Drivers of genetic diversity in secondary metabolic gene clusters within a fungal species

Abigail L. Lind[1‡], Jennifer H. Wisecaver[2], Catarina Lameiras[3], Philipp Wiemann[4§], Jonathan M. Palmer[5], Nancy P. Keller[4], Fernando Rodrigues[6,7], Gustavo H. Goldman[8], Antonis Rokas[1,2†]

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

2. Department of Biological Sciences, Vanderbilt University, Nashville, Tennessee, USA

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

4. Department of Medical Microbiology & Immunology, University of Wisconsin-Madison, Madison, Wisconsin, USA

5. Center for Forest Mycology Research, Northern Research Station, US Forest Service, Madison, Wisconsin, USA

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

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

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

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

‡Current address: Gladstone Institutes, San Francisco, California, USA

§Current address: Hexagon Bio, Menlo Park, California, USA

Summary

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

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

Introduction

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

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

In the last decade, several comparative studies have described macroevolutionary patterns of SM gene cluster diversity. For example, studies centered on genomic comparisons of closely related species have identified several different types of inter-species divergence, from single nucleotide substitutions (e.g., differences in fumonisins produced by *Fusarium* species are caused by variants in one gene [10]), to gene gain / loss events (e.g., the

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

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

To address these questions, we investigated the genetic diversity of all 36 known and predicted SM gene clusters in whole genome sequence data from 66 strains, 8 of which were sequenced in this study, of the opportunistic human pathogen *Aspergillus fumigatus*, a species with cosmopolitan distribution and panmictic population structure [28]. We found that 13 SM

gene clusters were generally conserved and harbored low amounts of variation. In contrast, the remaining 23 SM gene clusters were highly variable and contained one or more of five different types of genetic variation: single-nucleotide polymorphisms including nonsense and frameshift variants, individual gene gain and loss polymorphisms, entire cluster gain and loss polymorphisms, polymorphisms associated with changes in cluster genomic location, and clusters with non-homologous alleles resembling the idiomorphs of fungal mating loci. Many clusters contained interesting combinations of these types of polymorphisms, such as pseudogenization in some strains and entire cluster loss in others. The types of variants we find are likely generated by a combination of DNA replication and repair errors, recombination, genomic insertions and deletions, and horizontal transfer. We additionally find an enrichment for transposable elements (TEs) around horizontally transferred clusters, clusters that change in genomic locations, and idiomorphic clusters. Taken together, our results provide a guide to both the types of polymorphisms and the genetic drivers of SM gene cluster diversification in filamentous fungi. As most of the genetic variants that we observe have been previously associated with SM gene cluster diversity across much larger evolutionary distances and timescales, we argue that processes influencing SM gene cluster diversity within species are sufficient to explain SM cluster macroevolutionary patterns.

Results and Discussion

We analyzed the genomes of 66 globally distributed strains of *Aspergillus fumigatus* for polymorphisms in SM gene clusters. We performed whole-genome sequencing on 8 strains, and collected the remaining 58 strains from publicly available databases including NCBI Genome and NCBI Short Read Archive (Figure 1, Table S1) [28–32]. All publicly available strains of *A. fumigatus* with sequencing data passing quality thresholds (see Methods) or with assembled genomes were included in our analysis. The resulting dataset contains strains sampled from 12 sites world-wide and from clinical and environmental sources (Table S1).

We analyzed all strains for polymorphisms in 33 curated SM gene clusters present in the reference Af293 genome, and additionally searched for novel SM gene clusters (see Methods). These examinations revealed five distinct types of polymorphisms in SM gene clusters (Figure 1, Table 1):

a) Single nucleotide and short indel polymorphisms. 33 / 33 SM gene clusters (present in the reference Af293 strain) contained multiple genes with missense SNPs and short indel variants in one or more strains. 23 / 33 SM gene clusters contained one or more genes with frameshift or nonsense variants.
b) Gene content polymorphisms involving loss or gain of one or more genes. 6 / 33 SM gene clusters contained a gene content polymorphism.
c) Whole SM gene cluster gain and loss polymorphisms. 3 / 33 SM gene clusters were entirely absent in one or more strains and an additional 3 previously unknown SM gene clusters were discovered.
d) Idiomorphic polymorphisms. One locus contained multiple non-homologous SM gene cluster alleles in different strains.
e) Genomic location polymorphisms. 2 / 33 SM gene clusters were found on different chromosomes between strains.

Both genomic location polymorphisms and idiomorphic polymorphisms are novel types of variants that have not been previously described for secondary metabolic gene clusters, likely

158 because they can only be identified when genome-wide synteny and sequence conservation are
159 high. The remaining types of variants, including single-nucleotide changes and gene gain and
160 loss events, have been implicated at the species level as major drivers of secondary metabolic
161 gene cluster evolution (Table 1), suggesting that the diversity-generating processes observed
162 within a species are sufficient to explain SM gene cluster evolution across species.
163

164 **Single-nucleotide and indel polymorphisms**

165 It is well established that single nucleotide polymorphisms (SNPs) and short indel
166 polymorphisms, which are caused by errors in DNA replication and repair, are a major source of
167 genomic variation [33]. Non-synonymous SNPs and indels with missense, frameshift, and
168 nonsense effects were widespread across the 33 SM reference gene clusters (Figure 1, Table
169 S2). Every strain contained numerous missense mutations and at least one nonsense or
170 frameshift mutation in its SM gene clusters. Although missense mutations are likely to influence
171 SM production, the functional effects of nonsense and frameshift mutations are comparatively
172 easier to infer from genomic sequence data because they often result in truncated proteins and
173 loss of protein function.
174

175 SNPs and short indel polymorphisms can affect secondary metabolite production, as in
176 the case of the lack of trypacidin production in the A1163 strain because of a previously
177 identified frameshift mutation in the polyketide synthase (PKS) of the trypacidin gene cluster
178 [34]. Interestingly, we identified a premature stop codon (Gln273*) in a transcription factor
179 required for trypacidin production, *tpcD* (Afu4g14550), in a strain sequenced in this study
180 (MO79587EXP) (Table S2). These data suggest that function of this SM gene cluster has been
181 lost at least twice independently in *A. fumigatus*.
182

183 Individual nonsense or frameshift variants varied in frequency. For example, the non-
184 ribosomal peptide synthase (NRPS) *pes3* gene (Afu5g12730) in SM gene cluster 21 harbors 16
185 nonsense or frameshift polymorphisms in 55 strains, seven of which are common (present in
186 ≥10 strains) and another seven are rare (≤5 strains). Strains with lab-mutated null alleles of the

187 *pes3* gene are more virulent than strains with functional copies [35], which may explain the
188 widespread occurrence of null *pes3* alleles within *A. fumigatus*.
189

190 **Gene content polymorphisms**
191 We additionally identified several SM gene clusters that gained or lost genes in some
192 strains. These gene content polymorphisms were most likely generated through genomic
193 deletion or insertion events and were sometimes found at high frequencies among strains
194 (Figure 1, Table 1). In three cases, these polymorphisms impact backbone synthesis genes,
195 rendering the SM gene cluster non-functional.
196
197 One example involves SM gene cluster 14, whose standard composition includes a
198 pyoverdine synthase gene, an NRPS-like gene, an NRPS backbone gene, and several additional
199 modification genes (Figure 2A). Four of the 66 strains examined lack an 11-kb region on the 3'
200 end of the cluster, which normally contains an NRPS gene and two additional cluster genes, and
201 the first non-SM genes on the 3' end flanking the cluster. All *A. fumigatus* strains contain a
202 *copia* family TE [36,37] at the 3' end of the cluster, suggesting that TEs may have been involved
203 in the generation of this polymorphism. While this polymorphism could have arisen through a
204 deletion event, a homologous cluster lacking the 11-kb region is also present in the reference
205 genomes of *Aspergillus lentulus* and *Aspergillus fischeri*, close relatives of *A. fumigatus* (Figure
206 2A). The most parsimonious explanation is that the genome of the *A. fumigatus* ancestor
207 contained an SM gene cluster that lacked the 11-kb region, and that this genomic region was
208 subsequently gained and increased in frequency within *A. fumigatus*.
209
210 The remaining two gene content polymorphisms affecting SM backbone genes were
211 restricted to one strain each and appear to have arisen through genomic deletion events.
212 Specifically, strain IF1SWF4 lacks an 8-Kb region near the helvolic acid SM gene cluster,
213 resulting in the loss of the backbone oxidosqualene cyclase gene as well an upstream region
214 containing two non-SM genes (Figure S1). Strain LMB35Aa lacks a 54-kb region on the end of

chromosome 2, which includes five genes from the telomere-proximal fumigaclavine C cluster (Figure S1).

Three other cases of gene content polymorphisms involved gene loss or truncation events of non-backbone structural genes. The second half of the ORF of the *gliM O*-methyltransferase gene in the gliotoxin gene cluster has been lost in 2 / 66 strains (Figure S1) and the first half of the permease *fmqE* in the fumiquinazoline gene cluster has been lost in 4 / 66 strains (Figure S1). Finally, an ABC transporter gene in SM cluster 21 has been almost entirely lost in 21 / 66 strains (Figure S1). This deletion event is found in strains that are related in the SNP-based strain phylogeny but does not perfectly mirror the phylogeny (Figure 1).

Whole gene cluster loss polymorphisms

Several SM gene clusters were gained or lost entirely across strains. We observed several instances where a cluster present in the genome of either the reference Af293 or A1163 (also known as CEA10) strains was absent or pseudogenized in other strains, which we present in this section.

One of the novel SM gene clusters, cluster 34, was present in all but two of the strains (Af293 and F7763). Cluster 34 contains a PKS backbone gene, one PKS-like gene with a single PKS associated domain, nine genes with putative biosynthetic functions involved in secondary metabolism, and six hypothetical proteins (Figure 2B). The two strains that lack cluster 34 contain a likely non-functional cluster fragment that includes the PKS-like gene, two biosynthetic genes, and three hypothetical proteins. Interestingly, the 3' region flanking cluster 34 is syntenic across all 66 strains but the 5' region is not, suggesting that a recombination or deletion event may have resulted in its loss in the Af293 and F7763 strains. These two strains form a clade in the strain phylogeny (Figure 1), so it is likely that this deletion or recombination event occurred once.

One notable example of an SM gene cluster present in the Af293 reference genome but absent or pseudogenized in others was SM cluster 4. This cluster contains 5 genes on the tip of the Af293 chromosome 1, and contains orthologs to five of the six genes in the fusarielin-producing gene cluster in *Fusarium graminearum* [38]. Cluster 4 is also present in several other *Aspergillus* species, including *A. clavatus* and *A. niger* [38], as well as in whole or in part in other non-*Aspergillus* fungi in the class Eurotiomycetes and in fungi in the class Sordariomycetes (Figure S3) [30,39–47]. Phylogenetic analysis of the genes in cluster 4 does not provide a clear view of the origin of this cluster, which is consistent either with extensive gene loss in both Sordariomycetes and Eurotiomycetes, or alternatively with horizontal gene transfer (HGT) between fungi belonging to the two classes (Figure S2, Figure S3).

Cluster 4 is entirely absent in 4 / 66 strains, and its genes are undergoing pseudogenization in an additional 43 strains via multiple independent mutational events (Figure 3). The four strains lacking the cluster form a single clade on the strain phylogeny, suggesting that the cluster was lost in a single deletion event (Figure 1). Further, 19 strains shared a single frameshift variant in the polyketide synthase gene (4380_4381insAATGGGCT; frameshift at Glu1461 in Afu1g17740) and an additional 13 strains shared a single frameshift variant (242delG; frameshift at Gly81) in an aldose 1-epimerase gene (Afu1g17723) (Figure 3A, Table S2). Eleven other strains each contained one to several frameshift or nonsense polymorphisms involving nine unique mutational sites. Five of these strains contained multiple distinct frameshifts and premature stop codons in more than one gene in the cluster, indicating that the entire pathway is pseudogenized in these strains.

A phylogeny of the entire cluster 4 locus across all 62 strains with short-read data shows that two pseudogenizing variants shared across multiple strains, one in the aldose 1-epimerase gene and one in the polyketide synthase, are found in loci that form well-supported clades (Figure 3B), suggesting that these variants arose once. Similarly, a set of variants shared across three strains and one variant shared in two strains are found in loci that form well-supported clades in the locus phylogeny. Two strains sharing a pseudogenizing variant in the polyketide

synthase do not group together in the locus phylogeny, a discordance likely stemming from within-locus recombination events. Finally, functional alleles of cluster 4 are distributed throughout the locus phylogeny, suggesting that the functional allele is ancestral and the pseudogenized variants are derived.

Perhaps surprisingly, loss of function polymorphisms (from nonsense and frameshift mutations to wholesale cluster loss) are common and sometimes frequent within *A. fumigatus*. The majority of these polymorphisms are presumably neutral, and reflect the fact that any mutation is more likely to result in loss of a function than in gain. Consistent with this hypothesis is our observation that these loss events were often found at low frequencies. However, the possibility also exists that some of the high-frequency, recurrent loss of function polymorphisms may be adaptive. Given that many secondary metabolites are primarily secreted in the extracellular environment and can benefit nearby conspecifics that are not themselves producing the metabolite [51], individual strains may be circumventing the energetically costly process of producing the metabolite themselves in a situation analogous to the Black Queen hypothesis [52].

Whole gene cluster gain polymorphisms

By searching for novel SM gene clusters in the genomes of the other 65 *A. fumigatus* strains, we found three SM gene clusters that were absent from the genome of the Af293 reference strain. As SM gene clusters are often present in repeat-rich and subtelomeric regions that are challenging to assemble [48,49], the strains analyzed here might harbor additional novel SM gene clusters that were not captured here.

One of these SM gene clusters, cluster 34, was mentioned earlier as an example of whole gene cluster loss polymorphism (Figure 2B) and is present in most strains but has been lost in two strains. The other two SM gene clusters absent from the Af293 genome are present at lower frequencies and likely reflect gene cluster gain events; cluster 35 is present in 2 / 66 strains and cluster 36 in 4 / 66 strains. Cluster 35 is located in a region syntenic with an Af293

chromosome 4 region and is flanked on both sides by TEs (Figure S4). Eight of the 14 genes in this SM gene cluster are homologous to genes in an SM gene cluster in the genome of the insect pathogenic fungus *Metarhizium anisopliae* (Figure S4) [50]. Phylogenetic analysis of these 8 genes is consistent with a horizontal transfer event (Figure S5). The two strains that contain this novel cluster are not sister to each other on the strain phylogeny (Figure 1).

Cluster 36 is an NRPS-containing cluster located on shorter genomic scaffolds that lack homology to either the Af293 or A1163 genomes, making it impossible to determine on which chromosome this cluster is located (Figure S4). Two of the strains containing this novel cluster are sister to each other on the strain phylogeny, while the third is distantly related to these two (Figure 1). The evolutionary histories of the genes in the cluster are consistent with vertical inheritance, and these genes are present in multiple *Aspergillus* species.

Idiomorph polymorphisms

One of the most peculiar types of polymorphisms that we identified is a locus containing different unrelated alleles of SM gene clusters, reminiscent of the idiomorph alleles at the fungal mating loci [53]. This locus, which resides on chromosome 3 and corresponds to cluster 10 in the Af293 genome (Figure 4), was previously described as being strain-specific in a comparison between Af293 and A1163 strains [30] and is thought to reside in a recombination hot spot [28]. Our analysis showed that there are at least 6 different alleles of this cluster in *A. fumigatus*, containing 4 different types of key enzymes involved in natural product biosynthesis: a polyketide synthase (PKS) non-ribosomal peptide synthetase (NRPS) hybrid, a highly reducing (HR) PKS, a non-reducing (NR) PKS and an NRPS-like enzyme (Figure 4). Two additional alleles were present in only one strain each (Figure S6).

In the Af293 reference genome, the cluster present at the idiomorph locus contains one NR-PKS along with an NRPS-like gene (Allele B). In the A1163 reference genome and 17 other strains, there is a PKS-NRPS and an HR-NRPS at this locus (Allele E). These alleles show an almost complete lack of sequence similarity except for a conserved hypothetical protein and a

fragment of the HR-PKS in the Af293 allele; in contrast, the upstream and downstream flanking regions of the two alleles, which do not contain any backbone genes, are syntenic. Remarkably, another allele, present in 12 strains, contains all of the genes from both the Af293 and A1163 clusters (Allele D). The remaining three alleles contain various combinations of these genes. One allele found in 22 strains contains some A1163-specific genes, including the HR-PKS, and no Af293-specific genes (Allele F), while another allele found in 3 strains contains some Af293-specific genes, including the NRPS-like gene, but no A1163 genes (Allele A). The final allele, present in 8 strains, contains the entire Af293 allele as well as part of the A1163 allele containing the HR-PKS (Allele C). Every allele is littered with multiple long terminal repeat sequence fragments from *gypsy* and *copia* TE families as well as with sequence fragments from DNA transposons from the *mariner* family [36]. In some cases, these TEs correspond with breakpoints in synteny between alleles, suggesting that the diverse alleles of this SM gene cluster may have arisen via TE-driven recombination. Further, both of the alleles that are restricted to a single strain have an insertion event of several genes near a TE, while the rest of the locus is highly similar to one of the more common alleles (Figure S6).

Untargeted XCMS analysis [54] of an Allele D strain (08-19-02-30) and two Allele F strains (08-12-12-13 and 08-19-02-10) and comparison of their metabolite profiles revealed the presence of 2 unique masses in 08-19-02-30 (Table S4; Figure S7), raising the possibility that variation at the idiomorph locus is functional. Further analysis is underway to investigate whether any of these m/z can be directly linked to the Allele D sequence.

To gain insight into the evolutionary history of this locus, we constructed a phylogeny based on its conserved downstream flanking region (Figure 4B). The resulting phylogeny shows some grouping of strains that share alleles, but there are no clades that contain all instances of a particular allele. This is likely to be the consequence of within-locus recombination between strains of *A. fumigatus*, which has been previously described at this locus [28] and which is potentially driven by the high number of repetitive sequences at this locus.

While it is tempting to speculate that Allele D, the longest allele containing all observed genes, represents the ancestral state, this does not explain the presence of a shared hypothetical protein and PKS gene fragment between Allele C and Allele B. Further, two close relatives of *A. fumigatus*, *A. lentulus* and *A. fischeri,* contain a similar region with conserved upstream and downstream flanking genes that is highly dissimilar from any of the alleles observed in *A. fumigatus* (Figure S8). In both species, this locus contains numerous TEs as well as genes homologous to portions of allele E in *A. fumigatus* (Figure S8). *A. fischeri* additionally contains two hypothetical proteins from the PKS-NRPS region of *A. fumigatus* and an additional hybrid PKS-NRPS-containing gene cluster not found in either *A. lentulus* or any *A. fumigatus* strain (Figure S8). Other genes at this locus in both *A. lentulus* and *A. fischeri* have functions likely not related to SM. Interestingly, *A. lentulus* contains a gene with a heterokaryon incompatibility protein domain, which may be involved in determining vegetative incompatibility [55]. Only one representative genome from each species has been sequenced, but based on the high concentration of TEs and lack of sequence similarity with any *A. fumigatus* alleles, it is likely that this locus is highly variable within both *A. lentulus* and *A. fischeri*.

It is possible that polymorphism at this locus originated via SM gene cluster fusion or fission events driven by TEs, which are present in large numbers. Interestingly, two other previously described instances of SM gene cluster variation bear some resemblance to the *A. fumigatus* idiomorphic SM gene cluster 10 locus. The first is the presence of two non-homologous *Aspergillus flavus* alleles, where some strains contain a 9-gene sesquiterpene-like SM gene cluster and others contain a non-homologous 6-gene SM gene cluster at the same genomic location [56]. The second is the presence of two non-homologous SM gene clusters at the same, well-conserved, locus in a comparison of six species of dermatophyte fungi [57]. Based on these results, we hypothesize that idiomorphic clusters may be common in fungal populations and contribute to the broad diversity of SM gene clusters across filamentous fungi.

Genomic location polymorphisms

The final type of polymorphism that we observed is associated with SM gene clusters that are found in different genomic locations in different strains, suggesting that these SM gene clusters are behaving like mobile genetic elements. This type of polymorphism was observed in SM gene clusters 1 and 33, both of which produce as yet identified products, and are present at low frequencies in *A. fumigatus* strains.

SM gene cluster 1, which is present in six strains at three different genomic locations (Figure 5A), consists of a PKS and four other structural genes that are always flanked by a 15 Kb region (upstream) and a 43 Kb region (downstream) containing TEs. In the reference Af293 strain and in strain F7763, cluster 1 and its flanking regions are located on chromosome 1, while in strains 08-31-08-91, F13619, and Z5 they are located between Afu4g07320 and Afu4g07340 on chromosome 4. In contrast, in strain JCM10253, the cluster and flanking regions are located on chromosome 8 immediately adjacent to the 3' end of the intertwined fumagillin and pseurotin SM gene supercluster [58]. The strains containing the allele on Chromosome 1 are sister to each other on the strain phylogeny, while the other strains are scattered across the tree and do not reflect the phylogeny (Figure 1).

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

This "jumping" gene cluster is not present in any other sequenced genome in the genus *Aspergillus*, and phylogenetic analysis of its constituent genes is consistent with HGT between fungi (Figure S7). Specifically, this gene cluster is also present in *Phaeosphaeria nodorum* [59], a plant pathogen from the class Dothideomycetes, *Pseudogymnoascus pannorum* [60], a fungus isolated from permafrost from the Leotiomycetes, and *Escovopsis weberi* [61], a fungal parasite of fungus-growing ants from the Sordariomycetes (Figure 5B). One additional species, the endophyte *Hypoxylon* sp. CI4A from the class Sordariomycetes [62], contains four of the five

cluster genes but is missing Afu1g00970, an MFS drug transporter. However, this species contains a gene unrelated to Afu1g00970 that is annotated as an MFS drug transporter immediately adjacent to this cluster (Figure 5B). None of these fungi contain the upstream or downstream TE-rich flanking regions present in *A. fumigatus*, and each fungus contains additional unique genes with putative biosynthetic functions adjacent to the transferred cluster. The most likely explanation for this change in flanking regions is that this SM gene cluster was transferred into *A. fumigatus* once and has subsequently "jumped" in different genomic locations in different strains.

The second SM gene cluster that shows variation in its genomic location across strains, cluster 33, contains a terpene synthase. This cluster is present in only 5 strains at 3 distinct locations (Figure 5C). Similar to cluster 1, cluster 33 is also flanked by TEs, and in one strain the cluster is located in a new region 58 Kb from SM gene cluster 34. Two strains that contain the cluster in the same genomic location are sister to each other on the strain phylogeny, while the placement of the other three strains containing the cluster does not reflect the phylogeny (Figure 1). In contrast to cluster 1, cluster 33 does not appear to have been horizontally transferred between fungi and its genes are present in other sequenced *Aspergillus* species [63], suggesting that the mobility of clusters 1 and 33 may be driven by different mechanisms.

Interestingly, both cases of mobile gene clusters are located near or immediately adjacent to other SM gene clusters in some strains. Cluster 33 is located 58Kb away from cluster 34 in one strain and Cluster 1 is located immediately adjacent to the intertwined fumagillin and pseurotin supercluster [58] in another. This supercluster is regulated by the transcriptional factor *fapR* (Afu8g00420) and is located in a chromosomal region controlled by the master SM regulators *laeA* (Afu1g14660) and *veA* (Afu1g12490) [58,64], raising the hypothesis that mobile gene clusters might be co-opting the regulatory machinery acting on adjacent SM gene clusters. Previous work has hypothesized that the fumagillin and pseurotin supercluster formed through genomic rearrangement events placing the once-independent gene clusters in close proximity to each other [58]. Our observation that the mobile cluster 1 is

446 located in this same region not only supports this hypothesis but also implicates TEs as one of
447 the mechanisms by which superclusters are formed. These superclusters may also represent an
448 intermediate stage in the formation of new SM gene clusters. Supercluster formation,
449 potentially mediated by mobile gene clusters, and followed by gene loss, could explain
450 macroevolutionary patterns of SM gene clusters where clustered genes in one species are
451 found to be dispersed over multiple gene clusters in other species [9,11].
452
453 **Conclusions**
454 Our examination of the genomes of 66 strains of *Aspergillus fumigatus* revealed five
455 general types of polymorphisms that describe variation in SM gene clusters. These
456 polymorphisms include variation in SNPs and short indels, gene and gene cluster gains and
457 losses, non-homologous (idiomorph) gene clusters at the same genomic position, and mobile
458 clusters that differ in their genomic location across strains (Figure 6). Previous work has
459 demonstrated that SM gene clusters, like the metabolites that they produce, are highly
460 divergent between fungal species [8,9,19,63]. Our examination of genome-wide variation
461 shows that these SM gene clusters are also diverse across strains of a single fungal species.
462 These results also demonstrate that the diversity of SM gene clusters within *A. fumigatus*
463 cannot be captured by sequencing a single representative strain, which is the current standard
464 practice for determining the SM gene cluster content of a fungal species.
465
466 The quantification of diversity in SM gene clusters within a species is dependent on both
467 numbers and types of strains analyzed. The types of polymorphisms detected as well as their
468 observed frequency, especially for rare polymorphisms, will increase with the number of
469 genomes examined. In addition, both the frequencies of the different types of polymorphisms
470 and the polymorphisms themselves may also change with sampling design, or in a manner
471 corresponding to the population structure or ecology of the species under study. *A. fumigatus*
472 is a cosmopolitan species with panmictic population structure [28], characteristics that do not
473 always apply to other filamentous fungi. Fungi exhibiting strong population structure or fungi
474 adapted to different ecological niches might contain different patterns of genetic diversity.

475

476 Nevertheless, the variants and genetic drivers we observe at the within-species level are
477 also implicated as driving SM gene cluster variation at the between-species level, suggesting
478 that the observed microevolutionary processes are sufficient to explain macroevolutionary
479 patterns of SM gene cluster evolution. For example, the narrow and discontinuous distribution
480 of SM gene clusters across the fungal phylogeny has been attributed to HGT as well as to gene
481 cluster loss [13,15,20,22,30,65–67]. Here, we find evidence that both processes also influence
482 the distribution of SM gene clusters within a species (Figures 2, 5, S2-S5). Interestingly, the
483 fraction of SM gene clusters within *A. fumigatus* that harbor loss of function polymorphisms is
484 substantial, consistent with the macroevolutionary view that SM gene cluster loss is rampant
485 [18,19,67]. However, our within-species observations are also consistent with the
486 macroevolutionary importance of HGT to SM gene cluster evolution. Once thought to be non-
487 existent in eukaryotes, HGT is now considered to be responsible for the presence of several
488 different SM gene clusters in diverse filamentous fungi [13,67,68]. The instances of HGT of SM
489 gene clusters within *A. fumigatus* suggests that acquisition of foreign genetic material
490 containing SM gene clusters is likely a common and ongoing occurrence in fungal populations.
491

492 One recurring theme across different types of SM gene cluster polymorphisms in *A.*
493 *fumigatus* was the perpetual presence of TEs adjacent to or within clusters. One particularly
494 striking case is the "idiomorphic" Cluster 10, where TEs seem to correspond with breakpoints in
495 synteny both within *A. fumigatus* and also between *A. fumigatus* and its close relatives (Figure
496 4, Figure S8). TEs were also present flanking mobile and horizontally transferred SM gene
497 clusters and were located adjacent to gene gain sites. There are several potential explanations
498 for the observed TE enrichment. First, TE presence may promote repeat-driven recombination
499 and gene rearrangement, or the TEs themselves may be the agents of horizontally transferred
500 clusters (either on their own or through a viral vector). Alternatively, it may simply be the case
501 that SM gene clusters preferentially reside in TE-rich genomic regions.
502

In summary, examination of SM gene cluster variation within a single fungal species revealed five distinct types of polymorphism that are widespread across different types of SM gene clusters and are caused by many underlying genetic drivers, including errors in DNA transcription and repair, non-homologous recombination, gene duplication and loss, and HGT. The net effect of the observed variation raises the hypothesis that the chemical products of filamentous fungal species are in a state of evolutionary flux, each population constantly altering its SM gene cluster repertoire and consequently modifying its chemodiversity.

Methods

Strains analyzed

Eight strains of *A. fumigatus* were isolated from four patients with recurrent cases of aspergillosis in the Portuguese Oncology Institute in Porto, Portugal. Each strain was determined to be *A. fumigatus* using macroscopic features of the culture and microscopic morphology observed in the slide preparation from the colonies with lactophenol solution [69]. Based on the morphological characterization, all clinical strains were classified as *A. fumigatus complex*-Fumigati. After whole genome sequencing, retrieval and examination of the beta tubulin and calmodulin sequences of each strain confirmed that all strains belonged to *A. fumigatus* (see Phylogenetic Analysis and Figure S9). The genomes of all eight strains were sequenced using 150bp Illumina paired-end sequence reads at the Genomic Services Lab of Hudson Alpha (Huntsville, Alabama, USA). Genomic libraries were constructed with the Illumina TruSeq library kit and sequenced on an Illumina HiSeq 2500 sequencer. Samples of all eight strains were sequenced at greater than 180X coverage or depth (Table S1). Short read sequences for these 8 strains are available in the Short Read Archive under accession SRP109032.

In addition to the 8 strains sequenced in this study, we retrieved 58 *A. fumigatus* strains with publicly available whole genome sequencing data, resulting in a dataset of 66 strains (Table S1). The strains used included both environmental and clinical strains and were isolated from multiple continents. Genome assemblies for 10 of these strains, including the Af293 and A1163 reference strains, were available for download from GenBank [28–32,70]. For 6 of these strains, short read sequences were also available from the NCBI Short Read Archive (SRA), which were used for variant discovery only (see Single nucleotide variant (SNV) and indel discovery) and not for genome assembly. Short read sequences were not available for the remaining 4 strains. Short read sequences were downloaded for an additional 48 strains from the Short Read Archive if they were sequenced with paired-end reads and at greater than 30x coverage.

Single nucleotide variant (SNV) and indel discovery

All strains with available short read data (62 of 66 strains) were aligned to both the Af293 and A1163 reference genomes using BWA mem version 0.7.12-r1044 [71]. Coverage of genes present in the reference genome was calculated using bedtools v2.25.0 [72]. SNV and indel discovery and genotyping was performed relative to the Af293 reference genome and was conducted across all samples simultaneously using the Genome Analysis Toolkit version 3.5-0-g36282e4 with recommended hard filtering parameters [73–75] and annotated using snpEff version 4.2 [76].

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

All 56 strains without publicly available genome assemblies were *de novo* assembled using the iWGS pipeline [77]. Specifically, all strains were assembled using SPAdes v3.6.2 and MaSuRCA v3.1.3 and resulting assemblies were evaluated using QUAST v3.2 [78–80]. The average N50 of assemblies constructed with this strategy was 463 KB (Table S1). Genes were annotated in these assemblies as well as in five GenBank assemblies with no predicted genes using augustus v3.2.2 trained on *A. fumigatus* gene models [81]. Repetitive elements were annotated in all assemblies using RepeatMasker version open-4.0.6 [82].

Secondary metabolic gene cluster annotation and discovery

Secondary metabolic gene clusters in the Af293 reference genome were taken from two recent reviews, both of which considered computational and experimental data to delineate cluster boundaries [83,84] (Table S3). The genomes of the other 65 strains were scanned for novel SM gene clusters using antiSMASH v3.0.5.1 [85]. To prevent potential assembly errors from confounding the analysis, any inference about changes in genomic locations of genes or gene clusters was additionally verified by manually inspecting alignments and ensuring that paired end reads supported an alternative genomic location (see SNV and indel discovery). Cases where paired end reads did not support the change in genomic location (i.e., all 3' read mapping to Chromosome 1 and all 5' pairs mapping to Chromosome 8), or where mapping was ambiguous or low quality were discarded.

568

569 **Phylogenetic analysis**

570 To confirm all strains in this analysis belonged to the species *Aspergillus fumigatus*, the
571 genomic sequences of the beta tubulin and calmodulin genes were extracted from the
572 assembled genomes of all strains. Gene phylogenies were constructed using *Aspergillus*
573 *fischerianus* as an outgroup using RAxML v8.0.25 with the GTRGAMMA substitution model [86].
574 The tree was midpoint rooted and all branches with bootstrap support less than 80% were
575 collapsed (Figure S10).

576

577 To construct a SNP-based strain phylogeny, biallelic SNPs with no missing data were
578 pruned using SNPRelate v1.8.0 with a linkage disequilibrium threshold of 0.8 [87]. A total of
579 15,274 SNVs were used to create a phylogeny using RAxML v8.0.25 with the ASC_BINGAMMA
580 substitution model [86]. The tree was midpoint rooted and all branches with bootstrap support
581 less than 80% were collapsed. The phylogeny was visualized using ITOL version 3.0 [88].

582

583 To understand the evolutionary histories of specific SM gene clusters showing unusual
584 taxonomic distributions, we reconstructed the phylogenetic trees of their SM genes.
585 Specifically, SM cluster protein sequences were queried against a local copy of the NCBI non-
586 redundant protein database (downloaded May 30, 2017) using phmmer, a member of the
587 HMMER3 software suite [89] using acceleration parameters --F1 1e-5 --F2 1e-7 --F3 1e-10. A
588 custom perl script sorted the phmmer results based on the normalized bitscore (nbs), where
589 nbs was calculated as the bitscore of the single best-scoring domain in the hit sequence divided
590 by the best bitscore possible for the query sequence (i.e., the bitscore of the query aligned to
591 itself). No more than five hits were retained for each unique NCBI Taxonomy ID. Full-length
592 proteins corresponding to the top 100 hits (E-value < 1 × 10-10) to each query sequence were
593 extracted from the local database using esl-sfetch [89]. Sequences were aligned with MAFFT
594 v7.310 using the E-INS-i strategy and the BLOSUM30 amino acid scoring matrix [90] and
595 trimmed with trimAL v1.4.rev15 using its gappyout strategy [91]. The topologies were inferred
596 using maximum likelihood as implemented in RAxML v8.2.9 [86] using empirically determined

597 substitution models and rapid bootstrapping (1000 replications). The phylogenies were
598 midpoint rooted and branches with less than 80% bootstrap support were collapsed using the
599 ape and phangorn R packages [92,93]. Phylogenies were visualized using ITOL version 3.0 [88].
600
601 To understand the evolutionary histories of SM gene clusters 4 and 10, full-length
602 nucleotide sequences of all 62 strains with short read sequence data were extracted for the
603 entire cluster region (SM gene cluster 4) or the downstream flanking region (SM gene cluster
604 10) using the previously described SNV analysis procedure followed by Genome Analysis
605 Toolkit's "ExtractAlternativeReferenceFasta" tool [74]. The resulting nucleotide sequences were
606 aligned using MAFFT v7.310 [90]. Phylogenies were constructed using maximum likelihood as
607 implemented in RAxML v 8.0.25, using the GTRGAMMA substitution model and rapid
608 bootstrapping (1000 replications) [86]. Phylogenies were midpoint rooted and branches with
609 less than 80% bootstrap support were collapsed. Phylogenies were visualized using ITOL version
610 3.0 [88].
611

612 **Differential metabolite analysis**
613 For natural product analysis, 5 x 10^6 spores/mL for the indicated strains were grown in
614 50 mL liquid GMM [94] for five days at 25 °C and 250 rpm in duplicates. Supernatants were
615 extracted with equal volumes of ethyl acetate, dried down and resuspended in 20% acetonitrile
616 (ACN). Each sample was analyzed by ultra-high performance liquid chromatography (UHPLC)
617 coupled with mass spectrometry (MS). The samples were separated on a ZORBAX Eclipse XDB-
618 C18 column (Agilent, 2.1 x 150 mm with a 1.8 µM particle size using a binary gradient of 0.5 %
619 (v/v) formic acid (FA) as solvent A and 0.5 % (v/v) FA in ACN as solvent B that was delivered by a
620 VanquishTM UHPLC system (Thermo Scientific) with a flow rate of 0.2 mL/min. The binary
621 gradient started with 20% B that was increased with a linear gradient to 100% B in 15 min
622 followed by an isocratic step at 100% B for 5 min. Before every run, the system was equilibrated
623 for 5 min at 20% B. The UHPLC system was coupled to a Q Exactive hybrid quadrupole
624 OritrapTM MS (Thermo Scientific). For electrospray ionization, the ion voltage was set at +/-3.5
625 kV in positive and negative mode, respectively. Nitrogen was used as sheath gas at a flow rate

626 of 45 and as sweep gas at a flow rate of 2. Data analysis was performed using XCMS [54] and

627 Maven [95] software.

628

Acknowledgements

A.L.L. was supported by the U.S. National Library of Medicine training grant 2T15LM007450. This work was supported in part by the National Science Foundation (IOS-1401682 to J.H.W. and DEB-1442113 to A.R.), the National Institutes of Health (NIH grant R01 AI065728-01 to N.P.K), and by the Northern Portugal Regional Operational Programme (NORTE 2020) under the Portugal 2020 Partnership Agreement through the European Regional Development Fund (FEDER) (NORTE-01-0145-FEDER-000013 to F.R.). G.H.G. thanks the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), both from Brazil, for funding and support. This work was conducted in part using the resources of the Advanced Computing Center for Research and Education at Vanderbilt University (http://www.accre.vanderbilt.edu/).

Author contributions

Conceptualization, A.R., G.H.G., A.L.L.; Methodology, A.L.L., J.H.W, C.L, P.W., J.M.P.; Investigation, A.L.L.; Visualization, A.L.L., J.H.W, P.W.; Resources, G.H.G., F.R., N.P.K., C.L.; Writing, A.L.L, A.R.

647 **Table 1. Types and rates of SM gene cluster variants in *A. fumigatus* strains.**

648

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 *Botrytis* [17,27], aflatoxin in *Aspergillus oryzae* and *Aspergillus flavus* [26], fumonisins in *Fusarium* [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 *Fusarium*, aflatoxin and sterigmatocystin in *Aspergillus* [11–15], HC toxin in *Cochliobolus carbonarum* [96]
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 *Fusarium* [24,25]
Cluster 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 *Aspergillus flavus* and *Aspergillus oryzae* [56,57]
Mobile gene clusters	Potential for change in gene regulation	Transposable elements; horizontal gene transfer; other mechanisms?	2 gene clusters	8 / 66 strains	None

649

650

651 **Figure S1. Alignments showing deletion of genes in SM gene clusters.**
652
653 **Figure S2. Gene phylogenies of the fusarielin-like SM gene cluster 4.**
654
655 **Figure S3. Fusarielin-like gene clusters in Eurotiomycetes and Sordariomycetes.**
656
657 **Figure S4. Novel SM gene clusters in A. fumigatus strains.**
658
659 **Figure S5. Gene phylogenies of SM gene cluster 34.**
660
661 **Figure S6. Two alleles of the idiomorphic SM gene cluster 10 present in one strain each.**
662
663 **Figure S7. Metabolomics analysis of strains with different alleles of the idiomorphic cluster**
664 **indicates presence of different metabolites.**
665
666 **Figure S8. Idiomorphic locus in other species.**
667
668 **Figure S9. Gene phylogenies of the mobile SM gene cluster 1.**
669
670 **Figure S10. Marker gene phylogenies of all strains and *Aspergillus fischeri*.**
671
672 **Table S1. Summary of strains, sequence data, and assemblies used.**
673
674 **Table S2. All nonsynonymous variants in SM gene cluster genes.**
675
676 **Table S3. Description, locus tags, and gene annotations of all reference Af293 SM gene**
677 **clusters.**
678
679 **Table S4. XCMS analysis of extracted supernatants from 08-19-02-30, 08-12-12-13, and 08-19-**
680 **02-10.**
681
682

683

684 **References**

685 1. Vining LC. Functions of secondary metabolites. Annu Rev Microbiol. Annual Reviews 4139
686 El Camino Way, P.O. Box 10139, Palo Alto, CA 94303-0139, USA; 1990;44: 395–427.
687 doi:10.1146/annurev.mi.44.100190.002143

688 2. Schimek C. Evolution of Special Metabolism in Fungi: Concepts , Mechanisms , and
689 Pathways. In: Pöggler S, Wöstmeyer J, editors. Evolution of Fungi and Fungal-Like
690 Organisms, The Mycota. XIV. Berlin, Heidelberg: Springer-Verlag; 2011. pp. 293–328.

691 3. Fox EM, Howlett BJ. Secondary metabolism: regulation and role in fungal biology. Curr
692 Opin Microbiol. 2008;11: 481–7. doi:10.1016/j.mib.2008.10.007

693 4. Scharf DH, Heinekamp T, Brakhage AA. Human and Plant Fungal Pathogens: The Role of
694 Secondary Metabolites. Heitman J, editor. PLoS Pathog. Public Library of Science;
695 2014;10: e1003859. doi:10.1371/journal.ppat.1003859

696 5. Yim G, Wang HH, Davies J. Antibiotics as signalling molecules. Philos Trans R Soc Lond B
697 Biol Sci. 2007;362: 1195–200. doi:10.1098/rstb.2007.2044

698 6. Keller NP. Translating biosynthetic gene clusters into fungal armor and weaponry. Nat
699 Chem Biol. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights
700 Reserved.; 2015;11: 671–7. doi:10.1038/nchembio.1897

701 7. Bennett J, Bentley R. What's in a name?—Microbial secondary metabolism. Adv Appl
702 Microbiol. 1989;34. Available:
703 http://www.sciencedirect.com/science/article/pii/S0065216408703162/pdf?md5=07c2a
704 f4dcbe482f13a1b94eb2f886c2e&pid=1-s2.0-S0065216408703162-main.pdf&_valck=1

705 8. Khaldi N, Seifuddin FT, Turner G, Haft D, Nierman WC, Wolfe KH, et al. SMURF: Genomic
706 mapping of fungal secondary metabolite clusters. Fungal Genet Biol. 2010;47: 736–41.
707 doi:10.1016/j.fgb.2010.06.003

708 9. Lind AL, Wisecaver JH, Smith TD, Feng X, Calvo AM, Rokas A. Examining the evolution of
709 the regulatory circuit controlling secondary metabolism and development in the fungal
710 genus Aspergillus. PLoS Genet. Public Library of Science; 2015;11: e1005096.
711 doi:10.1371/journal.pgen.1005096

712 10. Proctor RH, Busman M, Seo J-A, Lee YW, Plattner RD. A fumonisin biosynthetic gene
713 cluster in Fusarium oxysporum strain O-1890 and the genetic basis for B versus C
714 fumonisin production. Fungal Genet Biol. 2008;45: 1016–1026.
715 doi:10.1016/j.fgb.2008.02.004

716 11. Proctor RH, McCormick SP, Alexander NJ, Desjardins AE. Evidence that a secondary
717 metabolic biosynthetic gene cluster has grown by gene relocation during evolution of the
718 filamentous fungus Fusarium. Mol Microbiol. 2009;74: 1128–42. doi:10.1111/j.1365-
719 2958.2009.06927.x

12. Berry DB, Guan Q, Hose J, Haroon S, Gebbia M, Heisler LE, et al. Multiple Means to the Same End: The Genetic Basis of Acquired Stress Resistance in Yeast. Copenhaver GP, editor. PLoS Genet. Public Library of Science; 2011;7: e1002353. doi:10.1371/journal.pgen.1002353

13. Slot JC, Rokas A. Horizontal transfer of a large and highly toxic secondary metabolic gene cluster between fungi. Curr Biol. Elsevier Ltd; 2011;21: 134–9. doi:10.1016/j.cub.2010.12.020

14. Ehrlich KC, Chang P-K, Yu J, Cotty PJ. Aflatoxin Biosynthesis Cluster Gene cypA Is Required for G Aflatoxin Formation. Appl Environ Microbiol. 2004;70: 6518–6524. doi:10.1128/AEM.70.11.6518-6524.2004

15. Carbone I, Ramirez-Prado JH, Jakobek JL, Horn BW. Gene duplication, modularity and adaptation in the evolution of the aflatoxin gene cluster. BMC Evol Biol. 2007;7: 111. doi:10.1186/1471-2148-7-111

16. Yu J, Chang P-K, Ehrlich KC, Cary JW, Bhatnagar D, Cleveland TE, et al. Clustered pathway genes in aflatoxin biosynthesis. Appl Environ Microbiol. American Society for Microbiology; 2004;70: 1253–62. doi:10.1128/AEM.70.3.1253-1262.2004

17. Campbell M a, Rokas A, Slot JC. Horizontal transfer and death of a fungal secondary metabolic gene cluster. Genome Biol Evol. 2012;4: 289–93. doi:10.1093/gbe/evs011

18. Kroken S, Glass NL, Taylor JW, Yoder OC, Turgeon BG. Phylogenomic analysis of type I polyketide synthase genes in pathogenic and saprobic ascomycetes. Proc Natl Acad Sci U S A. 2003;100: 15670–5. doi:10.1073/pnas.2532165100

19. Bushley KE, Turgeon BG. Phylogenomics reveals subfamilies of fungal nonribosomal peptide synthetases and their evolutionary relationships. BMC Evol Biol. 2010;10: 26. doi:10.1186/1471-2148-10-26

20. Patron NJ, Waller RF, Cozijnsen AJ, Straney DC, Gardiner DM, Nierman WC, et al. Origin and distribution of epipolythiodioxopiperazine (ETP) gene clusters in filamentous ascomycetes. BMC Evol Biol. 2007;7: 174. doi:10.1186/1471-2148-7-174

21. Khaldi N, Collemare J, Lebrun M-H, Wolfe KH. Evidence for horizontal transfer of a secondary metabolite gene cluster between fungi. Genome Biol. BioMed Central; 2008;9: R18. doi:10.1186/gb-2008-9-1-r18

22. Khaldi N, Wolfe KH. Evolutionary Origins of the Fumonisin Secondary Metabolite Gene Cluster in Fusarium verticillioides and Aspergillus niger. Int J Evol Biol. Hindawi Publishing Corporation; 2011;2011: 423821. doi:10.4061/2011/423821

23. Reynolds HT, Slot JC, Divon HH, Lysøe E, Proctor RH, Brown DW. Differential Retention of Gene Functions in a Secondary Metabolite Cluster. Mol Biol Evol. Oxford University Press (OUP); 2017;10: e1004816. doi:10.1093/molbev/msx145

24. Wiemann P, Sieber CMK, von Bargen KW, Studt L, Niehaus E-M, Espino JJ, et al. Deciphering the Cryptic Genome: Genome-wide Analyses of the Rice Pathogen Fusarium fujikuroi Reveal Complex Regulation of Secondary Metabolism and Novel Metabolites. Mitchell AP, editor. PLoS Pathog. Public Library of Science; 2013;9: e1003475. doi:10.1371/journal.ppat.1003475

25. Chiara M, Fanelli F, Mulè G, Logrieco AF, Pesole G, Leslie JF, et al. Genome Sequencing of Multiple Isolates Highlights Subtelomeric Genomic Diversity within *Fusarium fujikuroi*. Genome Biol Evol. 2015;7: 3062–3069. doi:10.1093/gbe/evv198

26. Chang P-K, Horn BW, Dorner JW. Sequence breakpoints in the aflatoxin biosynthesis gene cluster and flanking regions in nonaflatoxigenic Aspergillus flavus isolates. Fungal Genet Biol. 2005;42: 914–923. doi:10.1016/j.fgb.2005.07.004

27. Schumacher J, Gautier A, Morgant G, Studt L, Ducrot P-H, Le Pêcheur P, et al. A Functional Bikaverin Biosynthesis Gene Cluster in Rare Strains of Botrytis cinerea Is Positively Controlled by VELVET. Yun S-H, editor. PLoS One. Public Library of Science; 2013;8: e53729. doi:10.1371/journal.pone.0053729

28. Abdolrasouli A, Rhodes J, Beale MA, Hagen F, Rogers TR, Chowdhary A, et al. Genomic Context of Azole Resistance Mutations in Aspergillus fumigatus Determined Using Whole-Genome Sequencing. MBio. 2015;6. doi:10.1128/mBio.00536-15

29. Nierman WC, Pain A, Anderson MJ, Wortman JR, Kim HS, Arroyo J, et al. Genomic sequence of the pathogenic and allergenic filamentous fungus Aspergillus fumigatus. Nature. 2005;438: 1151–6. doi:10.1038/nature04332

30. Fedorova ND, Khaldi N, Joardar VS, Maiti R, Amedeo P, Anderson MJ, et al. Genomic islands in the pathogenic filamentous fungus Aspergillus fumigatus. Richardson PM, editor. PLoS Genet. Public Library of Science; 2008;4: e1000046. doi:10.1371/journal.pgen.1000046

31. Knox BP, Blachowicz A, Palmer JM, Romsdahl J, Huttenlocher A, Wang CCC, et al. Characterization of Aspergillus fumigatus Isolates from Air and Surfaces of the International Space Station. mSphere. 2016;1.

32. Paul S, Zhang A, Ludeña Y, Villena GK, Yu F, Sherman DH, et al. Insights from the genome of a high alkaline cellulase producing Aspergillus fumigatus strain obtained from Peruvian Amazon rainforest. J Biotechnol. 2017;251: 53–58. doi:10.1016/j.jbiotec.2017.04.010

33. Roberts JD, Kunkel TA. Fidelity of DNA replication. In: DePamphilis ML, editor. DNA replication in eukaryotic cells. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press; 1996. pp. 217–247.

34. Throckmorton K, Lim FY, Kontoyiannis DP, Zheng W, Keller NP. Redundant synthesis of a conidial polyketide by two distinct secondary metabolite clusters in Aspergillus fumigatus. Environ Microbiol. 2015; doi:10.1111/1462-2920.13007

793 35. O'Hanlon KA, Cairns T, Stack D, Schrettl M, Bignell EM, Kavanagh K, et al. Targeted
794 Disruption of Nonribosomal Peptide Synthetase pes3 Augments the Virulence of
795 Aspergillus fumigatus. Infect Immun. 2011;79: 3978–3992. doi:10.1128/IAI.00192-11

796 36. Kapitonov V V., Jurka J. A universal classification of eukaryotic transposable elements
797 implemented in Repbase. Nat Rev Genet. Nature Publishing Group; 2008;9: 411–412.
798 doi:10.1038/nrg2165-c1

799 37. Mount SM, Rubin GM. Complete nucleotide sequence of the Drosophila transposable
800 element copia: homology between copia and retroviral proteins. Mol Cell Biol. 1985;5:
801 1630–8. Available: http://www.ncbi.nlm.nih.gov/pubmed/2410772

802 38. Sørensen JL, Hansen FT, Sondergaard TE, Staerk D, Lee TV, Wimmer R, et al. Production
803 of novel fusarielins by ectopic activation of the polyketide synthase 9 cluster in Fusarium
804 graminearum. Environ Microbiol. Blackwell Publishing Ltd; 2012;14: 1159–1170.
805 doi:10.1111/j.1462-2920.2011.02696.x

806 39. Futagami T, Mori K, Yamashita A, Wada S, Kajiwara Y, Takashita H, et al. Genome
807 Sequence of the White Koji Mold Aspergillus kawachii IFO 4308, Used for Brewing the
808 Japanese Distilled Spirit Shochu. Eukaryot Cell. 2011;10: 1586–1587.
809 doi:10.1128/EC.05224-11

810 40. Andersen MR, Salazar MP, Schaap PJ, van de Vondervoort PJI, Culley D, Thykaer J, et al.
811 Comparative genomics of citric-acid-producing Aspergillus niger ATCC 1015 versus
812 enzyme-producing CBS 513.88. Genome Res. 2011;21: 885–97.
813 doi:10.1101/gr.112169.110

814 41. Yu J, Wu G, Jurick WM, Gaskins VL, Yin Y, Yin G, et al. Genome Sequence of *Penicillium*
815 *solitum* RS1, Which Causes Postharvest Apple Decay. Genome Announc. 2016;4: e00363-
816 16. doi:10.1128/genomeA.00363-16

817 42. Yamada O, Machida M, Hosoyama A, Goto M, Takahashi T, Futagami T, et al. Genome
818 sequence of *Aspergillus luchuensis* NBRC 4314. DNA Res. 2016;23: 507–515.
819 doi:10.1093/dnares/dsw032

820 43. Cheeseman K, Ropars J, Renault P, Dupont J, Gouzy J, Branca A, et al. Multiple recent
821 horizontal transfers of a large genomic region in cheese making fungi. Nat Commun.
822 2014;5: 2876. doi:10.1038/ncomms3876

823 44. Baroncelli R, Sreenivasaprasad S, Sukno SA, Thon MR, Holub E. Draft Genome Sequence
824 of Colletotrichum acutatum Sensu Lato (Colletotrichum fioriniae). Genome Announc.
825 2014;2: e00112-14-e00112-14. doi:10.1128/genomeA.00112-14

826 45. Hacquard S, Kracher B, Hiruma K, Münch PC, Garrido-Oter R, Thon MR, et al. Survival
827 trade-offs in plant roots during colonization by closely related beneficial and pathogenic
828 fungi. Nat Commun. 2016;7: 11362. doi:10.1038/ncomms11362

829 46. King R, Urban M, Hammond-Kosack MCU, Hassani-Pak K, Hammond-Kosack KE. The
830 completed genome sequence of the pathogenic ascomycete fungus Fusarium
831 graminearum. BMC Genomics. 2015;16: 544. doi:10.1186/s12864-015-1756-1

832 47. Kazan K, Gardiner DM, Manners JM. On the trail of a cereal killer: Recent advances in
833 Fusarium graminearum pathogenomics and host resistance [Internet]. Molecular Plant
834 Pathology. 2012. pp. 399–413. doi:10.1111/j.1364-3703.2011.00762.x

835 48. Treangen TJ, Salzberg SL. Repetitive DNA and next-generation sequencing: computational
836 challenges and solutions. Nat Rev Genet. NIH Public Access; 2011;13: 36–46.
837 doi:10.1038/nrg3117

838 49. Palmer JM, Keller NP. Secondary metabolism in fungi: does chromosomal location
839 matter? Curr Opin Microbiol. 2010;13: 431–6. doi:10.1016/j.mib.2010.04.008

840 50. Gao Q, Jin K, Ying S-H, Zhang Y, Xiao G, Shang Y, et al. Genome sequencing and
841 comparative transcriptomics of the model entomopathogenic fungi Metarhizium
842 anisopliae and M. acridum. Achtman M, editor. PLoS Genet. Public Library of Science;
843 2011;7: e1001264. doi:10.1371/journal.pgen.1001264

844 51. Richards TA, Talbot NJ. Horizontal gene transfer in osmotrophs: playing with public
845 goods. Nat Rev Microbiol. Nature Research; 2013;11: 720–727. doi:10.1038/nrmicro3108

846 52. Morris JJ, Lenski RE, Zinser ER. The Black Queen Hypothesis: evolution of dependencies
847 through adaptive gene loss. MBio. American Society for Microbiology; 2012;3: e00036-
848 12. doi:10.1128/mBio.00036-12

849 53. Metzenberg RL, Glass NL. Mating type and mating strategies inNeurospora. BioEssays.
850 Wiley Subscription Services, Inc., A Wiley Company; 1990;12: 53–59.
851 doi:10.1002/bies.950120202

852 54. Smith CA, Want EJ, O'Maille G, Abagyan R, Siuzdak G. XCMS: Processing Mass
853 Spectrometry Data for Metabolite Profiling Using Nonlinear Peak Alignment, Matching,
854 and Identification. Anal Chem. 2006;78: 779–787. doi:10.1021/ac051437y

855 55. Espagne E, Balhadère P, Penin M-L, Barreau C, Turcq B. HET-E and HET-D belong to a new
856 subfamily of WD40 proteins involved in vegetative incompatibility specificity in the
857 fungus Podospora anserina. Genetics. 2002;161: 71–81. Available:
858 http://www.ncbi.nlm.nih.gov/pubmed/12019224

859 56. Gibbons JG, Salichos L, Slot JC, Rinker DC, McGary KL, King JG, et al. The evolutionary
860 imprint of domestication on genome variation and function of the filamentous fungus
861 Aspergillus oryzae. Curr Biol. 2012;22: 1403–9. doi:10.1016/j.cub.2012.05.033

862 57. Zhang H, Rokas A, Slot JCJ. Two different secondary metabolism gene clusters occupied
863 the same ancestral locus in fungal dermatophytes of the Arthrodermataceae. PLoS One.
864 2012;7: e41903. doi:10.1371/journal.pone.0041903

865 58. Wiemann P, Guo C-J, Palmer JM, Sekonyela R, Wang CCC, Keller NP. Prototype of an
866 intertwined secondary-metabolite supercluster. Proc Natl Acad Sci U S A. 2013;110:
867 17065–70. doi:10.1073/pnas.1313258110

868 59. Syme RA, Hane JK, Friesen TL, Oliver RP. Resequencing and Comparative Genomics of
869 *Stagonospora nodorum*: Sectional Gene Absence and Effector Discovery. G3:
870 Genes|Genomes|Genetics. 2013;3: 959–969. doi:10.1534/g3.112.004994

871 60. Chibucos MC, Crabtree J, Nagaraj S, Chaturvedi S, Chaturvedi V. Draft Genome
872 Sequences of Human Pathogenic Fungus Geomyces pannorum Sensu Lato and Bat White
873 Nose Syndrome Pathogen Geomyces (Pseudogymnoascus) destructans. Genome
874 Announc. 2013;1: e01045-13-e01045-13. doi:10.1128/genomeA.01045-13

875 61. de Man TJB, Stajich JE, Kubicek CP, Teiling C, Chenthamara K, Atanasova L, et al. Small
876 genome of the fungus *Escovopsis weberi* , a specialized disease agent of ant agriculture.
877 Proc Natl Acad Sci. 2016;113: 3567–3572. doi:10.1073/pnas.1518501113

878 62. Wu W, Davis RW, Tran-Gyamfi MB, Kuo A, LaButti K, Mihaltcheva S, et al.
879 Characterization of four endophytic fungi as potential consolidated bioprocessing hosts
880 for conversion of lignocellulose into advanced biofuels. Appl Microbiol Biotechnol.
881 Springer Berlin Heidelberg; 2017;101: 2603–2618. doi:10.1007/s00253-017-8091-1

882 63. de Vries RP, Riley R, Wiebenga A, Aguilar-Osorio G, Amillis S, Uchima CA, et al.
883 Comparative genomics reveals high biological diversity and specific adaptations in the
884 industrially and medically important fungal genus Aspergillus. Genome Biol. BioMed
885 Central; 2017;18: 28. doi:10.1186/s13059-017-1151-0

886 64. Lin H-C, Chooi Y-H, Dhingra S, Xu W, Calvo AM, Tang Y. The Fumagillin Biosynthetic Gene
887 Cluster in Aspergillus fumigatus Encodes a Cryptic Terpene Cyclase Involved in the
888 Formation of β-trans-Bergamotene. J Am Chem Soc. 2013;135: 4616–9.
889 doi:10.1021/ja312503y

890 65. Proctor RH, Van Hove F, Susca A, Stea G, Busman M, van der Lee T, et al. Birth, death and
891 horizontal transfer of the fumonisin biosynthetic gene cluster during the evolutionary
892 diversification of Fusarium. Mol Microbiol. 2013;90: 290–306. doi:10.1111/mmi.12362

893 66. Wong S, Wolfe KH. Birth of a metabolic gene cluster in yeast by adaptive gene relocation.
894 Nat Genet. Nature Publishing Group; 2005;37: 777–782. doi:10.1038/ng1584

895 67. Khaldi N, Collemare J, Lebrun M-H, Wolfe KH. Evidence for horizontal transfer of a
896 secondary metabolite gene cluster between fungi. Genome Biol. 2008;9: R18.
897 doi:10.1186/gb-2008-9-1-r18

898 68. Wisecaver JH, Rokas A. Fungal metabolic gene clusters-caravans traveling across
899 genomes and environments [Internet]. Frontiers in Microbiology. Frontiers; 2015. p. 161.
900 doi:10.3389/fmicb.2015.00161

69. de Hoog G, Guarro J, Gené J, Figueras M. Atlas of Clinical Fungi. Washington, DC: ASM Press; 2001.

70. Liu D, Zhang R, Yang X, Wu H, Xu D, Tang Z, et al. Thermostable cellulase production of Aspergillus fumigatus Z5 under solid-state fermentation and its application in degradation of agricultural wastes. Int Biodeterior Biodegradation. 2011;65: 717–725. doi:10.1016/j.ibiod.2011.04.005

71. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25: 1754–1760. doi:10.1093/bioinformatics/btp324

72. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. Oxford University Press; 2010;26: 841–2. doi:10.1093/bioinformatics/btq033

73. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. Cold Spring Harbor Laboratory Press; 2010;20: 1297–303. doi:10.1101/gr.107524.110

74. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, del Angel G, Levy-Moonshine A, et al. From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. Current Protocols in Bioinformatics. Hoboken, NJ, USA: John Wiley & Sons, Inc.; 2013. p. 11.10.1-11.10.33. doi:10.1002/0471250953.bi1110s43

75. DePristo MA, Banks E, Poplin R, Garimella K V, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 2011;43: 491–498. doi:10.1038/ng.806

76. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin). Taylor & Francis; 2012;6: 80–92. doi:10.4161/fly.19695

77. Zhou X, Peris D, Kominek J, Kurtzman CP, Hittinger CT, Rokas A. in silico Whole Genome Sequencer & Analyzer (iWGS): A Computational Pipeline to Guide the Design and Analysis of de novo Genome Sequencing Studies. G3: Genes|Genomes|Genetics. 2016; doi:10.1534/g3.116.034249

78. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. Mary Ann Liebert, Inc.; 2012;19: 455–77. doi:10.1089/cmb.2012.0021

79. Zimin A V., Marcais G, Puiu D, Roberts M, Salzberg SL, Yorke JA. The MaSuRCA genome assembler. Bioinformatics. 2013;29: 2669–2677. doi:10.1093/bioinformatics/btt476

80. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome

assemblies. Bioinformatics. 2013;29: 1072–1075. doi:10.1093/bioinformatics/btt086

81. Stanke M, Morgenstern B. AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. Nucleic Acids Res. Oxford University Press; 2005;33: W465-7. doi:10.1093/nar/gki458

82. Smit A, Hubley R, Green P. Repeatmasker Open-4.0 [Internet]. Available: http://www.repeatmasker.org

83. Inglis DO, Binkley J, Skrzypek MS, Arnaud MB, Cerqueira GC, Shah P, et al. Comprehensive annotation of secondary metabolite biosynthetic genes and gene clusters of Aspergillus nidulans, A. fumigatus, A. niger and A. oryzae. BMC Microbiol. 2013;13: 91. doi:10.1186/1471-2180-13-91

84. Bignell E, Cairns TC, Throckmorton K, Nierman WC, Keller NP. Secondary metabolite arsenal of an opportunistic pathogenic fungus. Philos Trans R Soc B Biol Sci. 2016;371.

85. Medema MH, Blin K, Cimermancic P, de Jager V, Zakrzewski P, Fischbach M a, et al. antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. Nucleic Acids Res. 2011;39: W339-46. doi:10.1093/nar/gkr466

86. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30: 1312–1313. doi:10.1093/bioinformatics/btu033

87. Zheng X, Levine D, Shen J, Gogarten SM, Laurie C, Weir BS. A high-performance computing toolset for relatedness and principal component analysis of SNP data. Bioinformatics. Oxford University Press; 2012;28: 3326–3328. doi:10.1093/bioinformatics/bts606

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

89. Eddy SR. A new generation of homology search tools based on probabilistic inference. Genome Inform. 2009;23: 205–11. Available: http://www.ncbi.nlm.nih.gov/pubmed/20180275

90. Katoh K, Standley DM. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Mol Biol Evol. 2013;30: 772–780. doi:10.1093/molbev/mst010

91. Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. 2009;25: 1972–1973. doi:10.1093/bioinformatics/btp348

92. Paradis E, Claude J, Strimmer K. APE: Analyses of Phylogenetics and Evolution in R language. Bioinformatics. Oxford University Press; 2004;20: 289–290.

973 doi:10.1093/bioinformatics/btg412

974 93. Schliep KP. phangorn: phylogenetic analysis in R. Bioinformatics. Oxford University Press;
975 2011;27: 592–3. doi:10.1093/bioinformatics/btq706

976 94. Shimizu K, Keller NP. Genetic involvement of a cAMP-dependent protein kinase in a G
977 protein signaling pathway regulating morphological and chemical transitions in
978 Aspergillus nidulans. Genetics. 2001;157: 591–600. Available:
979 http://www.ncbi.nlm.nih.gov/pubmed/11156981

980 95. Melamud E, Vastag L, Rabinowitz JD. Metabolomic Analysis and Visualization Engine for
981 LC−MS Data. Anal Chem. 2010;82: 9818–9826. doi:10.1021/ac1021166

982 96. Ahn JH, Walton JD. Chromosomal organization of TOX2, a complex locus controlling host-
983 selective toxin biosynthesis in Cochliobolus carbonum. Plant Cell. American Society of
984 Plant Biologists; 1996;8: 887–97. doi:10.1105/tpc.8.5.887

985

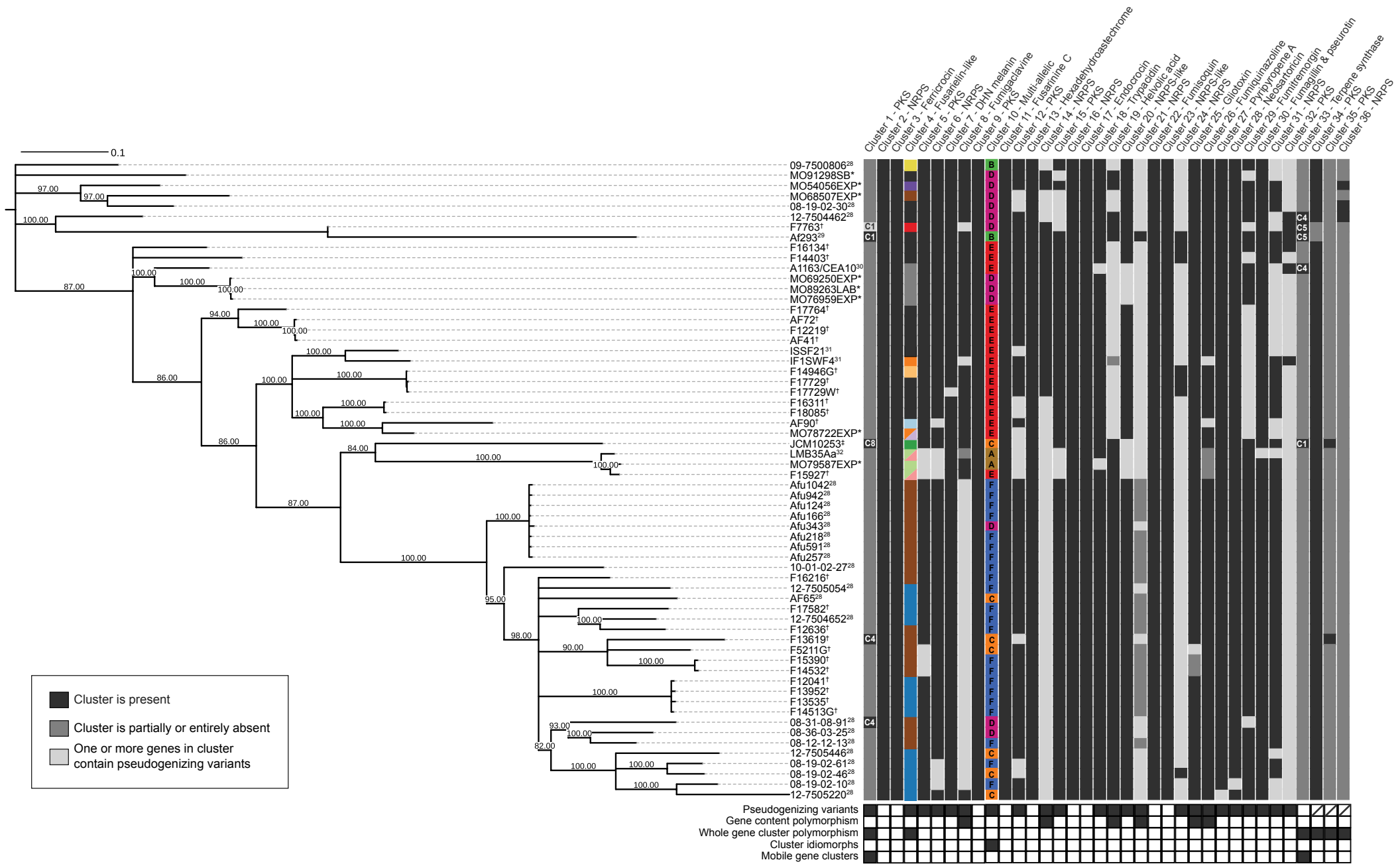


Figure 1. Genetic diversity of secondary metabolic gene clusters within a fungal species. The phylogeny was constructed using 15,274 biallelic SNPs with no missing data. The tree is midpoint rooted and all branches with bootstrap support less than 80% are collapsed. This phylogeny does not include strains Af10, Af210, Z5, or RP-2014 as short read data were not available. Superfixes following strain names indicate publications associated with DNA sequencing. * indicates strains sequenced in this study, † indicates strains sequenced at JCVI with no associated publication, and ‡indicates strains sequenced by RIKEN with no associated publications. Heatmaps show presence, absence and polymorphisms in SM gene clusters. Black indicates the cluster is present in a strain with no polymorphisms aside from missense variants, light gray indicates one or more genes in the cluster are pseudogenized, and dark gray indicates the cluster is partially or entirely absent. Colors for Cluster 4 indicate which pseudogenizing variants are present (see Figure 3) and colors for Cluster 10 indicate which allele of the cluster is present (see Figure 4). Chromosomal location of Clusters 1 and 33 are indicated. If more than one type of polymorphism is present within a cluster in a strain, only one is depicted. Types of polymorphisms found in each cluster are summarized below the cluster heatmap.

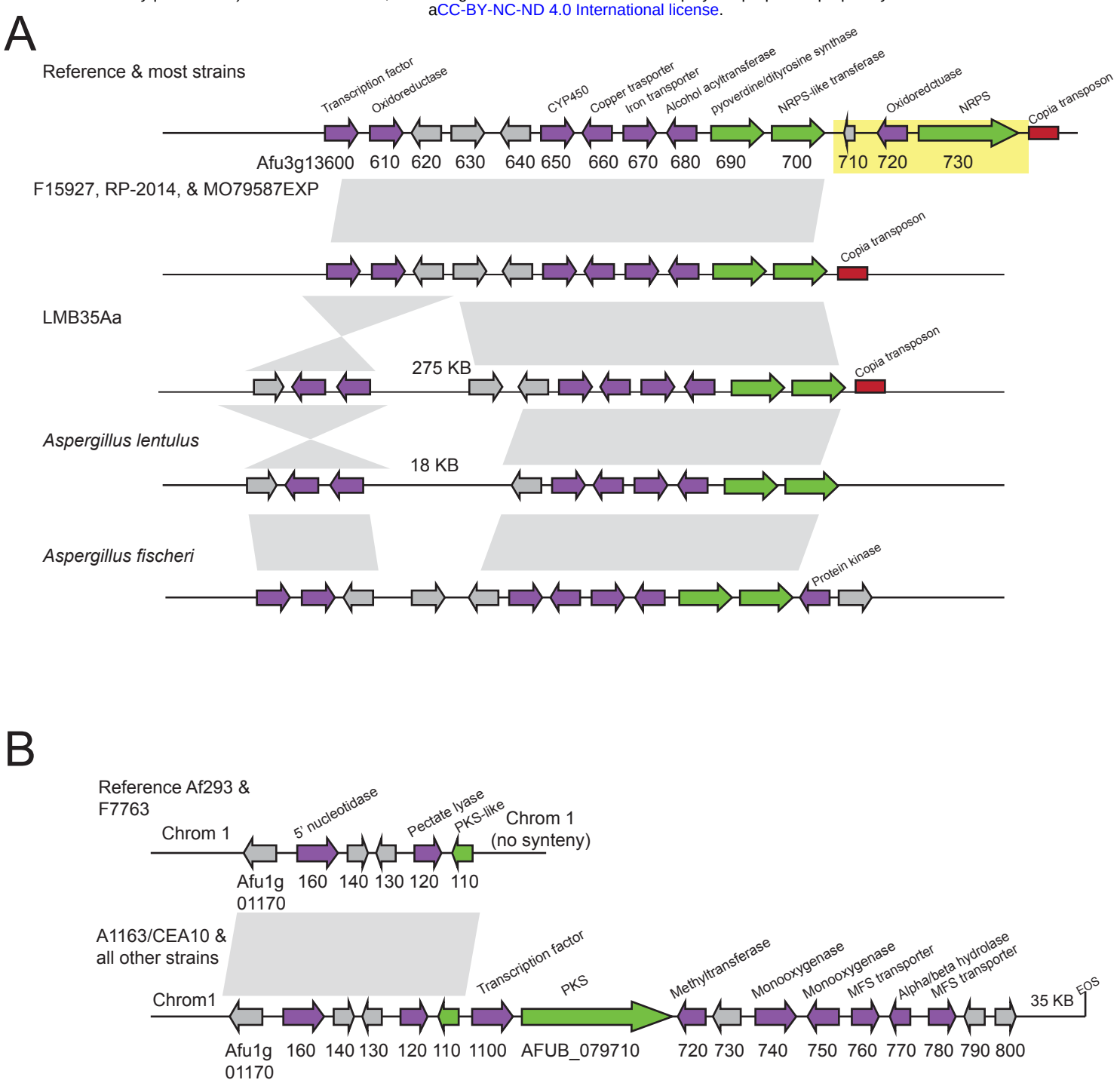
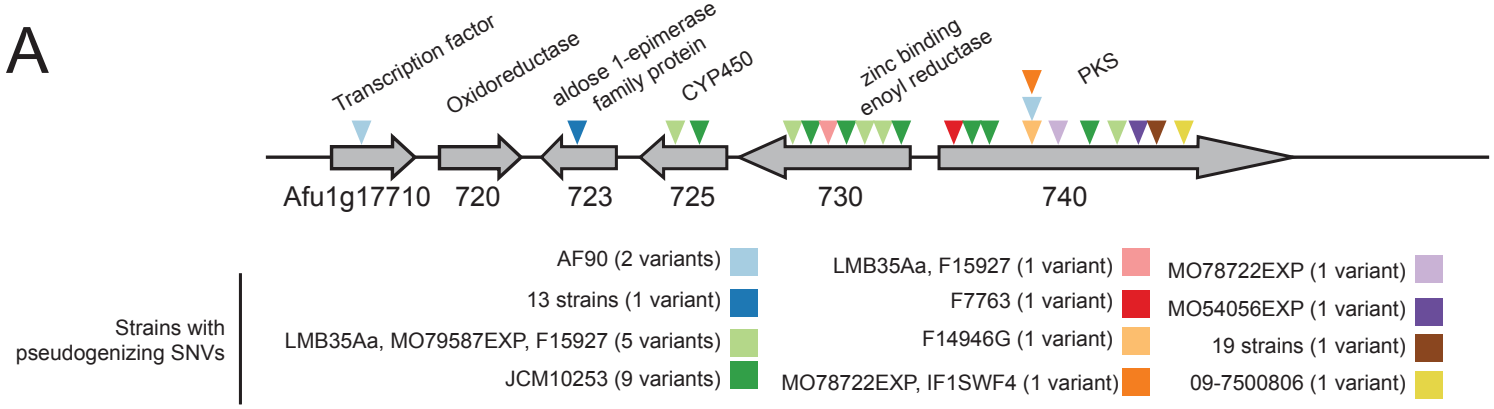


Figure 2. Gene gains and deletions in SM gene clusters. (A) Differences in gene content in SM gene cluster 14 in A. fumigatus strains and closely related species. Four A. fumigatus strains lack an 11-Kb region in this cluster, including an NRPS backbone gene, highlighted in yellow. Regions upstream and 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 fischeri and Aspergillus lentulus, close relatives of A. fumigatus, contain a cluster lacking the 11-kb region. (B) 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.



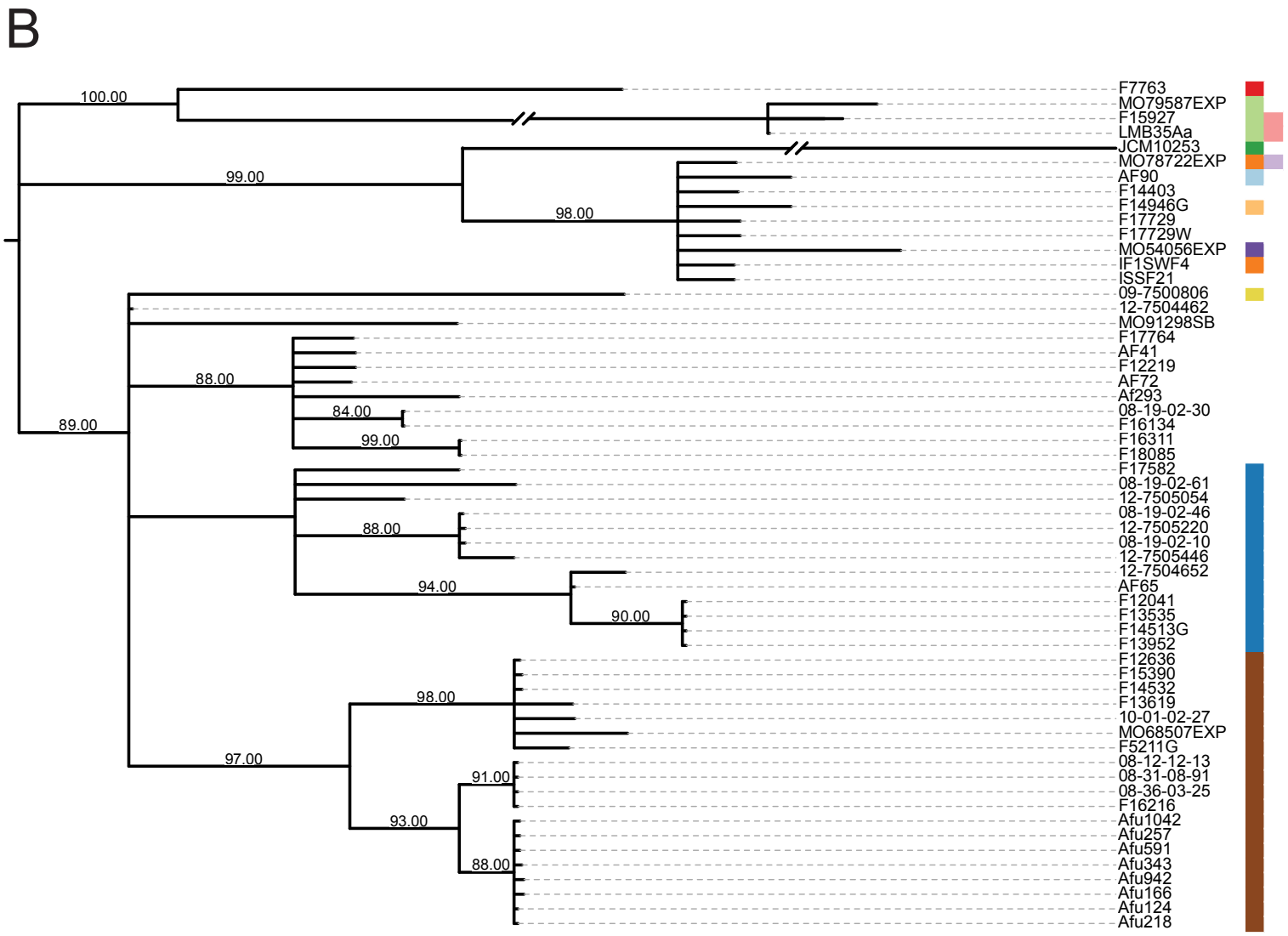


Figure 3. Pseudogenization in the fusarielin-like SM gene cluster. (A) Positions of frameshift variants and nonsense variants in the fusarielin-like SM gene cluster 4. (B) Locus phylogeny of the fusarielin-like SM gene cluster based on a nucleotide alignment of the entire gene cluster including intergenic and non-coding regions. The phylogeny is midpoint rooted and branches with bootstrap support <80% are collapsed. Two branches were shortened for visualization purposes. Strains with pseudogenizing variants are indicated with colored boxes. Colors correspond to variants shown in (A).

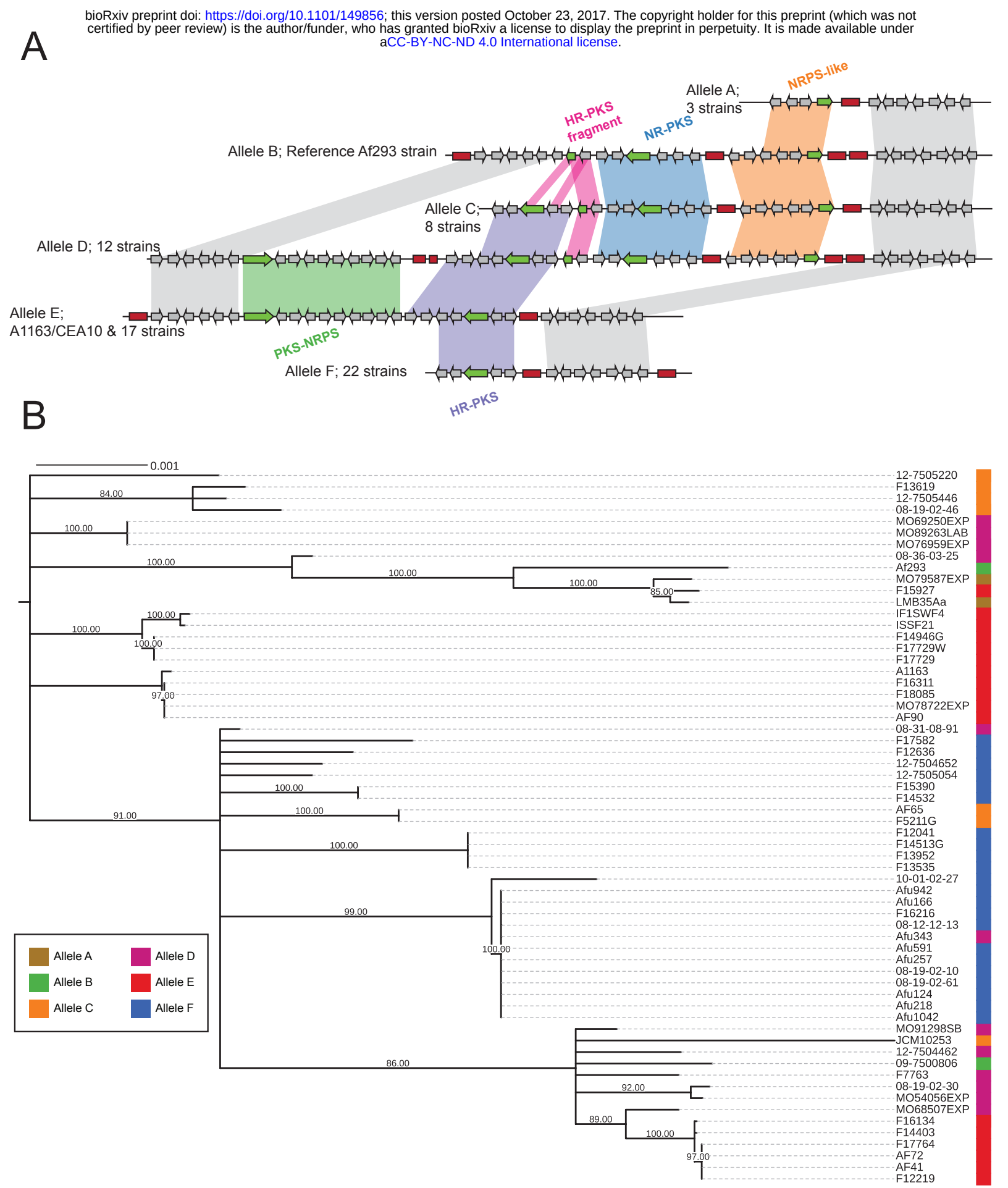


Figure 4. Six alleles of an idiomorphic SM gene cluster. (A) Alleles of SM gene cluster 10 on chromosome 3. Red boxes denote transposable elements. Green arrows denote backbone genes (PKS or NRPS). (B) Locus phylogeny of conserved downstream of the idiomorph cluster (highlighted in gray in A). Phylogeny was constructed using a 48 MB nucleotide alignment with the GTRGAMMA model and midpoint rooted. Branches with bootstrap support < 80% were collapsed.

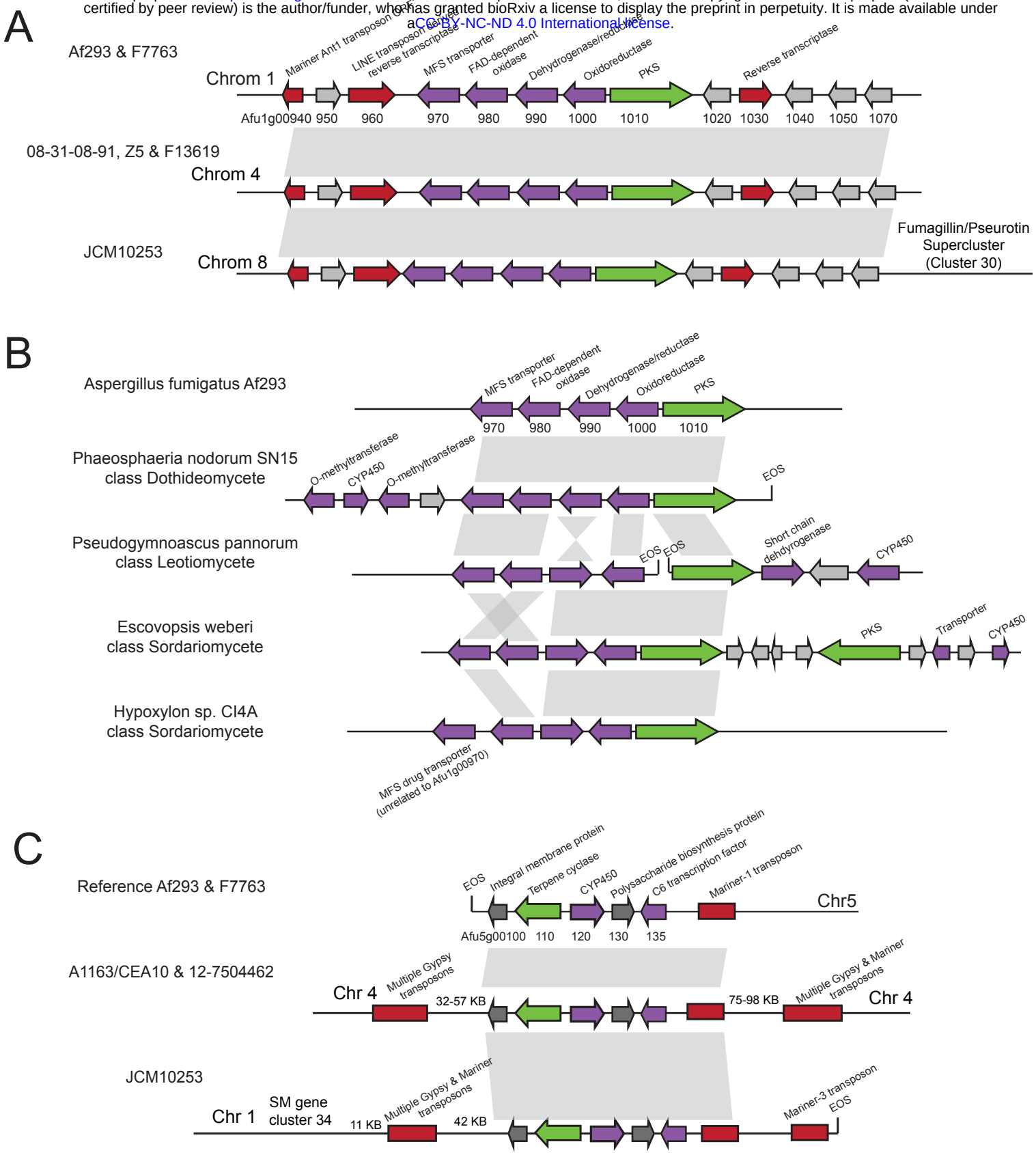


Figure 5. Multiple genomic locations of two SM gene clusters. (A) SM gene cluster 1 (Afu1g00970-01010) and flanking region is found in different genomic locations.The flanking regions contain transposon-derived open reading frames including two putative reverse transcriptases. In one strain, SM gene cluster 1 is found adjacent to SM gene cluster 30. (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. (C) SM gene cluster 33 (Afu5g00100-00135) is found in different genomic locations in diferent strains. In one strain, the cluster is adjacent to SM gene cluster 34. Multiple transposable elements flank the cluster in each strain.

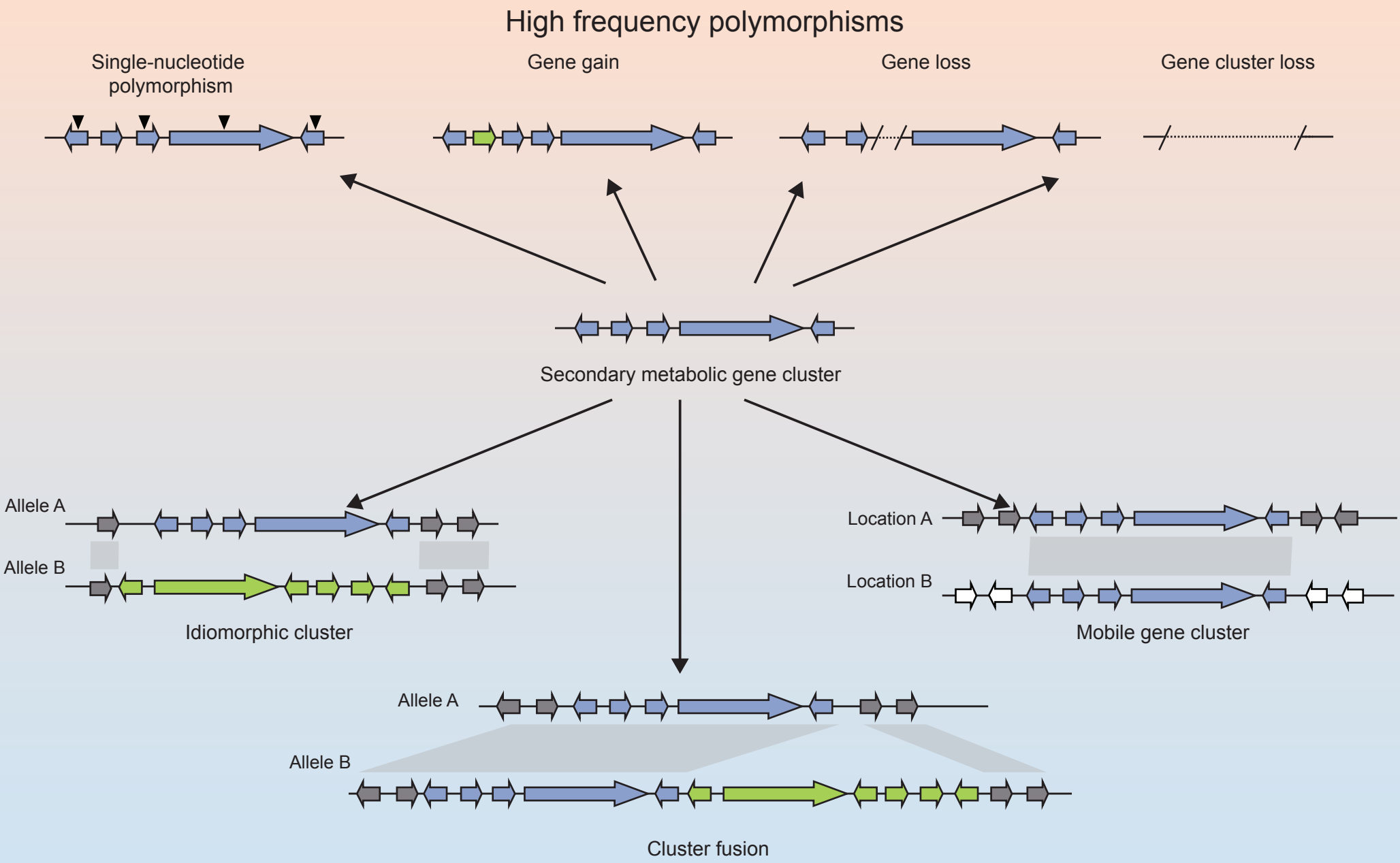


Figure 6. Types and frequencies of all SM gene cluster variants within *A. fumigatus*.