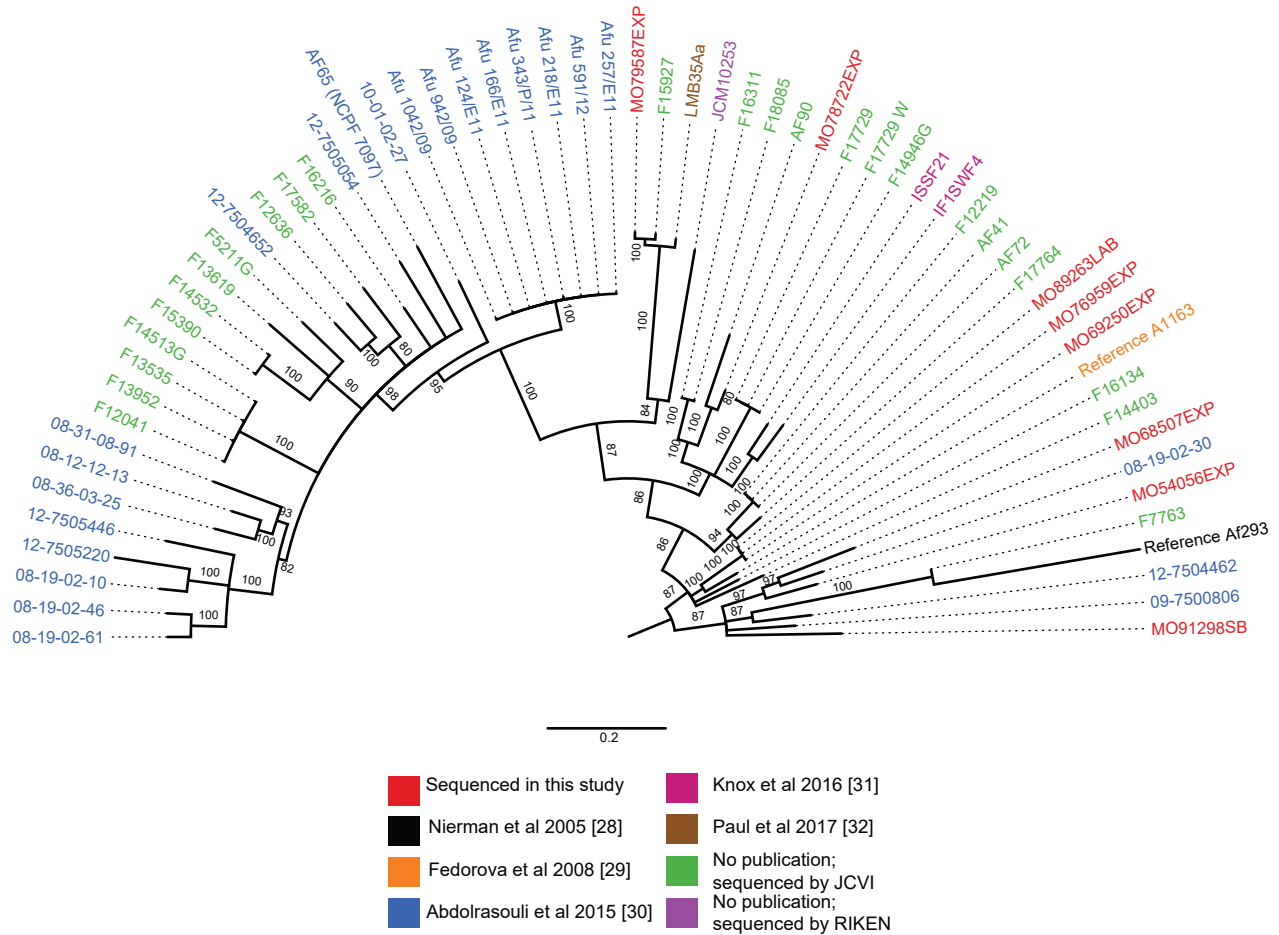


Figure 1. SNP-based phylogeny of *A. fumigatus* strains. The phylogeny was constructed using biallelic SNPs with no missing data. The tree is midpoint rooted and all branches with bootstrap support less than 80% are collapsed.

Reference & most strains

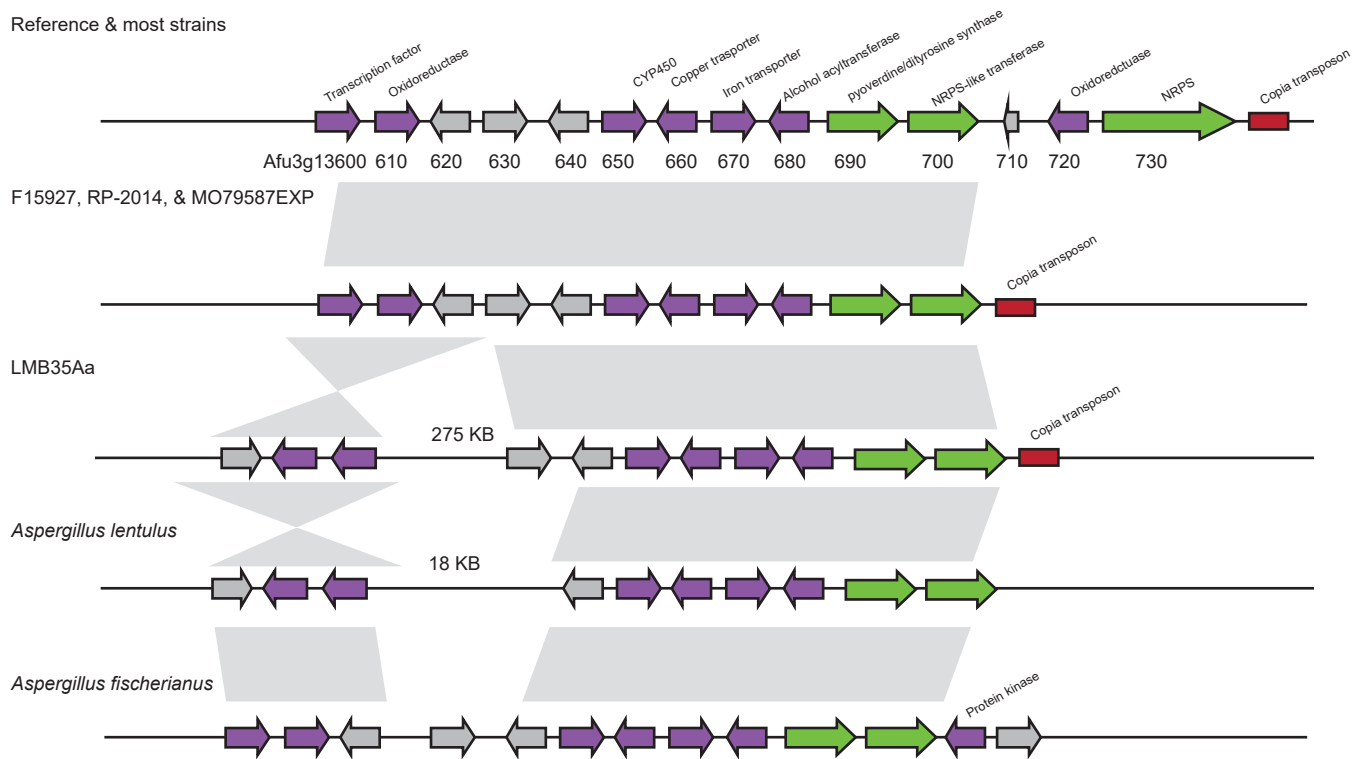


Figure 2. Differences in gene content in SM gene cluster 14 in *A. fumigatus* strains and closely related species. Four *A. fumigatus* strains lack an 11-Kb region in this cluster, including an NRPS backbone gene. Regions upstream and downstream of this cluster are syntenic. LMB35Aa also contains a large inversion that moves a transcription factor, oxidoreductase, and hypothetical protein 275 kb away from the cluster. *Aspergillus fischerianus* and *Aspergillus lentulus*, close relatives of *A. fumigatus*, contain a cluster lacking the 11-kb region.

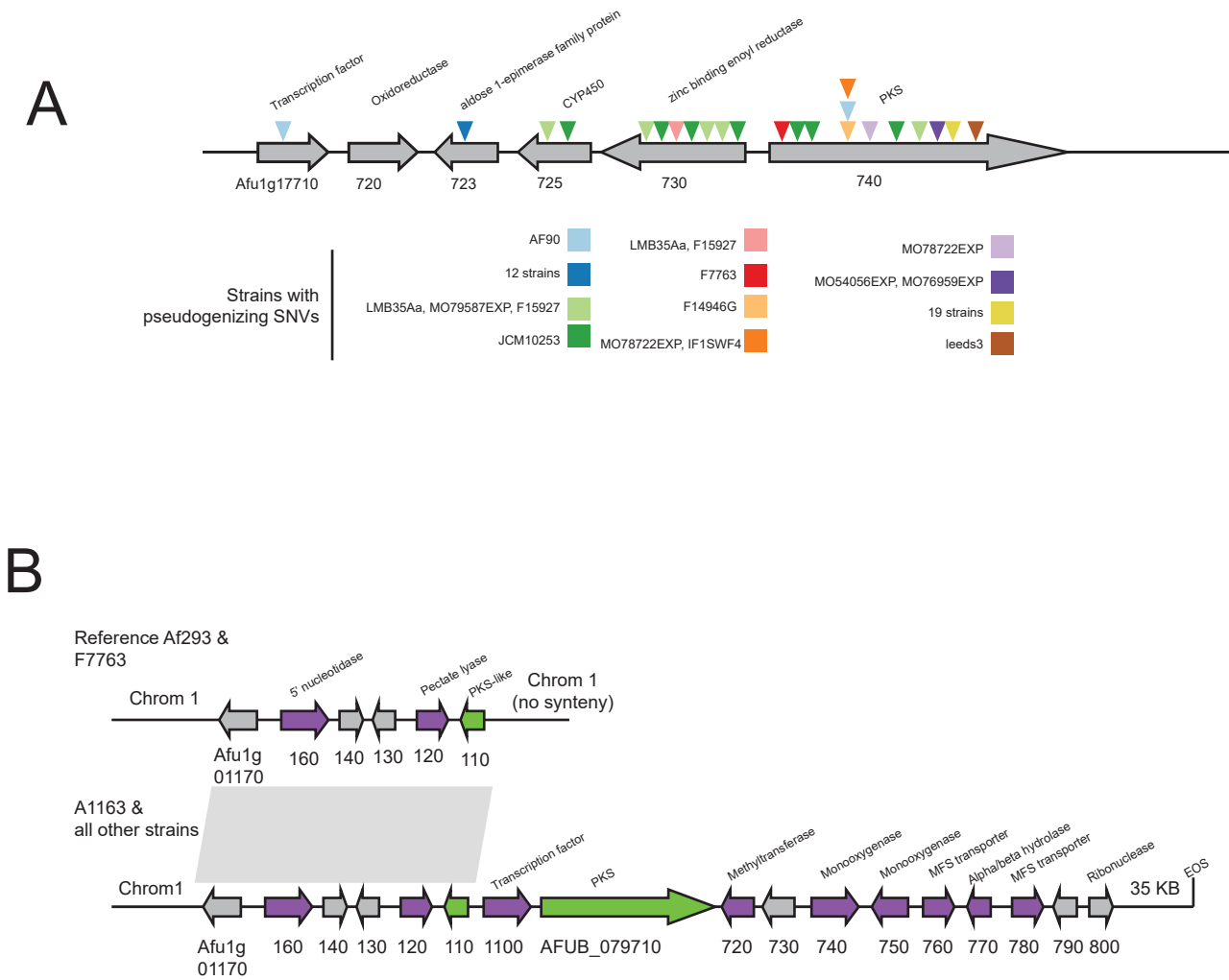


Figure 3. Pseudogenization and gene loss in SM gene clusters. A) SM gene cluster found in most *A. fumigatus* strains but absent from the Af293 reference and from the F7763 strain. EOS denotes end of scaffold. B) Positions of frameshift variants and nonsense variants in the fusarielin-like SM gene cluster 4.

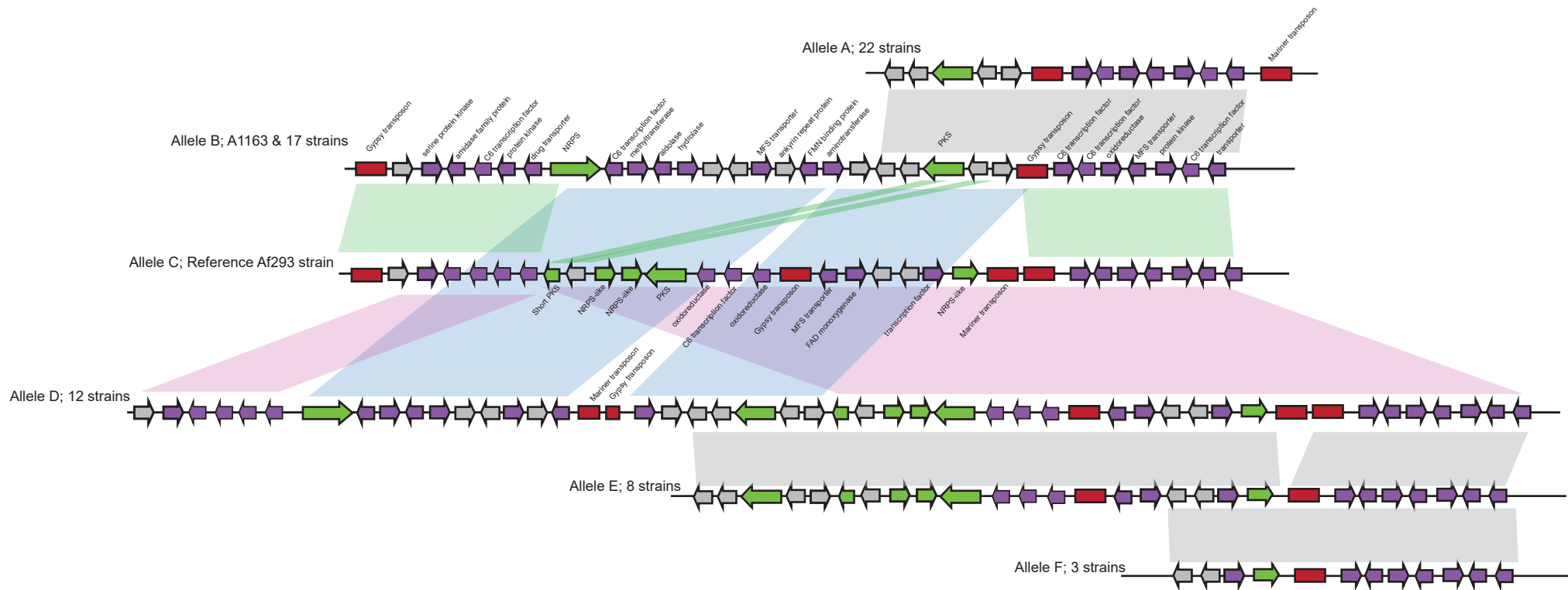


Figure 4. Six alleles of an idiomorphic SM gene cluster. Alleles of SM gene cluster 10 on chromosome 3. Red boxes denote transposable elements. Green arrows denote backbone genes (PKS or NRPS) or genes containing domains associated with PKS or NRPS genes. Purple arrows denote genes likely involved in SM biosynthesis, transport, or cluster regulation. Gray arrows denote genes without a putative function or with functions unrelated to SM.

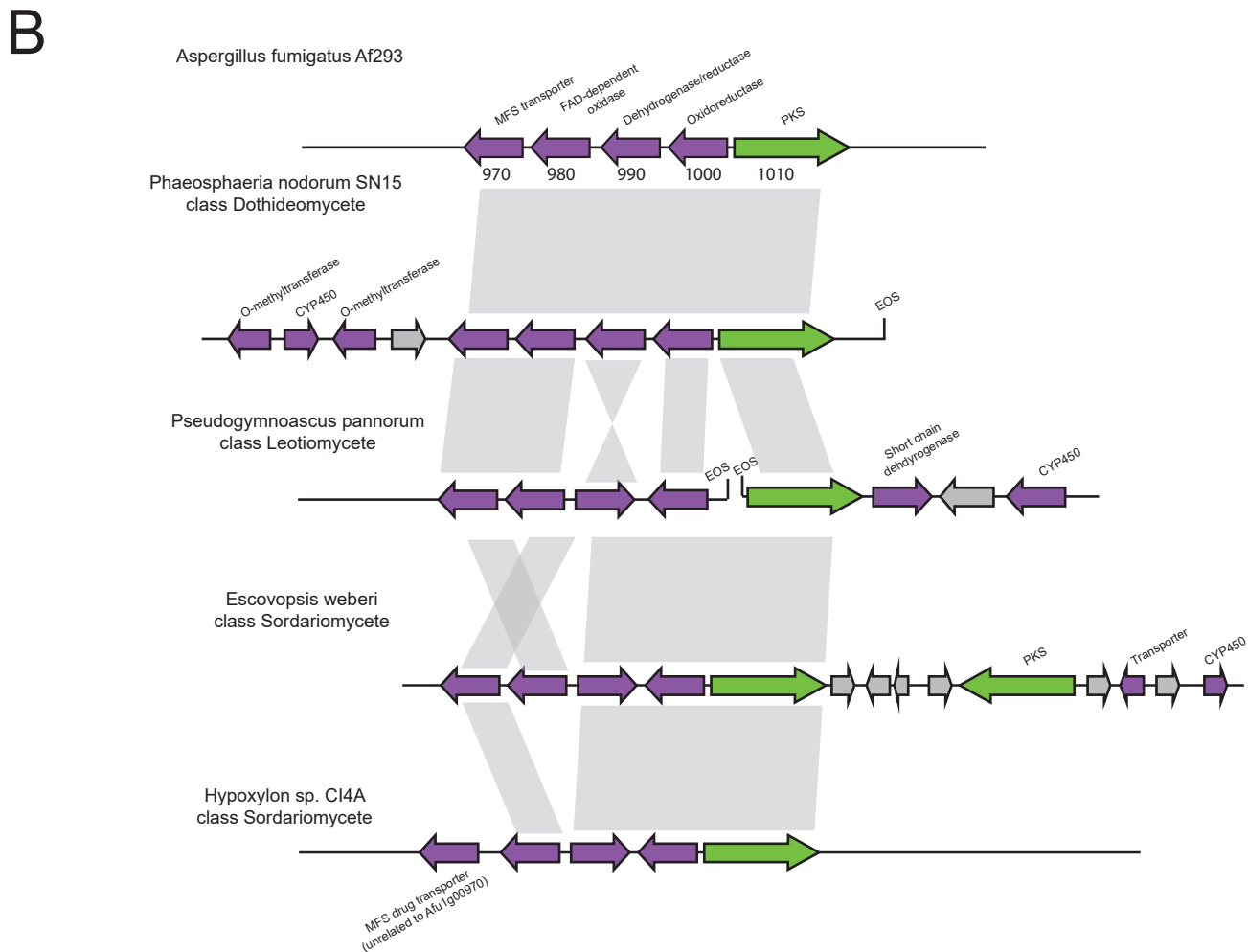
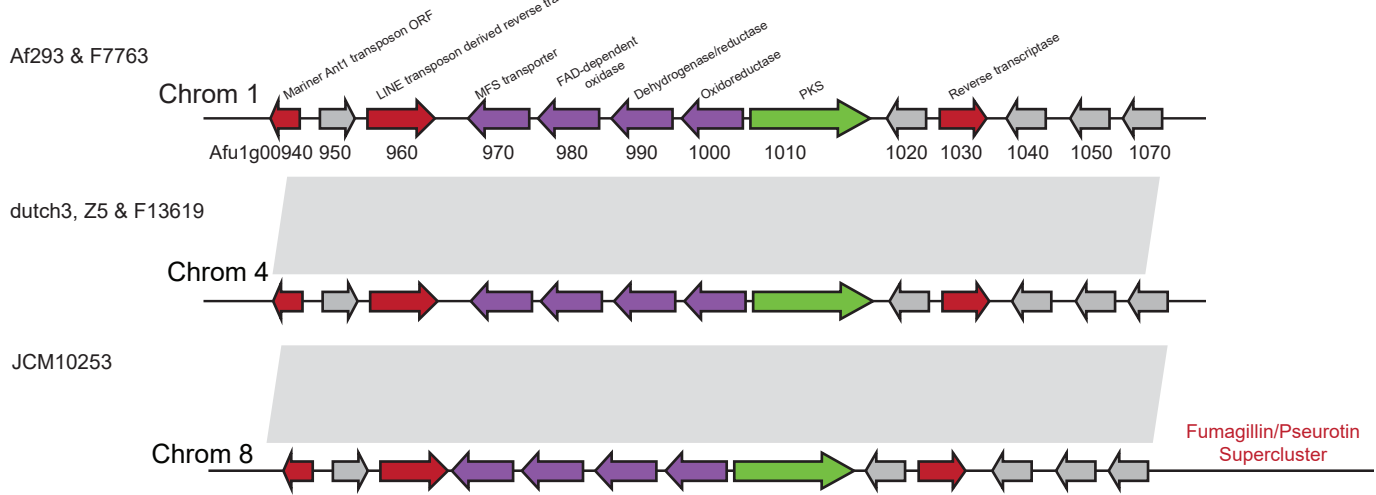


Figure 5. Multiple genomic locations of a horizontally transferred SM gene cluster.

A) Genomic location of SM gene cluster 1 (Afu1g00970-01010) and flanking region in all strains. This cluster is on chromosome 1 in two strains, chromosome 4 in three strains, and adjacent to the intertwined fumagillin and pseurotin supercluster on chromosome 8 in one strain. The flanking regions contain transposon-derived open reading frames including two putative reverse transcriptases.

B) Synteny of *A. fumigatus* SM gene cluster 1 with clusters in *Phaeosphaeria nodorum*, *Pseudogymnoascus pannorum*, *Escovopsis weberi*, and *Hypoxylon sp. CI4A*. EOS denotes end of scaffold. All species contain non-syntenic genes predicted by antiSMASH to be part of a biosynthetic gene cluster.

High frequency polymorphisms

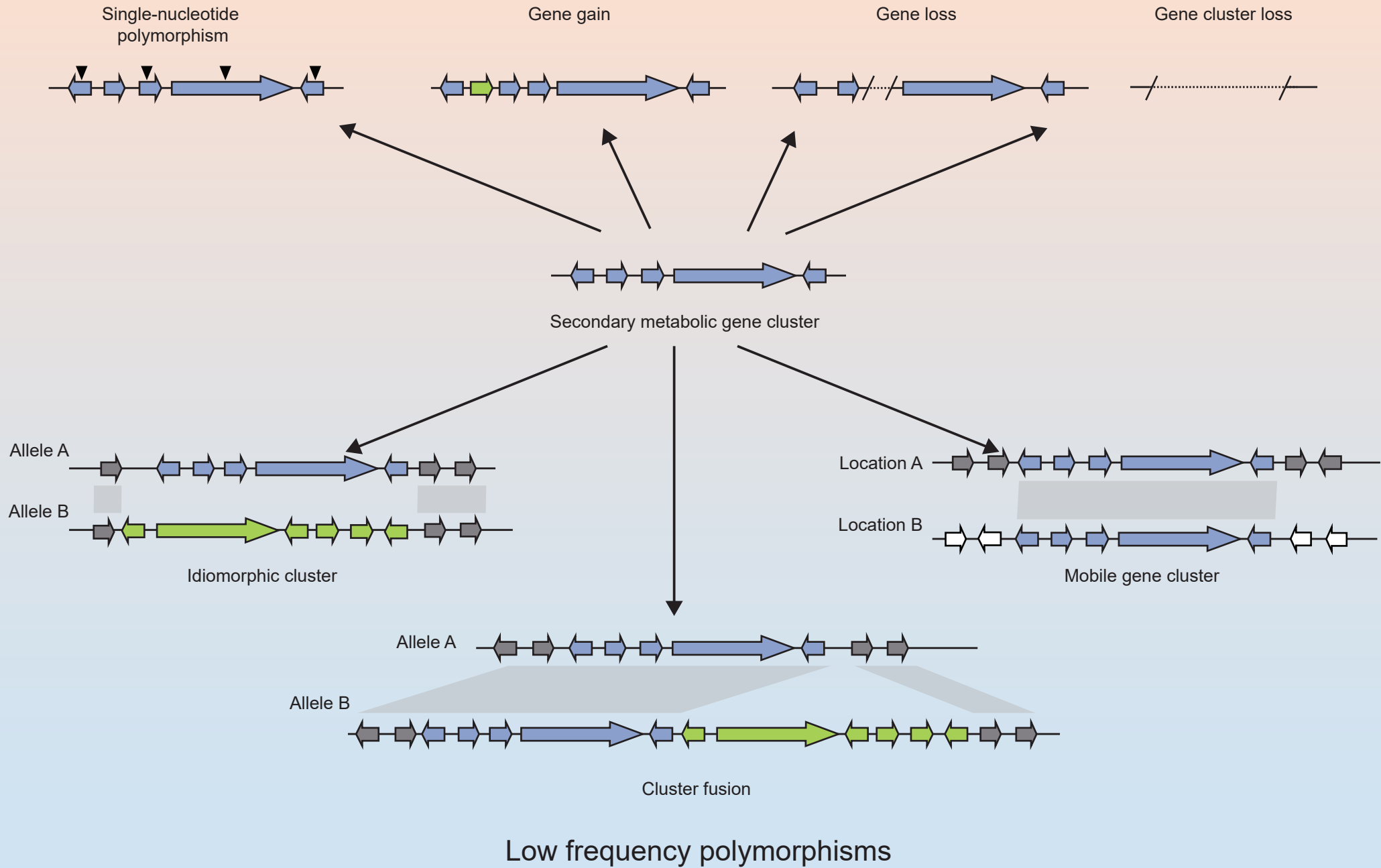


Figure 6. Types and frequencies of all SM gene cluster variants in the *A. fumigatus* population.