Genomic localization bias of secondary metabolite gene clusters and association with histone modifications in *Aspergillus*

Xin Zhang[1,2], Iseult Leahy[1], Jérôme Collemare[2,*,+], Michael F. Seidl[1,*,+]

[1]Theoretical Biology & Bioinformatics Group, Department of Biology, Utrecht University, Padualaan 8, 3584 CH Utrecht, the Netherlands; [2]Westerdijk Fungal Biodiversity Institute, Uppsalalaan 8, 3584 CT Utrecht, the Netherlands

[+]These authors contributed equally

Keywords: Aspergilli, secondary metabolite backbone gene, synteny, histone post-transcriptional modification, RNA-seq

Abstract

Fungi are well-known producers of bioactive secondary metabolites (SMs), which have been exploited for decades by humankind for various medical applications like therapeutics and antibiotics. SMs are synthesized by biosynthetic gene clusters (BGCs) – physically co-localized and co-regulated genes. Because BGCs are often regulated by histone post-translational modifications (PTMs), it was suggested that their chromosomal location is important for their expression. Studies in a few fungal species indicated an enrichment of BGCs in sub-telomeric regions; however, there is no evidence that BGCs with distinct genomic localization are regulated by different histone PTMs. Here, we used 174 *Aspergillus* species covering 22 sections to determine the correlation between BGC genomic localization, gene expression and histone PTMs. We found a high abundance and diversity of SM backbone genes across the *Aspergillus* genus, with notable unique genes within sections. Being unique or conserved in many species, BGCs showed a strong bias for being localized in low-synteny regions, regardless of their position in chromosomes. Using chromosome-level assemblies, we also confirmed a significantly biased localization in sub-telomeric regions. Notably, SM backbone genes in sub-telomeric regions and about half of those in low-synteny regions exhibit higher gene expression variability, likely due to the similar higher variability in H3K4me3 and H3K36me3 histone PTMs; while variations in histone H3 acetylation and H3K9me3 are not correlated to genomic localization and expression variation, as analyzed in two *Aspergillus* species. Expression variability across four *Aspergillus* species further supports that BGCs tend to be located in low-synteny regions and that regulation of expression in those regions likely involves different histone PTMs than the most commonly studied modifications.

37 **Significance**

38 Fungi are known for producing an array of bioactive compounds with medical benefits, yet our
39 understanding of how the production of these compounds is regulated remains limited. Here,
40 we focused on the fungal genus *Aspergillus*, containing many species known to be prolific
41 producers of bioactive compounds, to systematically uncover the diversity and genomic
42 localization of biosynthetic pathways. By expanding our knowledge beyond the few commonly
43 studied fungal species, this research offers novel insights into how the genomic localization of
44 biosynthetic pathways matters for the regulation of their expression. Thanks to a new view on
45 BGC localization and expression in relation to histone modifications, our results are expected
46 to stimulate functional research on neglected histone modifications that will support the
47 discovery and harnessing of new fungal metabolites for medical and industrial applications.

48 **Introduction**

49 Filamentous fungi are prolific producers of bioactive molecules, termed secondary metabolites
50 (SMs) or natural products (Mosunova et al., 2021). While these metabolites are not directly
51 necessary to sustain fungal growth, development or reproduction, they play indispensable
52 roles in inter-organismal interactions and fungal survival in diverse ecological niches (Keller,
53 2019). SMs have multifaceted functions, from acting as pigments that shield fungal spores
54 from UV radiation, and serving as virulence factors in pathogenic interactions, to providing
55 competitive edges through antibiotic properties (Keller, 2019). Humans have harnessed the
56 bioactive potential of these metabolites to significantly enhance their quality of life (Cairns et
57 al., 2021). Fungal SMs have found significant applications in medical treatments, serving as
58 therapeutic agents, exemplified by lovastatin and ergotamine, as well as immunosuppressants
59 like cyclosporine (Newman & Cragg, 2016; Wiemann & Keller, 2014). They also play crucial
60 roles as antibiotics, with penicillin being the most notable example, and as antifungal agents,
61 such as griseofulvin (Newman & Cragg, 2016).

62

63 The biosynthesis of fungal SMs is orchestrated by a series of enzymatic reactions. First,
64 backbone (or core) enzymes use precursors derived from primary metabolism to synthesize
65 an intermediate that is then further modified by so-called tailoring enzymes to produce the
66 final compound(s) (Mosunova et al., 2021). Backbone enzymes responsible for the synthesis
67 of diverse SMs have been classified into categories based on their enzymatic functions such
68 as polyketide synthase (PKS), non-ribosomal peptide synthetase (NRPS), PKS-NRPS hybrid
69 enzymes, dimethylallyl tryptophan synthetase (DMATS), terpene cyclase (TC) (Mosunova et
70 al., 2021), and ribosomally synthesized and post-translationally modified peptides (RiPPs)
71 synthetase (Kessler & Chooi, 2022). Genes encoding backbone and tailoring enzymes, as
72 well as genes responsible for transcription control, transport, and/or self-resistance, are co-
73 regulated and often physically co-localized along the chromosome in biosynthetic gene
74 clusters (BGCs) (Mosunova et al., 2021).

75

The discrepancy between the predicted bioactive potential based on genome mining and the number of known molecules effectively produced can be explained by the tight transcriptional control of SM biosynthetic pathways (Brakhage, 2013). In addition to the regulation by local and global transcription factors (Chiang et al., 2009; Macheleidt et al., 2016), SM production is regulated globally by chromatin dynamics that, in response to both internal and external stimuli, transitions between different chromatin states: namely the 'open' euchromatin, loosely packed and transcriptionally active state, and the 'closed' heterochromatin, tightly packed and transcriptionally silent state (Collemare & Seidl, 2019; Gacek & Strauss, 2012). Histone post-translational modifications (PTMs) – corresponding to chemical modifications on specific amino acids localized on the 'tails' of histone proteins - contribute to the control of the chromatin status. Most histone acetylations and some histone methylations on H3K4 (histone protein H3 at lysine 4) and H3K36 are associated with euchromatin, while histone methylations on H3K9 and H3K27 are associated with heterochromatin (Lawrence et al., 2016). The deletion of histone deacetylases, which changes the chromatin from 'closed' heterochromatin to 'open' euchromatin, increases the activity of nine out of 14 NRPSs in *Aspergillus fumigatus* (Lee et al., 2009) and the production of kojic acid in *Aspergillus niger* (Li et al., 2019). The knockout of histone acetyltransferases also leads to defects in SM biosynthesis in *Aspergillus nidulans* (Cánovas et al., 2014). Using inhibitors of histone PTMs like SAHA (suberoylanilide hydroxamic acid, inhibiting histone deacetylases) and 5-aza (5-azacytidine, inhibiting DNA methylation) has been widely employed to deregulate the expression of BGCs in diverse fungi (Brakhage, 2013).

Following the discovery that histone PTMs play an important role in the regulation of SM biosynthesis, it was hypothesized that the chromosomal location of BGCs was crucial for their regulation (Palmer and Keller, 2010). Indeed, several studies reported that BGCs tend to be preferentially located at, or close to, sub-telomeric regions, such as in *Fusarium graminearum* (Connolly et al., 2013), and in some well-studied Aspergilli: namely *Aspergillus oryzae* (Kjærbølling et al., 2020), *Aspergillus flavus* (Yang et al., 2022), *A. fumigatus* (Keller, 2019;

104 Perrin et al., 2007), and *A. nidulans* (Klejnstrup et al., 2012). Sub-telomeric regions in fungi
105 are rich in repetitive sequences such as transposable elements and are typically associated
106 with histone PTMs linked to heterochromatin, especially with the presence of H3K9me3 and
107 H3K27me3 (Keller, 2019; Palmer & Keller, 2010; Wiemann et al., 2013). For instance, in
108 *Fusarium fujikuroi*, the PKS19 BGC, which produces fujikurins, is embedded within a region
109 rich in H3K9me3 on the long arm of chromosome VIII (Wiemann et al., 2009). Similarly, in *A.*
110 *nidulans*, several BGCs occur in proximity to H3K9me3-rich regions (Gacek-Matthews et al.,
111 2016). Thus, the peculiar genomic localization of BGCs might explain why these BGCs remain
112 transcriptionally silent. Moreover, because of their high repeat content, sub-telomeric regions
113 are prone to rearrangements and typically exhibit low synteny, characterized by reduced
114 conservation of physical co-localization of genes compared to other species (Nakao et al.,
115 2009). In *A. nidulans*, compared to *A. oryzae* and *A. fumigatus*, BGCs often occur in less
116 conserved and heavily rearranged genomic regions, even though the majority (77–79%) of
117 the genomes show conserved synteny (Inglis et al., 2013). However, not all SM gene clusters
118 are associated with heterochromatin and sub-telomeric regions. While previous studies have
119 predominantly focused on a limited number of well-known *Aspergillus* sections and species,
120 often finding an association of SM gene clusters with heterochromatin and sub-telomeric
121 regions, it is crucial to note that not all SM gene clusters are confined to these areas. To
122 determine how the genomic localization and regulation of BGCs are linked, we examined the
123 correlation among the BGC diversity, their genomic organizations, gene expression, and
124 histone PTMs together. We here include 174 diverse Aspergilli species covering 22 sections
125 to provide a more nuanced analysis of the distribution, conservation, and genomic localization
126 of SM genes, particularly in relation to genome-wide syntenic regions and histone PTMs using
127 a sub-set of species with chromosome-level assemblies. This extensive approach allows us
128 to explore the implications of BGC localization and histone PTMs on SM gene expression
129 regulation.
130
131 **Results and discussions**

132 **A comprehensive phylogenetic analysis of the fungal genus *Aspergillus***

133 To provide a comprehensive analysis of BGCs in Aspergilli, we included 174 publicly available
134 Aspergilli genomes (Kjærbølling et al., 2018, 2020; Rasmussen, 2017; Theobald, 2018;
135 Theobald et al., 2018; Vesth et al., 2018) and three *Penicillium* species derived from the Joint
136 Genome Institute (Table S1) as outgroups. The assembled genomes range in size from 23.2
137 to 42.9 Mb and 90% of them fall within the range of 25.8 to 38.8 Mb (Figure 1, Table S1). Four
138 *Aspergillus* species - *A. fumigatus*, *A. nidulans*, *A. niger*, and *Aspergillus ochraceus* - have
139 chromosomal-level genome assemblies (Nierman et al., 2005), which are comprised of eight
140 chromosomes (Figure 1, Table S1). The other well-studied model system, *A. oryzae*, has also
141 a good genome assembly of eleven scaffolds. Twenty-four Aspergilli assemblies have 100 or
142 fewer scaffolds, while *Aspergillus rambellii* has the most fragmented assembly with 4,177
143 scaffolds (Figure 1, Table S1). The number of predicted protein-coding genes ranges from
144 7,761 to 15,687, with 90% of Aspergilli having between 9,561 and 14,061 genes (Figure 1,
145 Table S1). Because the genome contiguity and the number of genes varied between different
146 Aspergilli, we sought to calculate the genome completeness by querying each genome
147 assembly for the occurrence of single-copy orthologous BUSCO genes (Table S2).
148 Importantly, we observed very high genome completeness with 90% of all Aspergilli scoring
149 higher than 98.0% BUSCO completeness (Figure 1, Table S1). Even *A. rambellii* with the most
150 fragmented assembly and the fewest genes (7,761) attains a high genome completeness of
151 97.1%. Similarly, four other Aspergilli with more than 1,000 scaffolds also have high genome
152 completeness (*Aspergillus udagawae* (97.8%), *Aspergillus avenaceus* (99.1%), *Aspergillus*
153 *haitiensis* (98.3%), and *Aspergillus coremiiformis* (99.5%)) (Figure 1, Table S1). Although
154 some regions might have been missed in highly fragmented genomes, we considered all
155 species included in this study exhibit sufficient quality to perform accurate comparative
156 analyses.

157

158 To provide a robust phylogenetic framework to study BGC and genomic diversity in Aspergilli,
159 we constructed a species phylogeny based on the concatenated alignment of 758 conserved,

160 single-copy BUSCO genes (in total 132,397 informative sites in gene alignments). The tree
161 confirms 22 distinct *Aspergillus* sections, each supported by bootstrap values exceeding 99%
162 (Figure 1), suggesting that our phylogenic analysis is robust. Moreover, this species phylogeny
163 confirms the associations among distinct *Aspergillus* sections as reported in previous
164 studies(Steenwyk et al., 2024). The tree enlarges the current taxonomic breadth as we
165 included a much larger set of diverse 174 *Aspergillus* species compared to less than 90
166 Aspergilli in previous publications (Steenwyk et al., 2019; Visagie et al., 2014, 2023; Zhang et
167 al., 2022). The 22 sections exhibit notable variation in the number of sequenced species:
168 section *Nigri* has 28 sequenced species, whereas sections *Janorum*, *Bispori*, and *Silvati* are
169 each represented by only a single sequenced species (Figure 1). Despite these differences,
170 the genome assemblies in most sections are relatively uniform in regard to assembly length,
171 number of scaffolds, genome completeness, and gene count, and thus provide a
172 comprehensive and robust reference to study the diversity and evolution of Aspergilli and as
173 well as a solid framework for further analysis.
174
175 **Aspergilli harbor an abundant secondary metabolite biosynthetic gene cluster**
176 **repertoire**
177 To evaluate the potential for SM biosynthesis encoded in the genomes of *Aspergillus* species,
178 we focused on SM backbone genes as these genes are crucial for the first step in SM
179 biosynthesis (Keller, 2019). The focus on SM backbone genes also prevents inaccurate
180 comparison of BGC split over several scaffolds in fragmented assemblies. We used
181 antiSMASH (Blin et al., 2021) to identify regions that comprise a predicted BGC, and employed
182 an in-house BGCtoolkit to further classify SM backbone enzymes into groups that correspond
183 to major chemical families and to retrieve their sequences.
184
185 We identified in total 16,243 SM backbone genes across 174 *Aspergillus* species and the
186 three selected *Penicillium*, ranging from 18 in *Aspergillus caninus* to up to 160 in *Aspergillus*
187 *implicatus* (Figure 1, Table S3). On average, we identified 92 SM backbone genes per

188 Aspergilli compared to only 77 per *Penicillium* species. These SM backbone genes are
189 unevenly distributed. The 41 early-diverging Aspergilli (comprising sections *Polypaecilum*,
190 *Cremei*, *Restricti*, *Aspergillus*, *Cervini*, *Clavati*, and *Fumigati*) have on average 65 SM
191 backbone genes, while the other 133 Aspergilli have on average 99. The *Circumdati* section
192 has the highest average number of SM backbone genes, with its 18 species encoding an
193 average of 129 SM backbone genes. As one BGC can contain several SM backbone genes,
194 the average number of SM backbone genes in Aspergilli is higher than the number of predicted
195 BGCs that have been reported in previous studies. For instance, studies that focused on
196 sections *Nigri* (Theobald et al., 2018), *Flavi* (Kjærbølling et al., 2020) using a customized
197 SMURF approach have detected on average 82 and 70 BGCs, respectively. Here, we found
198 these sections have an average of 104 (section *Nigri*) and 101 (section *Flavi*) SM backbone
199 genes detected with antiSMASH. The tool for BGC prediction greatly influences the numbers
200 as an average of 83 SM backbone genes were reported in section *Terrei* with the SMURF-
201 derived prediction(Theobald et al., 2024), while we predicted 100 SM backbone genes on
202 average. Yet, the trends are similar as both studies report *Aspergillus floccosus* as the richest
203 species in SM backbone genes within the *Terrei* section (Theobald et al., 2024). Five groups
204 of SM backbone genes stood out due to their abundance in all genomes: NRPS-like with 3,888,
205 rPKS (reducing Polyketide Synthase) with 3,175, NRPS with 2,943, nrPKS (non-reducing
206 Polyketide Synthase) with 1,395, and DMATS with 1,203 backbone genes. In 145 species,
207 and the three selected *Penicillium*, NRPS-like, rPKS, and NRPS are the three most abundant
208 SM backbone genes. In contrast, ten species in sections *Flavipedes* and *Fumigati* have highly
209 abundant DMATS, and other 22 Aspergilli have also many nrPKSs. Interestingly, we also
210 observed variation in the number of SM backbone genes within sections. For example, within
211 the *Nigri* section, two sub-groups could be defined based on their morphology, biseriates and
212 uniseriates which differ in the attachment of phialides to the vesicle. Among them biseriates
213 tend to have more SM backbone genes harbored in their genomes compared to uniseriates
214 (Nielsen et al., 2009). Similarly, two sub-groups in the *Fumigati* section show a diverse pattern
215 of SM backbones, the clade to which *A. fumigatus* belongs having a higher number of SM

backbone genes. In summary, our findings underscore that the SM repertoires are diverse and particularly vary between sections.

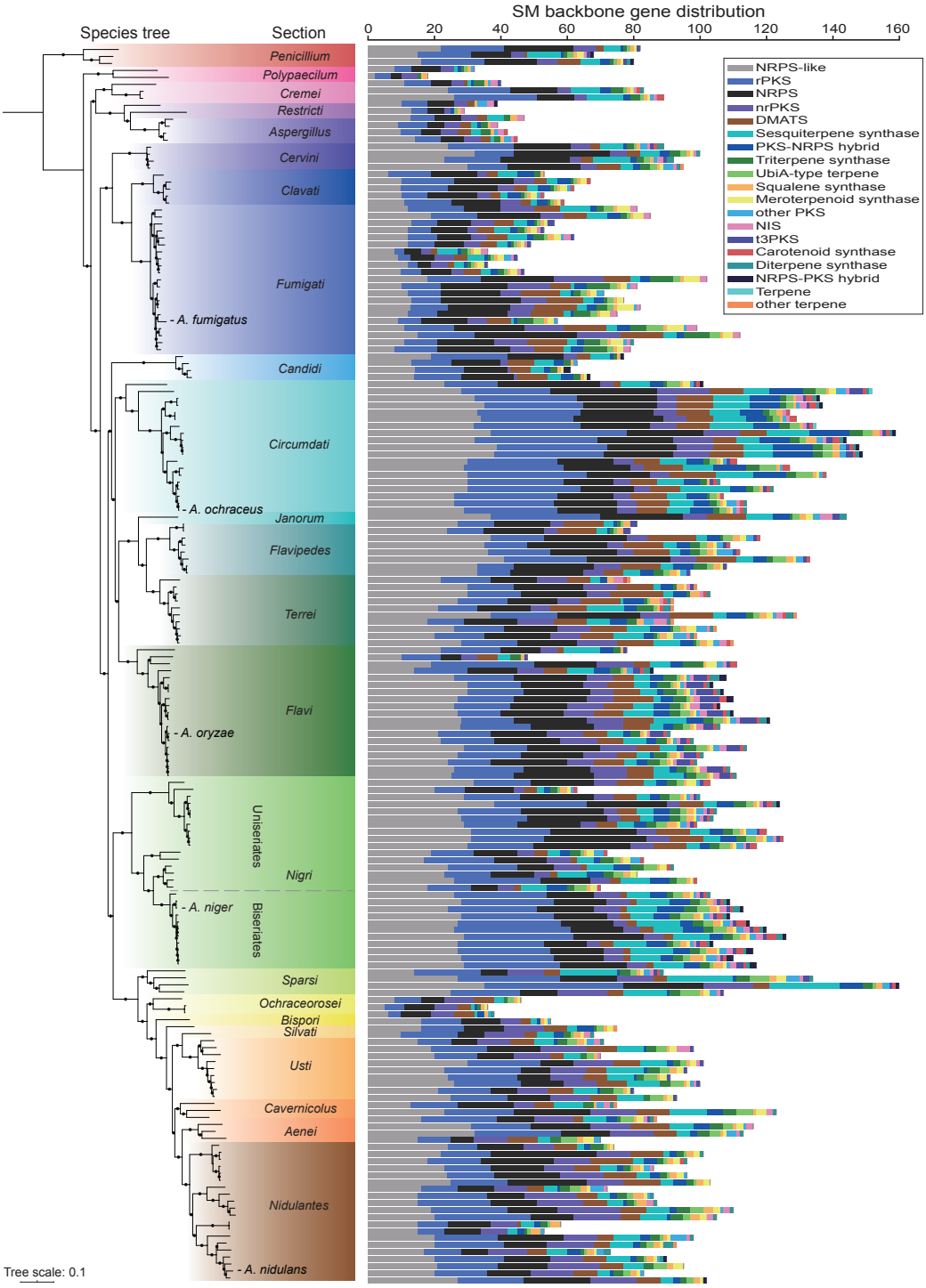


Figure 1 Phylogenetic relationship between *Aspergillus* specie and their secondary metabolite (SM) backbone gene complement. The phylogenetic relationships between 174 *Aspergillus* and three *Penicillium* species are built using a maximum-likelihood framework implemented in IQ-TREE using a concatenated alignment of 758 single-copy BUSCO genes. The dots on the branches indicate ultrafast bootstrap values over 95 and SH-aLRT (Shimodaira-

225 Hasegawa-like approximate likelihood ratio test) values over 80. Species from the same section
226 are highlighted by the same background color. The names of five well-studied Aspergilli are shown
227 in the tree and a dashed line highlights the separation of sub-groups within the *Nigri* section. The
228 smallest and largest dot sizes correspond to the minimum and maximum number in each category,
229 respectively. Genome completeness is indicated by the percentage of identified complete, single-
230 copy BUSCO genes in each genome assembly. SM backbone genes were first detected with
231 antiSMASH, then reclassified by BGCtoolkit. Bar plots representing different SM backbone groups
232 are arranged from left to right from the most to the least abundant.
233

234 Previous studies used BGC network approaches to assess BGC conservation within
235 *Aspergillus* sections (Kjærbølling et al., 2020; Vesth et al., 2018), which overestimates the
236 BGC diversity as variation in gene content at the loci of predicted BGCs will directly result in
237 non-conserved BGCs. In contrast, we here utilized a phylogeny-aware method (Figure S1) to
238 determine orthologous groups for each type of SM backbone gene. To obtain an estimate of
239 gene conservation in Aspergilli, we used OrthoFinder to identify orthologous groups for all
240 genes among all genomes. We included in total 1,546,732 orthologs as defined by
241 OrthoFinder (Table S4) and 16,638 SM backbone genes to test their uniqueness among
242 Aspergilli. From all OrthoFinder entries, we identified 34,144 unique genes (2.2%) among all
243 species, for which the estimated growth rate β is 0.27 (Figure 2A), where β represents the rate
244 of growth of unique genes as the number of sampled species increases (Heaps, 1978). For
245 SM backbone genes, 2,426 out of 16,638 (14.6%) are unique, for which the growth rate β is
246 0.45 (Figure 2B). These β rates imply that sequencing more *Aspergillus* species is expected
247 to expand the overall genomic diversity and specifically increase the diversity within the set of
248 SM backbone genes. Interestingly, a recent study observed more diverse SM genes
249 compared to pangenomes at the population level (Steenwyk et al., 2023), suggesting a large,
250 yet untapped diversity of SMs in Aspergilli. However, in our study, saturation is reached for
251 many SM groups, but not for NRPS and NRPS-like, which indicates that we might have found
252 nearly all polyketide and terpene biosynthetic pathways, but many non-ribosomal peptides
253 likely remain to be discovered. Lastly, we observed a similar trend for both unique orthologues
254 and SM backbone genes with continuous slight increases within sections and small jumps

255 between sections (Figures 2C and 2D), suggesting that sections are indeed characterized by
256 a sudden increase in BGC diversity that is consistent with divergence.
257
258 We observed that SM backbone genes are diverse and that increased diversity is observed
259 across different *Aspergillus* sections. However, the number of unique genes is lower than
260 estimated in three recent studies focused solely on the *Nigri* (Vesth et al., 2018), *Flavi*
261 (Kjærbølling et al., 2020) or *Terrei* (Theobald et al., 2024) sections that found around half of
262 the SM BGCs were predicted to be species-specific. Here, we only focused on SM backbone
263 genes and implemented a phylogeny-aware approach to evaluate SM backbone gene
264 diversity across a diverse set of *Aspergillus* species, while these previous studies used
265 complete SM gene cluster diversity and relied on a limited taxonomic sampling. These
266 previous studies thus overestimated the BGC diversity within *Aspergillus* sections. In contrast,
267 our results are more conservative and underestimate the true biochemical diversity that is
268 provided by the diversity of additional tailoring genes within BGCs that can lead to the
269 production of structurally related yet different secondary metabolites. However, our approach
270 provides broader and more detailed estimations regarding the type of metabolite and BGCs
271 that remain to be discovered in *Aspergillus* species.
272

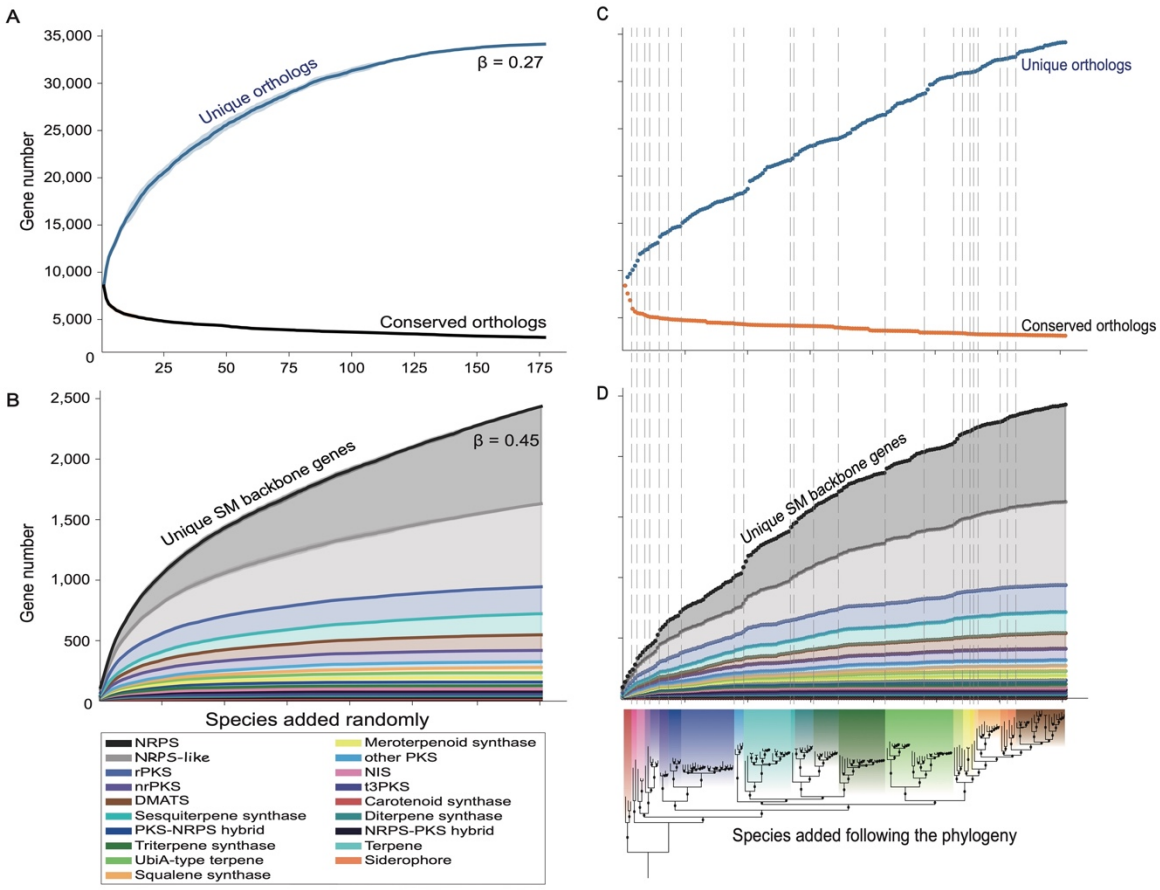


Figure 2 Evaluation of the repertoire of unique secondary metabolite (SM) backbone genes in Aspergilli. (A) The saturation curve plots the number of unique and conserved orthologous genes when species are added randomly. The calculations were performed ten times, and the standard deviations are shown as shadows around the curve. (B) Saturation curves plot the stacked number of unique SM backbone genes when species are added randomly. Different colors indicate distinct groups of SM backbone genes. The calculations were performed 10 times, and the standard deviations are shown as shadows around the curve. (C) Saturation curves of unique and conserved orthologous genes when species are added according to their taxonomy, starting with the outgroup, and the dash lines indicate distinct taxonomic sections. (D) Saturation curves of unique SM backbone genes with the species added according to their taxonomy and the dash lines indicate distinct taxonomic sections. A phylogenetic species tree is shown at the bottom for reference. PKS: polyketide synthase; NRPS: non-ribosomal peptide synthetase; DMATS: dimethylallyl tryptophan synthase; NIS: NRPS-independent siderophore.

Genome-wide synteny in Aspergilli is higher within than between sections

The tight regulation of BGC expression and the role of chromatin conformation in this regulation suggested that the genomic localization of BGCs might be important (Palmer et al., 2010). Several studies have indicated an over-representation of SM BGCs in sub-telomeric regions (Guzmán-Chávez et al., 2018; Keller, 2019; Kjærbølling et al., 2020; Klejnstrup et al.,

294 2012). These regions are known to be variable between strains and species, and
295 consequently, they display low levels of synteny. The term 'synteny' here refers to the
296 conservation of genomic regions that are maintained in a similar arrangement across different
297 genomes through evolutionary time (Drillon et al., 2013) (Figure 3A). As most of the
298 *Aspergillus* species do not have chromosome-level assemblies, we assessed BGC
299 localization by examining synteny throughout the *Aspergillus* genus. We performed an all-
300 versus-all pairwise comparison, and we uncovered, in general, that the amount of shared
301 synteny between species (synteny percentage) correlates with the divergence and is
302 consistent with the phylogenetic relationship between species (Figure 3B); *i.e*., two *Aspergillus*
303 sections that are more syntenic indeed diverged more recently. Based on this observation, we
304 propose that the *Nigri* section could be split in two, which is also coherent with the phylogeny.
305 Similarly, the *Circumdati*, *Flavi*, and *Nidulantes* sections show distinct groups of higher
306 synteny that might indicate ongoing divergence within these sections. Section *Nidulantes*
307 exhibits lower synteny, which seems coherent with the branch lengths in this clade compared
308 to other clades in which the divergence between species appears more recent than in the
309 *Nidulantes* section. Finally, early-diverging sections (*i.e., Polypaecilum*, *Cremei*, *Restricti*,
310 *Aspergillus*, *Cervini*, *Clavati*, and *Fumigati*) show high synteny within sections compared with
311 other sections (Figure 3C). Overall, the observed synteny between sections is low, ranging
312 from only 34% to 54%, with a median of 46%. The same analysis performed with only highly
313 contiguous assemblies yielded the same results (Figure S2B) (Liu et al., 2018).
314

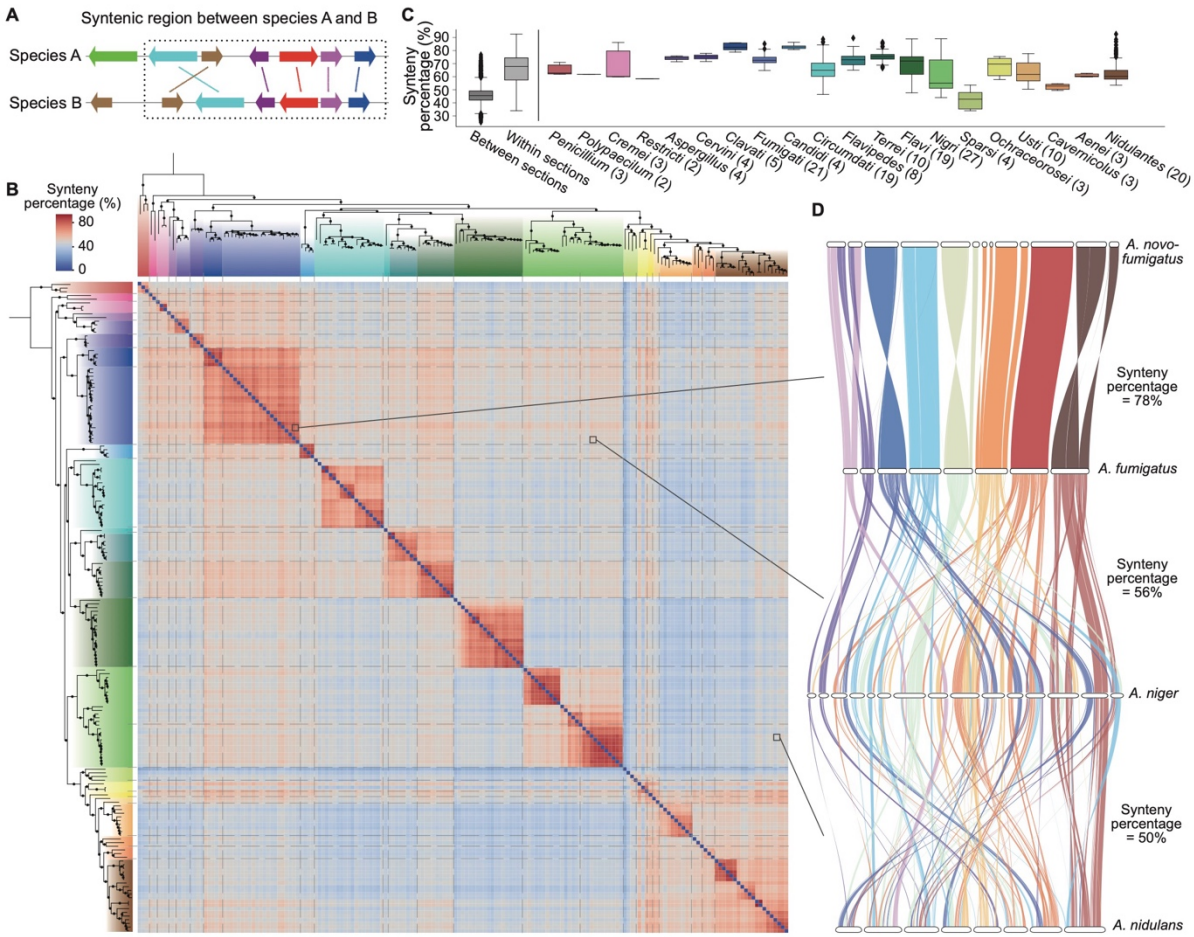


Figure 3 Conservation of synteny within and between *Aspergillus* sections. (A) A graphical representation of a hypothetical syntenic region between two species is demarcated by a dashed box. Arrows of the same color, linked by a line, represent homologous genes from two different species, with the arrow direction indicating the gene's orientation. (B) The heatmap shows pairwise synteny expressed as the synteny percentage based on all-versus-all comparisons, arranged based on the phylogenetic tree (Figure 1). The species tree, displayed to the left and top of the heatmap, uses colors to differentiate between sections (see legend for Figure 1 for the color coding). (C) Boxplots show the quantification of synteny by the 'synteny percentage' between sections, within sections, and in each section of the *Aspergillus* genus. Numbers in the parentheses indicate the number of species in the sections, and only sections with more than one species are shown. (D) Genome alignment between two species from the same section (top) and different sections (middle and bottom) highlights the degradation of synteny when taxonomic distance increases.

In Ascomycota, synteny varies across distantly related species and is categorized into micro-, meso-, and macro-synteny (Hane et al., 2011). Micro-synteny refers to the conservation of a small number of successive genes, typically ranging from two to ten. Meso-synteny represents larger, yet fragmented, synteny blocks compared to micro-synteny, but shorter than 20 thousand base pairs (kbp). Macro-synteny, on the other hand, is characterized by large and

335 continuous blocks extending over 20 kbp (Hane et al., 2011). Our findings align with the
336 expectations, revealing longer synteny blocks within sections than between them (Figure S2).
337 For instance, species in the *Fumigati* section show extensive regions of shared synteny across
338 chromosomes (Figure 3D). In contrast, species from different sections exhibit markedly
339 reduced synteny, both in percentage and block length (Figures 3D and S2). A pattern of
340 'degraded macrosynteny' was previously identified in several filamentous Ascomycota fungi,
341 including some Aspergilli (Hane et al., 2011). Our data suggest that within *Aspergillus* sections,
342 synteny blocks with an approximate length of 150 kbp are indicative of macrosynteny. In
343 contrast, between different sections, synteny blocks averaging 70 kbp to 150 kbp represent
344 degraded macrosynteny, likely reflecting the divergence time among these sections.
345
346 **SM backbone genes tend to reside in low-synteny and sub-telomeric regions**
347 To investigate the genomic localizations of SM genes in Aspergilli, we analyzed four species
348 with chromosomal-level assemblies: *A. nidulans*, *A. fumigatus*, *A. niger*, and *A. oryzae*
349 (Figures 4 and S3-S5). In each of the four species, the gene density shows that predicted
350 genes are evenly distributed throughout the genome, except for putative centromere and
351 telomere regions (as defined by the terminal 10% regions of chromosomes), which are known
352 to contain fewer genes but harbor repetitive sequences, and exhibit unique structural and
353 functional characteristics distinct from the rest of the genome (Xu et al., 2022). BUSCO genes,
354 which are conserved across the fungal kingdom and crucial for organismal function, are
355 significantly enriched in regions with high synteny, both within individual sections (91% of
356 BUSCO genes; Chi^2 *p*-value = 3.15^{e-117}) and across the *Aspergillus* genus (83%; Chi^2 *p*-value
357 = 0) compared with all genes (59% and 42%, respectively) (Figure 5B and 5C). In contrast,
358 SM backbone genes are significantly enriched in low-synteny regions in all four species at the
359 *Aspergillus* genus level (81% of backbone genes; Chi^2 *p*-value = 6.12^{e-58}) (Figures 5B and 5C,
360 Table S6). For the section-level synteny scores, though only 24% of SM backbone genes are
361 found in low-synteny regions, the trend remains significant compared with all genes (14%;
362 Chi^2 *p*-value = 7.77^{e-28}) and clearly differs from BUSCO genes (2%) (Figures 5B and 5C,

363 Tables S6). In regard to chromosomal location, only 21% of all genes are located in sub-
364 telomeric regions, and BUSCO genes are significantly located outside of these regions (7% in
365 sub-telomeric regions; Chi^2 *p*-value = 4.34^{e-72}) (Figures 5D and 5E). In contrast, SM backbone
366 genes show a trend to be located in the sub-telomeric regions (36%; Chi^2 *p*-value = 1.28^{e-12})
367 (Figures 5D and 5E), aligning with previous reports in *A. nidulans* (Klejnstrup et al., 2012), *A.*
368 *fumigatus* (Keller, 2019; Perrin et al., 2007), *A. oryzae* (Kjærbølling et al., 2020), as well as in
369 other fungal species such as *F. fujikuroi* (Studt et al., 2017) and *Penicillium chrysogenum*
370 (Guzmán-Chávez et al., 2018). However, the enrichment of SM backbone genes in sub-
371 telomeric regions is less significant than in low-synteny regions, suggesting that association
372 with low-synteny regions is an attribute of SM BGCs. The localization of SM backbone genes
373 in low-synteny and sub-telomeric regions of *Aspergillus* genomes may facilitate rapid genetic
374 rearrangement of these BGCs, thereby enabling rapid adaptation to diverse ecological niches.
375 Indeed, such a localization is coherent with the observed isolate- or population-specific
376 presence-absence pattern reported in *A. nidulans* (Drott et al., 2020), *A. fumigatus* (Lind et al.,
377 2017) and *A. flavus* (Drott et al., 2021). Most polymorphisms detected in *A. fumigatus*
378 population were suggested to involve transposable elements and/or recombination (Lind et
379 al., 2017), which are mechanisms likely to break synteny and could favor this bias in BGC
380 genomic localization. Noteworthy, SM backbone genes with known metabolite structures, as
381 listed in the MIBiG database (Table S5), tend to occur more frequently in sub-telomeric areas.
382 For instance, four out of six SM backbone genes in *A. nidulans* (Figure 4), 18 out of 27 in *A.*
383 *fumigatus* (Figure S3), and seven out of eleven in *A. niger* (Figure S4) are found in sub-
384 telomeric regions. *A. oryzae* represents a unique case, with only two out of 18 SM backbone
385 genes with known structures located in sub-telomeric regions (Figure S5). The observation
386 that most characterized BGCs are localized in sub-telomeric regions suggests that they are
387 more frequently expressed than those in other regions.
388

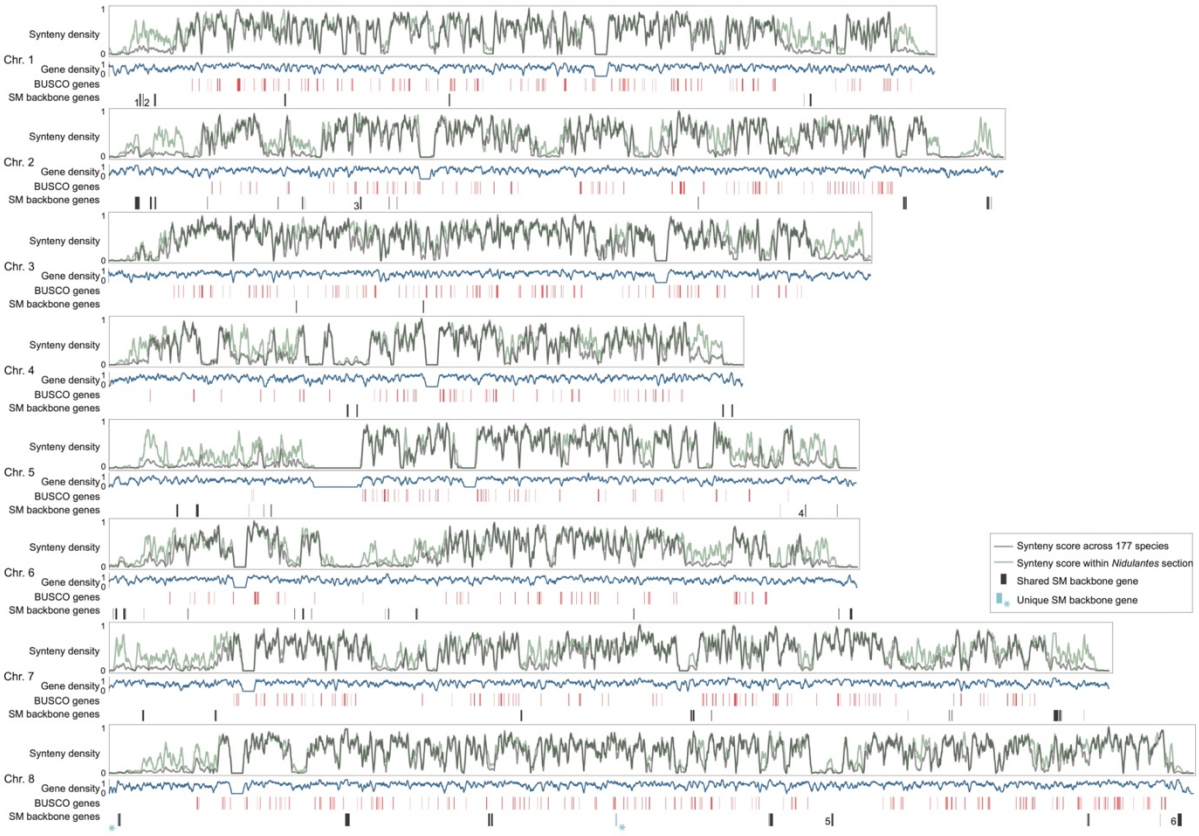
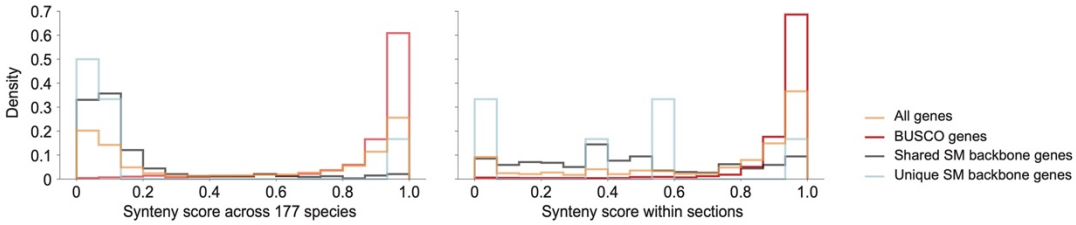


Figure 4 Chromosomal localization of secondary metabolite (SM) backbone genes in *Aspergillus nidulans*. This figure illustrates the genomic localization of BUSCO genes and SM backbone genes in *Aspergillus nidulans*. Lines in black and green represent synteny density across 177 species and within the *Nidulantes* section, respectively. The gene density is shown along the chromosomes. Red bars indicate the chromosomal locations of BUSCO genes, with bar width proportional to gene length. Black and blue bars mark the positions of shared and unique SM backbone genes, respectively, with their widths corresponding to gene length. Numbers adjacent to the black bars denote SM backbone genes associated with known metabolites as listed on the MIBiG website (Table S5).

A

Genes / species	*A. nidulans*	*A. niger*	*A. fumigatus*	*A. oryzae*
All genes	10,680	11,846	9,781	12,030
BUSCO genes	745	756	745	711
Shared SM backbone genes	81	113	56	89
Unique SM backbone genes	2	1	1	2

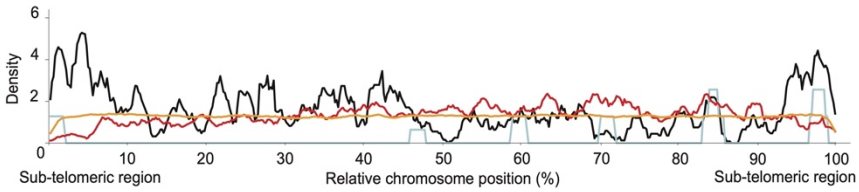
B



C

Gene groups	Genes with low genus-level synteny score	Genes with high genus-level synteny score	p value	Genes with low section-level synteny score	Genes with high section-level synteny score	p value
All genes	17,431	18,872	-	6,307	26,354	-
BUSCO genes	64	2,466	0	48	2,700	3.2e-117
SM backbone genes	279	14	6.1e-58	83	68	7.8e-28

D



E

Gene groups	Genes in sub-telomeric regions	Genes outside sub-telomeric regions	p value
All genes	9,139	35,186	-
BUSCO genes	207	2,746	4.3e-72
SM backbone genes	125	219	1.3e-12

Figure 5 Analysis of secondary metabolite (SM) backbone gene localization in low-synteny and sub-telomeric regions in four *Aspergillus* species. (A) Summary table of the different groups of genes considered in four *Aspergillus* species. (B) Distribution of different groups of genes in relation to synteny score analyzed across 177 species and within individual sections. (C) Summary table of the number of SM backbone genes, BUSCO genes, and all genes with high-synteny score (larger than 80%) and low-synteny score (lower than 20%) at the genus-level and section-level. (D) Distribution of different groups of genes relative to chromosomal localization, considering the terminal 10% of chromosomal regions as sub-telomeric. (E) Summary table of the number of SM backbone genes, BUSCO genes, and all genes within and outside of sub-telomeric regions. Chi-square statistic tests (Virtanen et al., 2020) were performed on SM backbone genes and BUSCO genes, determining whether their distributions differ significantly from the distribution of all genes in terms of genus-level synteny, section-level synteny, and sub-telomeric localization.

Exploring the impact of genomic localization and histone PTMs on SM gene expression variability in Aspergilli

It has been proposed that genomic localization of SM BGCs is important for the regulation of their expression through histone modifications as several histone PTMs were shown to

419 influence SM production in Aspergilli (Cánovas et al., 2014; Lee et al., 2009; Li et al., 2019).
420 Thus, we sought to explore the relationship between gene expression, histone PTMs, and
421 genomic and syntenic localization of SM backbone genes compared with BUSCO genes. To
422 this end, we utilized publicly available RNA-seq data from *A. nidulans*, *A. fumigatus*, *A. niger*,
423 and *A. oryzae*, and ChIP-seq (Chromatin ImmunoPrecipitation Sequencing) data from the first
424 two Aspergilli (Tables S7 and S8). We used the coefficient of variance (CV) to assess the
425 relative variability of gene expression and histone modification of different gene groups across
426 diverse experimental conditions. For BUSCO genes, we observed stable gene expression in
427 all species regardless of the synteny score (Figures 6A and S6), which is accompanied by
428 limited variation in all histone PTMs included in this study (H3K4me3, H3K9me3, H3K36me3,
429 and H3Kac) (Figures 6B-E, S6, S7). Compared to BUSCO genes, SM backbone genes exhibit
430 greater variability in gene expression (Figures 6A and S6). For histone PTMs, H3K36me3 in
431 *A. nidulans* and H3K4me3 in both *A. nidulans* and *A. fumigatus* show higher variability at SM
432 backbone genes compared with other gene groups (Figures 6B-C, S7 and S8), suggesting a
433 role for these histone modifications in regulating the expression of SM genes. Especially, SM
434 backbone genes with known structure and variable expression in *A. fumigatus* also exhibit
435 high CVs in H3K4me3, with four of them ranking among the top five most variable SM genes
436 (Table S5). However, the observed gene expression variation of SM genes is not explained
437 by H3K9me3 (in *A. nidulans* and *A. fumigatus*) nor H3ac (in *A. nidulans*) as SM genes do not
438 differ from BUSCO genes for these PTMs (Figures 6D-E, 6H-I, and S8). The finding that gene
439 expression seems independent from H3ac contrasts with the results obtained with *A. nidulans*
440 histone deacetylase mutants in which the expression of many BGCs is affected (Lee et al.,
441 2009). A possible explanation is that the inability to deacetylate histone tails prevents silencing
442 through histone methylation. It also suggests that specific histone methylations are a stronger
443 determinant of BGC expression than their acetylation. When comparing to other accessory
444 genes, as defined by genes located in regions with a synteny score lower than 0.2, SM genes
445 do not seem to differ in regard to RNA CVs and ChIP-seq CVs, indicating that the genomic
446 localization is particularly important for the regulation of the expression.

447

448 When taking the chromosome localization into account, SM backbone genes within the sub-
449 telomeric regions show slightly higher variation in gene expression in *A. nidulans*, *A. fumigatus*,
450 and *A. niger*, but not in *A. oryzae* (Figures 6A and S6). The difference in *A. oryzae* could be
451 due to the fact that it was domesticated for industrial purposes such as producing sake, soy
452 sauce, and miso, and it is generally recognized for its simplified metabolic profile and inability
453 to produce aflatoxins (Gibbons et al., 2012). A similar influence of domestication on the
454 downregulation of metabolite production was also observed in wild *Penicillium* molds on
455 cheese (Bodinaku et al., 2019). However, the sub-telomeric localization does not fully explain
456 a higher variation in both gene expression and histone PTMs because the distribution of these
457 variations for SM backbone genes outside sub-telomeric regions is bimodal and overlap with
458 sub-telomeric backbone genes (Figures 6A, 6C, S6, and S7). Lastly, SM backbone genes with
459 the known metabolite structures listed on the MIBiG website show various expression patterns
460 (Table S5). In *A. fumigatus*, known SM genes show high variability as their average CV (1.20)
461 is higher than the average CV for both all SM backbone genes (1.09) and SM backbone genes
462 outside sub-telomeric regions (0.94) (Table S5). Two SM backbone genes responsible for
463 sartorypyrone A, located in a low-synteny region at the genus level but not in a sub-telomeric
464 region (genes 25-27 in Figure S3), are in the top five most variably expressed SM backbone
465 genes. By contrast, in other Aspergilli, either no (in *A. nidulans* and *A. niger*) or only one (in *A.*
466 *oryzae*) SM backbone gene with known structure is in the top five most variable SM backbone
467 genes (Table S5). Based on our results, SM backbone genes that are still uncharacterized but
468 exhibit higher variation in expression deserve further attention as they are likely to produce
469 SMs under laboratory conditions. It must be noted that our conclusions rely on a very limited
470 set of available RNA-seq and ChIP-seq data. Gene expression and histone PTMs are highly
471 context-dependent, and the calculated CVs could be biased. We recommend that future
472 studies aim to incorporate more histone PTMs and expression data from diverse conditions to
473 mitigate these limitations and enhance the robustness of the conclusions drawn.

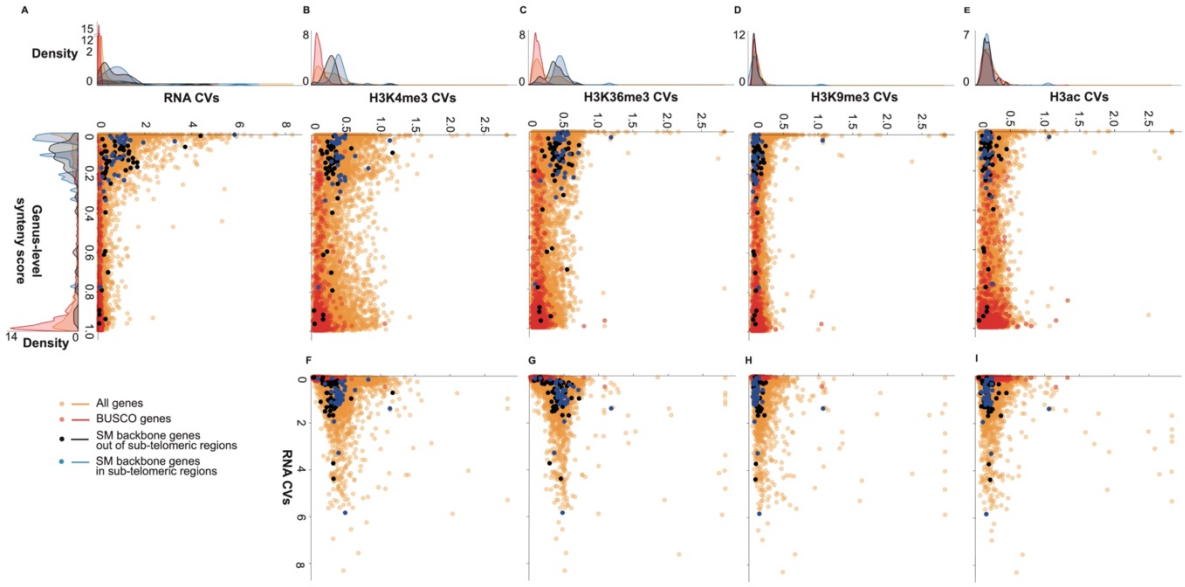


Figure 6 Correlation between genomic localization of secondary metabolite (SM) backbone genes and variation in gene expression and histone modifications in *Aspergillus nidulans*. We employed the coefficient of variance (CV) to assess the relative variability in gene expression (RNA) and histone post-translational modifications of all genes, BUSCO genes, and SM backbone genes in *Aspergillus nidulans*. (A-E) Correlation between CVs and genus-level synteny. (F-I) Correlation between histone modification CVs and gene expression CV. Me3: trimethylation; ac: acetylation.

Conclusions

Altogether, our results highlighted that SM backbone genes are quite abundant and diverse in the fungal genus *Aspergillus*, especially between different sections. Moreover, based on the four *Aspergillus* species with chromosomal-level assemblies and publicly available RNA-seq and ChIP-seq data, we showed that SM backbone genes exhibit a strong bias for low-synteny regions, and the association between their localization, variation in expression, and histone PTMs is restricted to H3K4me3 and H3K36me3 only. It is important to note that our study is limited to the few histone PTMs that are publicly available and other PTMs might play a more important role and need to be investigated. In particular, H3K79me is a relevant candidate as this modification was reported to localize at transition regions between heterochromatin and euchromatin, and we recently reported for this modification differences in abundances between different Aspergilli (Zhang et al., 2023). While our understanding of how genomic

496 localization and histone PTMs regulate SM production remains partial, our study points at new
497 research directions on understudied histone modifications that might prove key to unlock the
498 full potential of SM production in Aspergilli.
499
500 **Materials and methods**
501 **Acquisition of genomes and predicted proteomes for 177 fungal species**
502 The genomes and predicted proteomes of 177 fungal species were retrieved from the Joint
503 Genome Institute MycoCosm (https://mycocosm.jgi.doe.gov/mycocosm/home) repository
504 (Grigoriev et al., 2014) in January 2022. These species comprise 174 Aspergilli as well as
505 three *Penicillium* species (*Penicillium brevicompactum*, *Penicillium griseofulvum*, and
506 *Penicillium chrysogenum*) that were used as an outgroup in our analyses (Table S1).
507
508 **Species tree reconstruction**
509 To reconstruct the phylogenetic relationship between the here analyzed *Aspergillus* species,
510 we used the fungal database (Fungi Odb10, 758 orthologs) in BUSCO (Benchmarking
511 Universal Single-Copy Ortholog assessment tool) v4.0.1 (Simão et al., 2015; Zdobnov et al.,
512 2021) to retrieve fungal single-copy genes (Table S2) that are present in the 177 genomes.
513 For homologs of each single-copy BUSCO gene, we first generated multi-sequence alignment
514 using MAFFT v7.271 (parameter: --auto) (Katoh & Standley, 2013), then removed positions in
515 the alignments with gaps in more than 90% of the sequences using '-gt 0.1' in TrimAl v1.2
516 (Capella-Gutiérrez et al., 2009) for better tree construction (Steenwyk et al., 2020). We
517 concatenated the 758 alignments into one 'super-alignment' and employed IQ-TREE v1.6.10
518 (Chernomor et al., 2016; Nguyen et al., 2015) to construct the maximum-likelihood phylogeny
519 with partitioned model selection to allow selecting a substitution model (Table S2) for each
520 BUSCO gene alignment, and used both ultrafast bootstrap approximations and SH-aLRT
521 (Shimodaira–Hasegawa approximate Likelihood Ratio Test) with 1,000 pseudo-replicates to
522 calculate branch supports (Hoang et al., 2017).
523

524 **Identification and conservation analysis of putative SM backbone genes and**
525 **orthologous groups**
526 We used antiSMASH (antibiotics and Secondary Metabolite Analysis SHell) v6.1.0 (Blin et al.,
527 2021) (parameters: –minimal) to identify all regions with predicted BGCs in the 177 genome
528 assemblies, and refined the classification of SM backbone genes with our in-house tool
529 BGCtoolkit (https://github.com/WesterdijkInstitute/BGClib) based on conserved domain
530 organization (Figure 1 and Table S3). Then, we extracted the protein sequence of the
531 predicted backbone enzymes or of the adenylation domains for NRPS, NRPS-like, and PKS-
532 NRPS hybrid enzymes using BGCtoolkit. For each type of backbone enzyme, protein
533 sequences were aligned and trimmed using MAFFT and TrimAl, respectively, as mentioned
534 above. We constructed phylogenetic trees using FastTree v2.1.11 with default parameters
535 (Price et al., 2009). We subsequently used Newick Utilities v1.6.0 (Junier & Zdobnov, 2010)
536 to identify orthologs *via* three steps: i) re-root the tree based on the mid-point (nw_reroot); ii)
537 manually select the clades having more than 88 leaves and with trustworthy bootstrap value
538 (>= 0.99) and mark them as orthologs; iii) for the remaining leaves, automatically split the
539 clades with fewer than 88 leaves and with trustworthy bootstrap value (>= 0.99) (nw_ed -n -r).
540 In the construction of the three adenylation domain trees, an extra step was undertaken to
541 consolidate multiple domains originating from the same gene. This involved examining all
542 leaves in the domain tree to identify the genes associated with these domains. Clades were
543 collapsed and assigned a new clade name when the domains were found to originate from
544 the same gene. For any remaining domains, the original clade names were retained. We built
545 a saturation curve with the matplotlib.pyplot visualization library (Hunter, 2007) and used the
546 function from the SciPy library (Virtanen et al., 2020) to implement a curve-fitting process,
547 which determines the rate of growth (β) of unique gene numbers under Heap's Law (Heaps,
548 1978).
549

550 We used OrthoFinder v2.3.8 (Emms & Kelly, 2019) to identify orthologous groups across 177
551 predicted proteomes using default settings (Table S4). A saturation curve based on these
552 orthologous groups was built as described above.
553
554 **Genome synteny analysis**
555 We employed MCScanX (Multiple Collinearity Scan Toolkit X version) (Wang et al., 2012) to
556 conduct an all-versus-all pairwise comparison among all genomes to identify syntenic blocks
557 (match size: -s 3 and max gaps: -m 15). We defined the 'synteny percentage' between a pair
558 of species as the ratio of syntenic genes detected between two compared species to the total
559 genes. We defined the 'between sections' synteny percentage by considering all pairs of
560 species that are associated with different taxonomic sections, and the 'within sections' synteny
561 percentage by considering all pairs from the same taxonomic section. For the synteny
562 percentage of each taxonomic section, we considered all comparisons for species within that
563 specific section. For visual representation of genome alignment between two species, we used
564 the R package GENESPACE v1.2.3 (Lovell et al., 2022).
565
566 To calculate the average length of syntenic regions, we took all syntenic blocks identified by
567 the MCScanX in all-versus-all pairwise comparisons and excluded the contigs with less than
568 50 genes. Subsequently, we employed the hypergeometric distribution (Rice, 2003) utilizing
569 the 'hypergeom.cdf(x, M, n, N)' function from the SciPy library (Virtanen et al., 2020) to
570 evaluate the likelihood of successfully detecting a syntenic region between two given species.
571 In this function, 'M' denotes the total number of homologous gene pairs identified when
572 comparing the complete genomes of species A and B, 'n' represents the count of genes on
573 scaffold B1 with homologs in species B, and 'N' is the number of genes on scaffold A1 with
574 homologs in species A. This cumulative distribution function aggregates the probabilities of
575 identifying up to 'x' successful syntenic matches, thereby yielding the cumulative probability
576 of detecting 'x' or fewer successes. The analysis considered (1) the total number of homologs
577 between the two species, (2) the number of homologs on each contig relative to the other

578 species, and (3) the number of homologs between the contigs themselves. Setting a p-value
579 threshold at 1e-5, we then computed the average length of syntenic blocks for each pair of
580 species.
581
582 We defined the synteny score at the genus level as the synteny score for a specific gene by
583 dividing the number of species containing that gene by the total number of species (177
584 species). At the section level, the synteny score for a gene is determined by the ratio of the
585 number of species within the section harboring that gene to the total species count in that
586 section. We used the function 'chi2_contingency' from the SciPy library (Virtanen et al., 2020)
587 to perform chi-square statistic tests on SM backbone genes and BUSCO genes, determining
588 whether their distributions differ significantly from the distribution of all genes in terms of
589 genus-level synteny, section-level synteny, and sub-telomeric localization (Table S6).
590
591 **RNA-seq sample preparation**
592 To perform RNA sequencing, we obtained the fine powder of young mycelium of *A. fumigatus*
593 (strain: Af293, CBS 126847; grown on malt extract agar), *A. nidulans* (strain: Wtpaba
594 (genotype: pabaA1); grown on yeast agar glucose), and *A. niger* (strain: NRRL3, CBS 120.49;
595 grown on malt extract agar) as described previously (Zhang et al., 2023). Subsequently, we
596 extracted high-quality total RNA from each species in four replicates, following our in-house
597 RNA extraction protocol. Briefly, we placed around 100 mg of fine powder in a 1.5 mL tube,
598 added 1 mL TRIzol reagent (Fisher, 15596026), and incubated for 5 mins at 25°C. Following
599 that, we added 0.2 mL chloroform-d (Sigma, 416754), mixed gently by hand 15 times, and
600 incubated for 5 mins at room temperature. After centrifuging the samples at 10,000 RPM for
601 15 mins, we transferred the aqueous supernatant to a new tube. We then added 0.5 volume
602 of 100% ethanol (BOOM, 200-578-6) and proceeded from the 5th step of the NucleoSpin RNA
603 Kit protocol (Bioke, 740955). Finally, we eliminated DNA contamination using DNase I (Sigma,
604 AMPD1) and sent samples to the Beijing Genome Institute (BGI, Hong Kong, China) for library
605 construction and DNBseq (DNA NanoBall sequencing).

606

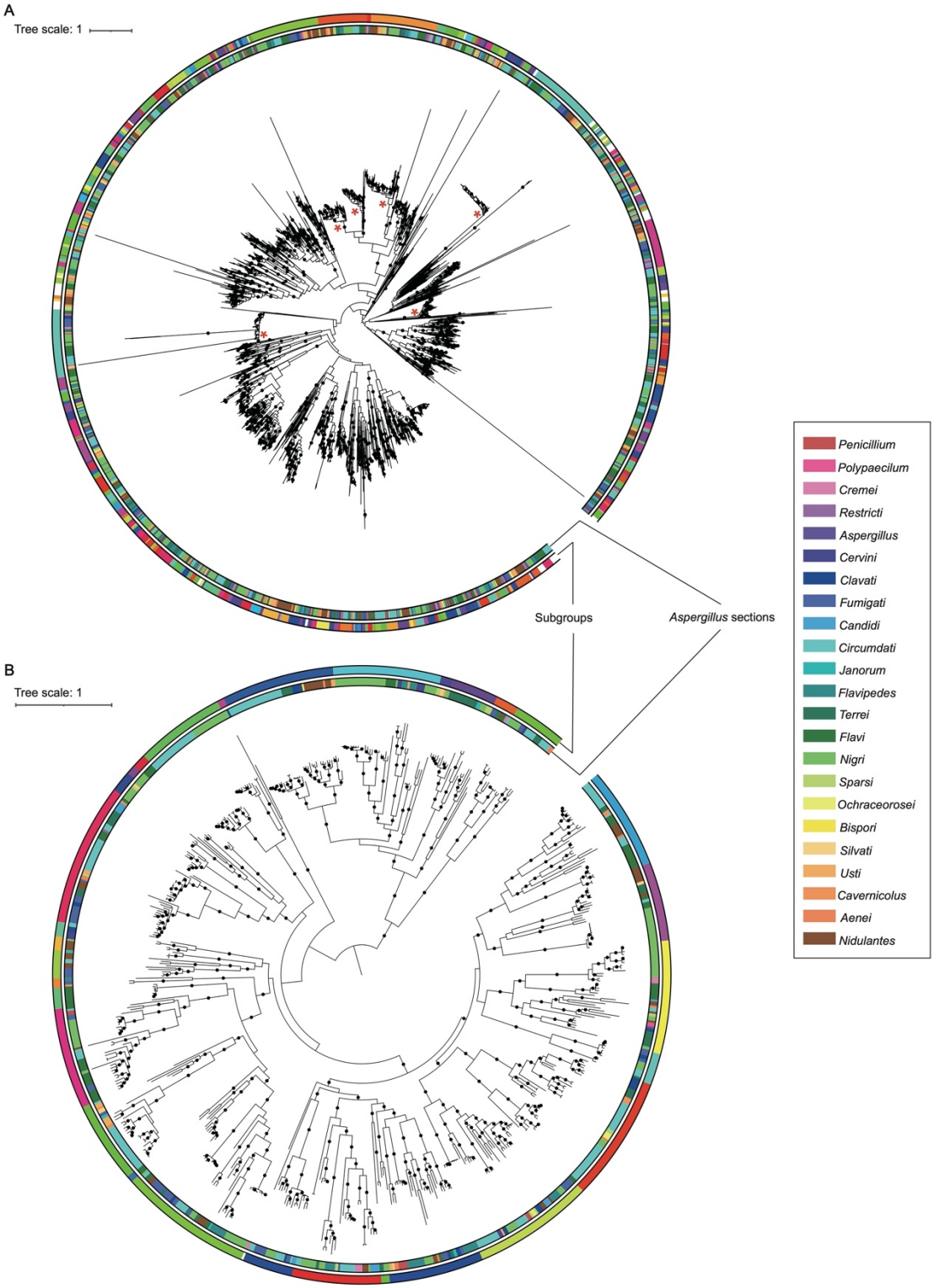
607 **Acquisition and analysis of public RNA-seq and ChIP-seq datasets**

608 To retrieve publicly available RNA-seq data (Table S7) from *A. niger*, *A. fumigatus*, *A. nidulans*,
609 and *A. oryzae*, we utilized prefetch v2.10.0 and fasterq-dump v2.10.0 from SRA Toolkit
610 (https://github.com/ncbi/sra-tools) to extract FASTQ files from the SRA database
611 (https://www.ncbi.nlm.nih.gov/sra). For all reads from both our BGI sequencing projects and
612 the public database, we used the pseudoalignment tool kallisto v0.46.0 (Bray et al., 2016) to
613 build the index and quantify raw read counts per gene. To remove the potential batch effect,
614 we used ComBat-seq (Zhang et al., 2020) in R v3.6.1 (Team, 2010) and obtained batch-effect
615 corrected read counts. Following this, we calculated the TPM (Transcripts Per Million) and
616 used log-transformed TPMs to calculate the coefficient of variation (CV) to assess the
617 variability of gene expression levels across different samples. We used seaborn (Waskom,
618 2021) and Matplotlib (Hunter, 2007) to plot the scatter plot and density plots.

619

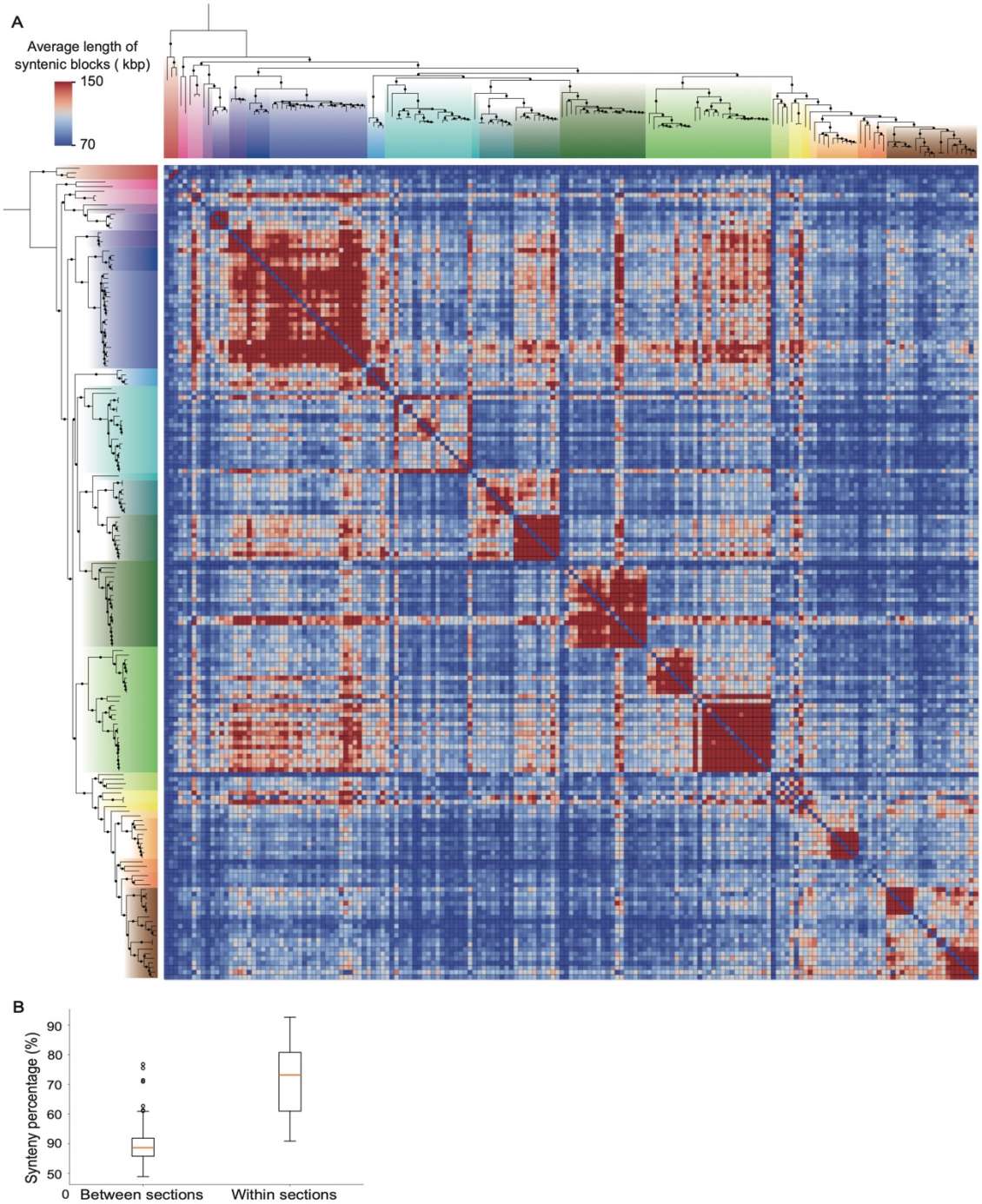
620 We utilized ChIP-seq data from two studies (Colabardini et al., 2022; Gacek-Matthews et al.,
621 2016), which included *A. nidulans* (H3ac, H3K4me3, H3K9me3, and H3K36me3) and *A.*
622 *fumigatus* (H3K4me3 and H3K9me3) (Table S8). The analyses (downloading and processing)
623 of this data were the same as the RNA-seq analysis mentioned above, yet we did not perform
624 the batch effect removal step because the two studies were conducted in different species.

625 **Supplementary materials**

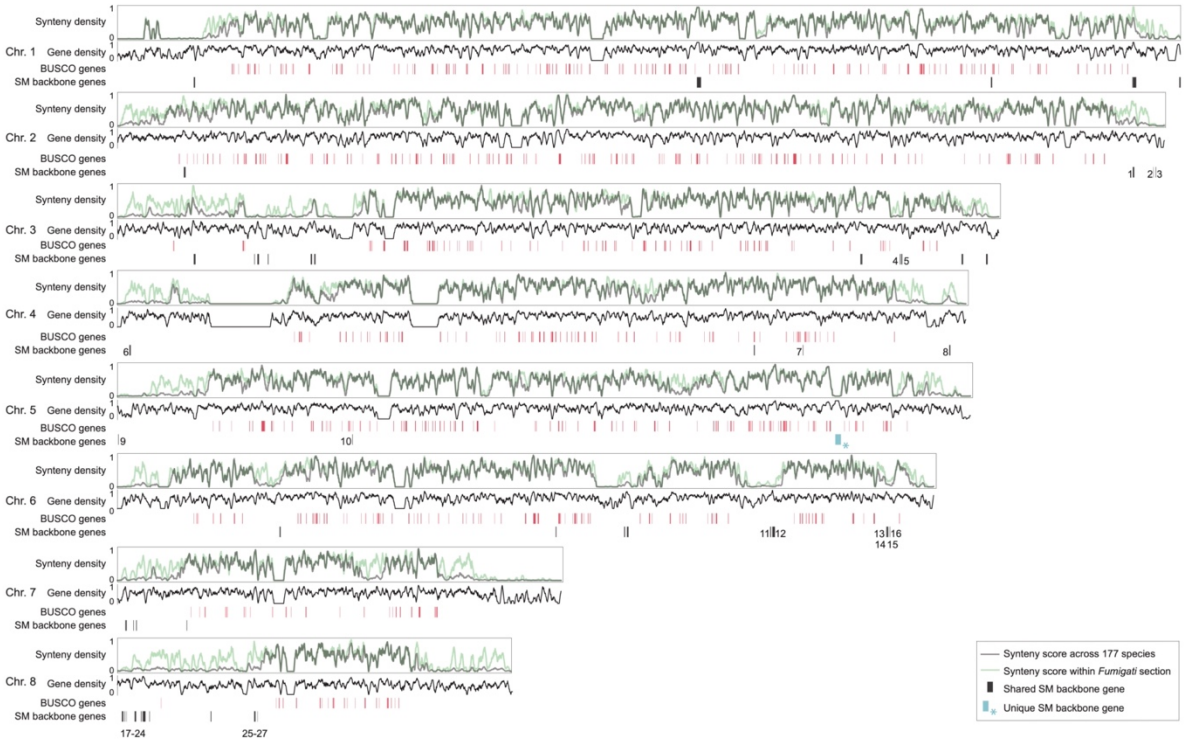


626
627 **Supplementary Figure 1 The phylogeny-aware method to detect secondary metabolite (SM)**
628 **backbone gene diversity.** We used FastTree (Price et al., 2009) to construct the phylogeny of
629 each type of SM backbone genes and annotated the trees with iTOL. The inner colored ring
630 indicates *Aspergillus* sections and the outer colored ring showed the splitting results by Newick
631 Utilities (Junier & Zdobnov, 2010). (A) Manual selection is implemented first to pick up the clades

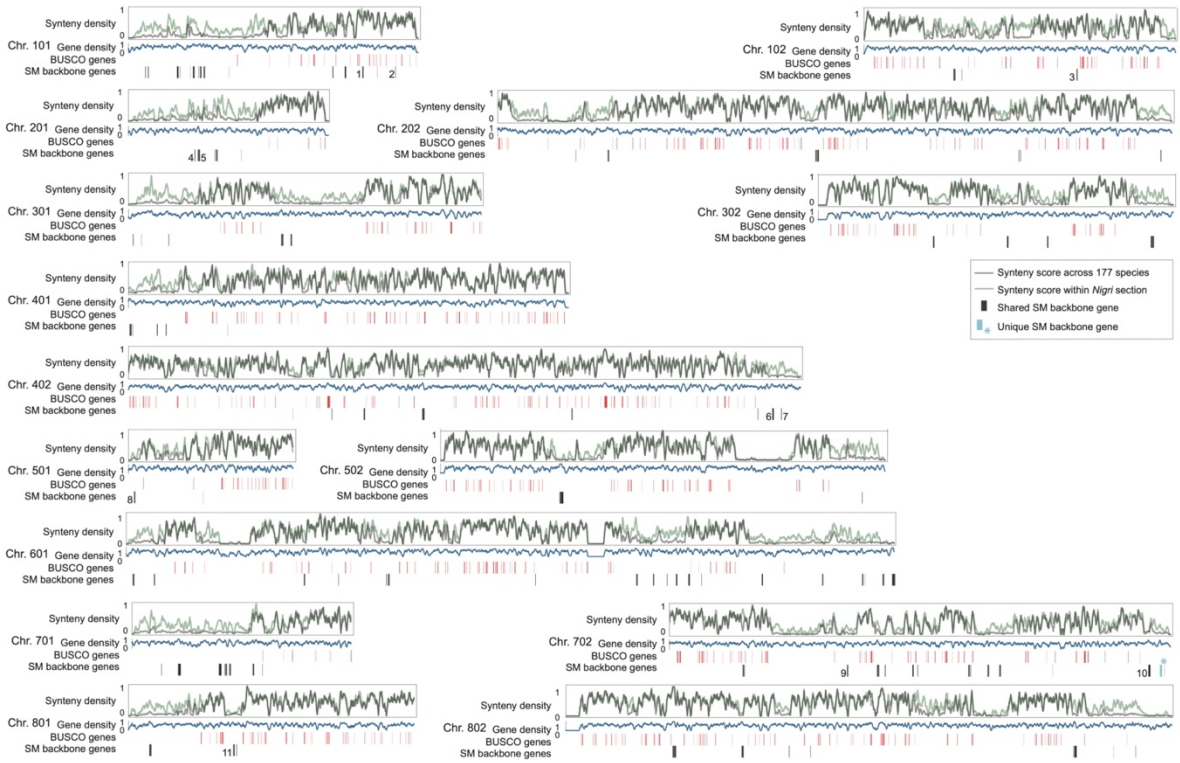
632 larger than 88 leaves (red "*" in the tree). For remaining leaves and trees that do not need manual
633 trimming (B), automatic splitting was performed with a trustworthy bootstrap value of >= 0.99.
634



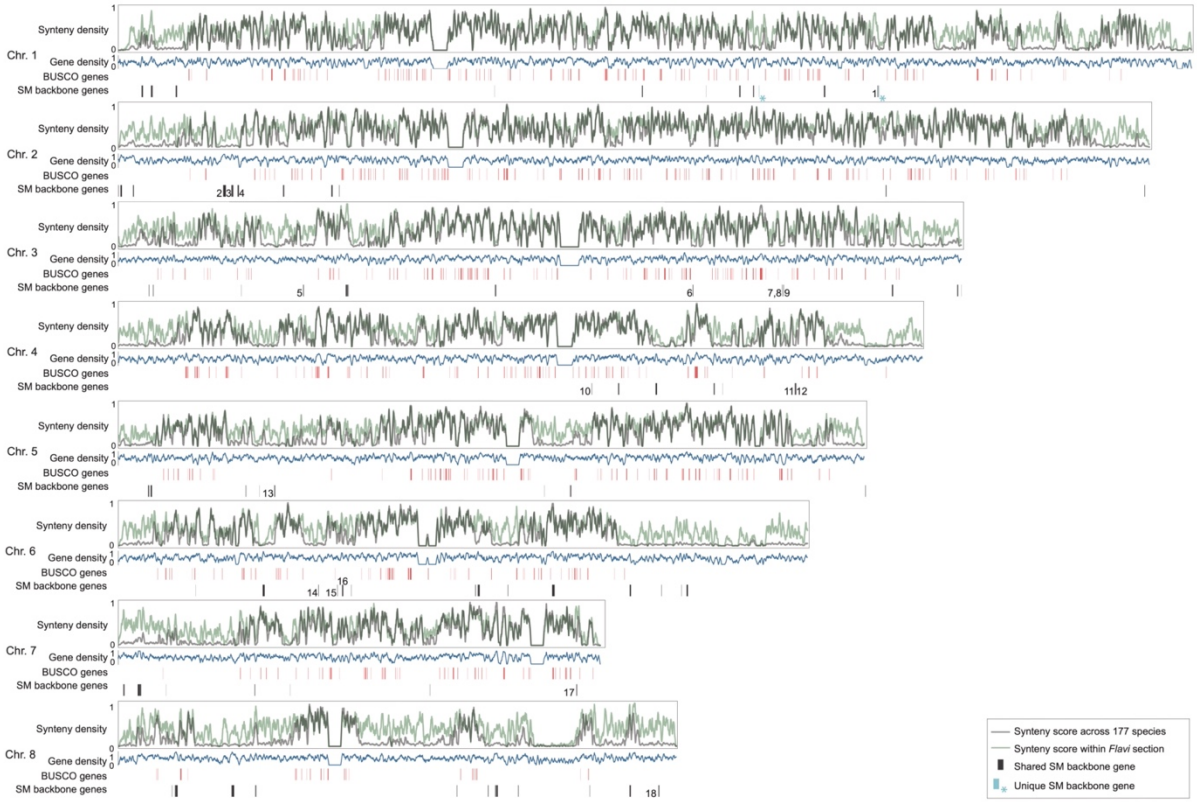
635
636 **Supplementary Figure 2 Length of syntenic blocks across *Aspergillus* sections.** (A) The
637 heatmap displays the all-versus-all pairwise comparisons based on the average length of syntenic
638 blocks across all species; species were arranged according to the phylogenetic tree. The
639 accompanying species tree, displayed to the left and top, uses colors to differentiate between
640 sections. (B) We used the "highly contiguous" genome assemblies defined by (Liu et al., 2018) to
641 calculate the synteny percentage between or within sections and obtained similar results as Figure
642 3C.
643



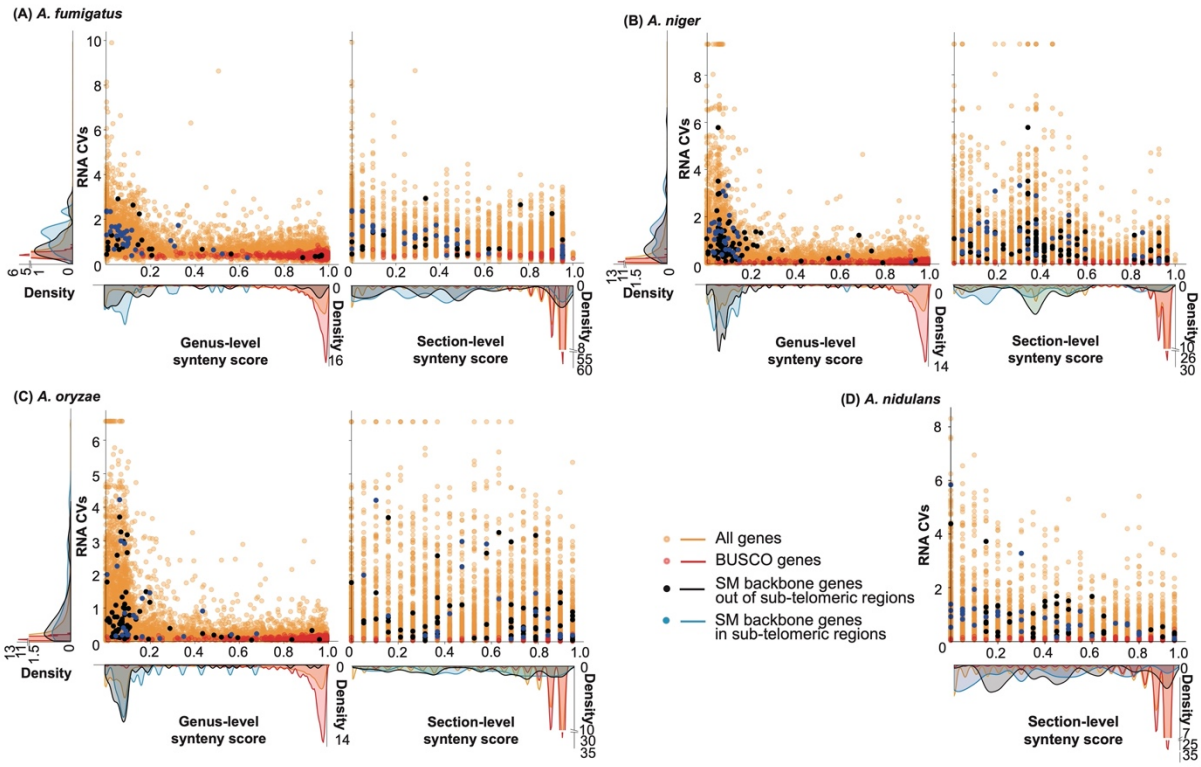
Supplementary Figure 3 Chromosomal localization of secondary metabolite (SM) backbone genes in *Aspergillus fumigatus*. This figure illustrates the genomic localization of BUSCO genes and SM backbone genes in *Aspergillus fumigatus*. Lines in black and green represent synteny density across 177 species and within the *Fumigati* section, respectively. The gene density is shown along the chromosomes. Red bars indicate the chromosomal locations of BUSCO genes, with bar width proportional to gene length. Black and blue bars mark the positions of shared and unique SM backbone genes, respectively, with their widths corresponding to gene length. Numbers adjacent to the black bars denote SM backbone genes associated with known metabolites as listed on the MIBiG database (Table S5).



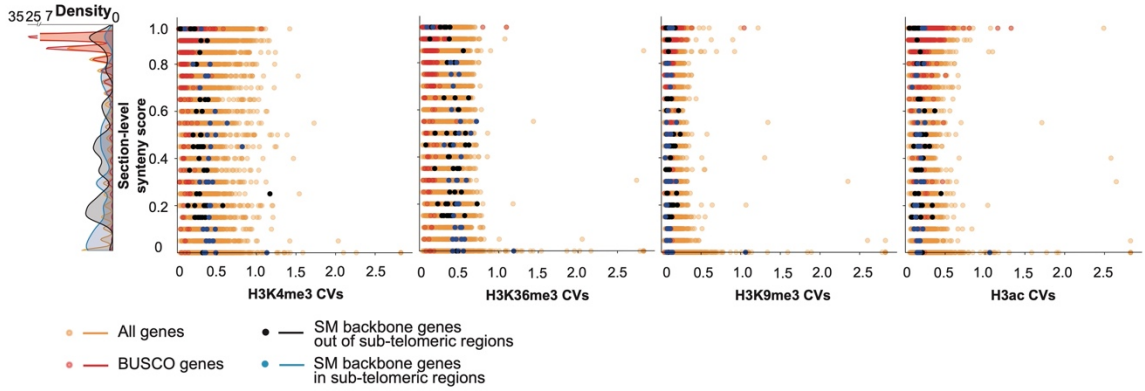
Supplementary Figure 4 Chromosomal localization of secondary metabolite (SM) backbone genes in *Aspergillus niger*. This figure illustrates the genomic localization of BUSCO genes and SM backbone genes in *Aspergillus niger*. Lines in black and green represent synteny density across 177 species and within the *Nigri* section, respectively. The gene density is shown along the chromosomes. Red bars indicate the chromosomal locations of BUSCO genes, with bar width proportional to gene length. Black and blue bars mark the positions of shared and unique SM backbone genes, respectively, with their widths corresponding to gene length. Numbers adjacent to the black bars denote SM backbone genes associated with known metabolites as listed on the MIBiG website (Table S5).



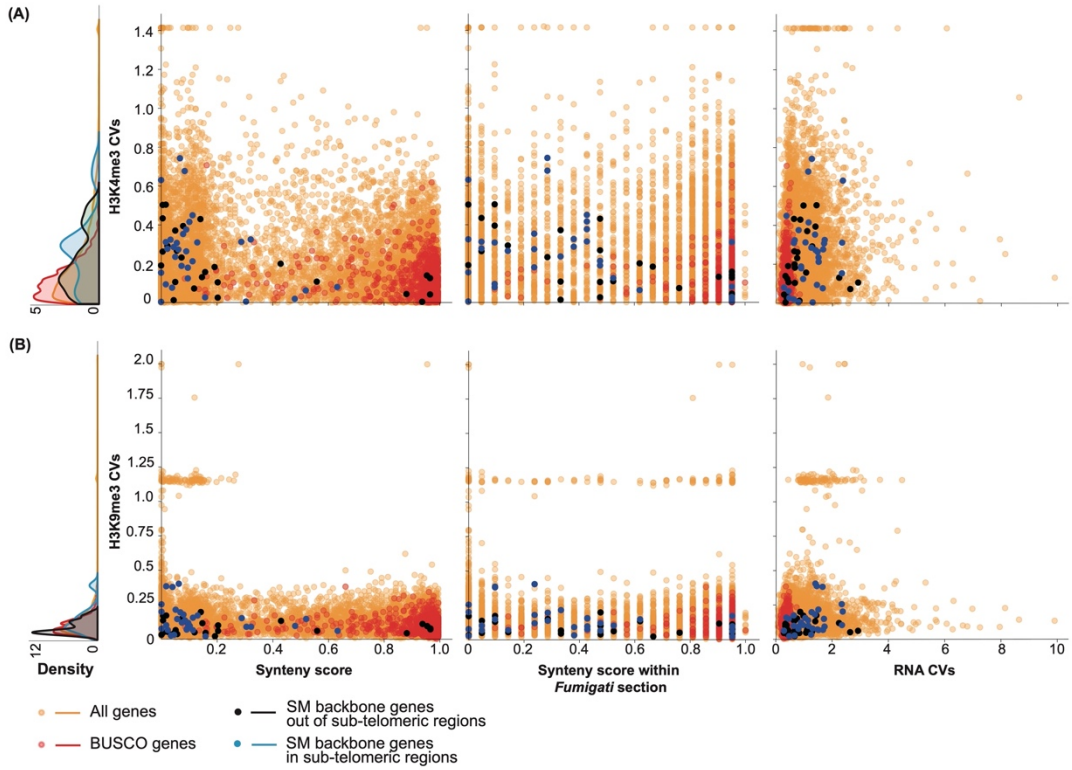
Supplementary Figure 5 Chromosomal localization of secondary metabolite (SM) backbone genes in *Aspergillus oryzae*. This figure illustrates the genomic localization of BUSCO genes and SM backbone genes in *Aspergillus oryzae*. Lines in black and green represent synteny density across 177 species and within the *Flavi* section, respectively. The gene density is shown along the chromosomes. Red bars indicate the chromosomal locations of BUSCO genes, with bar width proportional to gene length. Black and blue bars mark the positions of shared and unique SM backbone genes, respectively, with their widths corresponding to gene length. Numbers adjacent to the black bars denote SM backbone genes associated with known metabolites as listed on the MIBiG website (Table S5).



Supplementary Figure 6 Correlation between genomic localization of secondary metabolite (SM) backbone genes and variation in gene expression in four Aspergilli. We employed the coefficient of variance (CV) to assess the relative variability of all genes, BUSCO genes, and SM backbone genes within (A) *A fumigatus*, (B) *A. niger*, (C) *A. oryzae*, and (D) *A. nidulans*. The variability is depicted using distinct colors for dots and density lines to differentiate between the gene groups. RNA-seq data was downloaded from SRA (Table S7) and ChIP-seq data is from the publication (Gacek-Matthews et al., 2016).



Supplementary Figure 7 Correlation between variation in histone post-translation modifications and section-level synteny score in *Aspergillus nidulans*. We employed the coefficient of variance (CV) to assess the relative variability of all genes, BUSCO genes, and SM backbone genes in *A. nidulans*. The variability is depicted using distinct colors for dots and density lines to differentiate between the gene groups. ChIP-seq data is from the publication (Gacek-Matthews et al., 2016).



Supplementary Figure 8 Correlation between genomic localization of secondary metabolite (SM) backbone genes and variation in gene expression and histone modifications in *Aspergillus fumigatus*. We employed the coefficient of variance (CV) of H3K4me3 and H3K9me3 to assess the relative variability of all genes, BUSCO genes, and SM backbone genes. Data is from publication (Colabardini et al., 2022). The variability is depicted using distinct colors for dots and density lines to differentiate between the gene groups.

Data Summary

The genomes used in this study are publicly available at the JGI (Joint Genome Institute) MycoCosm repository (Grigoriev et al., 2014); the species names and abbreviations are listed in Table S1. To evaluate the completeness of the predicted proteomes and to obtain a species phylogeny, 758 fungal BUSCO (Benchmarking Universal Single-Copy Ortholog) genes were used, and their names are listed in Table S2. RNA-seq data is deposited in the SRA with the dataset identifier (RJNA1078405). Table S4 (Distribution of Orthologous groups) is deposited at Zenodo (DOI: 10.5281/zenodo.10638673).

Acknowledgment

The sequence data were produced by the US Department of Energy Joint Genome Institute (http://www.jgi.doe.gov/) in collaboration with the user community. Especially, we thank Scott Baker, Mikael Andersen, and Joseph Spatafora for their permission to use the sequence data. We thank Jos Houbraken for validating the *Aspergillus* species tree.

Funding information

Xin Zhang is funded by the Chinese Scholarship Council (CSC) (201907720028).

Authors and contribution

X.Z. formal analysis, investigation, methodology, visualization, writing original draft; I.L. formal analysis, investigation, methodology, visualization; J.C. conceptualization, funding acquisition, methodology, project administrations, supervision, visualization, writing original draft; M.F.S. conceptualization, funding acquisition, methodology, project administrations, supervision, visualization, writing original draft.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Blin, K., Shaw, S., Kloosterman, A. M., Charlop-Powers, Z., Van Wezel, G. P., Medema, M. H. & Weber, T. (2021). AntiSMASH 6.0: Improving cluster detection and comparison capabilities. *Nucleic Acids Research*, *49*(W1), W29–W35.

Bodinaku, I., Shaffer, J., Connors, A. B., Steenwyk, J. L., Biango-Daniels, M. N., Kastman, E. K., Rokas, A., Robbat, A. & Wolfe, B. E. (2019). Rapid phenotypic and metabolomic domestication of wild *Penicillium* molds on cheese. *MBio*, *10*(5), 10–1128.

Brakhage, A. A. (2013). Regulation of fungal secondary metabolism. *Nature Reviews Microbiology*, *11*(1), 21–32.

Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. *Nature Biotechnology*, *34*(5), 525–527.

Cairns, T. C., Barthel, L. & Meyer, V. (2021). Something old, something new: challenges and developments in *Aspergillus niger* biotechnology. *Essays in Biochemistry*, *65*(2), 213–224.

Cánovas, D., Marcos, A. T., Gacek, A., Ramos, M. S., Gutiérrez, G., Reyes-Domínguez, Y. & Strauss, J. (2014). The histone acetyltransferase GcnE (GCN5) plays a central role in the regulation of *Aspergillus* asexual development. *Genetics*, *197*(4), 1175–1189.

Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, *25*(15), 1972–1973.

Chernomor, O., Von Haeseler, A. & Minh, B. Q. (2016). Terrace aware data structure for phylogenomic inference from supermatrices. *Systematic Biology*, *65*(6), 997–1008.

Chiang, Y.-M., Szewczyk, E., Davidson, A. D., Keller, N., Oakley, B. R. & Wang, C. C. C. (2009). A gene cluster containing two fungal polyketide synthases encodes the biosynthetic pathway for a polyketide, asperfuranone, in *Aspergillus nidulans*. *Journal of the American Chemical Society*, *131*(8), 2965–2970.

Colabardini, A. C., Wang, F., Miao, Z., Pardeshi, L., Valero, C., de Castro, P. A., Akiyama, D. Y., Tan, K., Nora, L. C., Silva-Rocha, R., Marcet-Houben, M., Gabaldón, T., Fill, T., Wong, K. H. & Goldman, G. H. (2022). Chromatin profiling reveals heterogeneity in clinical isolates of the human pathogen *Aspergillus fumigatus*. In *PLOS Genetics* (Vol. 18).

Collemare, J. & Seidl, M. F. (2019). Chromatin-dependent regulation of secondary metabolite biosynthesis in fungi: is the picture complete? *FEMS Microbiology Reviews*, *43*(6), 591–607.

Connolly, L. R., Smith, K. M. & Freitag, M. (2013). The *Fusarium graminearum* histone H3 K27 methyltransferase KMT6 regulates development and expression of secondary metabolite gene clusters. *PLoS Genetics*, *9*(10), e1003916.

Drillon, G., Carbone, A. & Fischer, G. (2013). Combinatorics of chromosomal rearrangements based on synteny blocks and synteny packs. *Journal of Logic and Computation*, *23*(4), 815–838.

Drott, M. T., Bastos, R. W., Rokas, A., Ries, L. N. A., Gabaldón, T., Goldman, G. H., Keller, N. P. & Greco, C. (2020). Diversity of secondary metabolism in *Aspergillus nidulans* clinical isolates. *Msphere*, *5*(2), e00156-20.

Drott, M. T., Rush, T. A., Satterlee, T. R., Giannone, R. J., Abraham, P. E., Greco, C., Venkatesh, N., Skerker, J. M., Glass, N. L. & Labbé, J. L. (2021). Microevolution in the pansecondary metabolome of *Aspergillus flavus* and its potential macroevolutionary implications for filamentous fungi. *Proceedings of the National Academy of Sciences*, *118*(21), e2021683118.

Emms, D. M. & Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biology*, *20*(1), 1–14.

Gacek-Matthews, A., Berger, H., Sasaki, T., Wittstein, K., Gruber, C., Lewis, Z. A. & Strauss, J. (2016). KdmB, a Jumonji histone H3 demethylase, regulates genome-wide H3K4 trimethylation and is required for normal induction of secondary metabolism in *Aspergillus nidulans*. *PLoS Genetics*, *12*(8), 1–29.

Gacek, A. & Strauss, J. (2012). The chromatin code of fungal secondary metabolite gene clusters. *Applied Microbiology and Biotechnology*, *95*(6), 1389–1404.

Gibbons, J. G., Salichos, L., Slot, J. C., Rinker, D. C., McGary, K. L., King, J. G., Klich, M. A., Tabb, D. L., McDonald, W. H. & Rokas, A. (2012). The evolutionary imprint of domestication on genome variation and function of the filamentous fungus *Aspergillus oryzae*. *Current Biology*, *22*(15), 1403–1409.

Grigoriev, I. V, Nikitin, R., Haridas, S., Kuo, A., Ohm, R., Otillar, R., Riley, R., Salamov, A., Zhao, X. & Korzeniewski, F. (2014). MycoCosm portal: gearing up for 1000 fungal genomes. *Nucleic Acids Research*, *42*(D1), D699–D704.

Guzmán-Chávez, F., Zwahlen, R. D., Bovenberg, R. A. L. & Driessen, A. J. M. (2018). Engineering of the filamentous fungus *Penicillium chrysogenum* as cell factory for natural products. *Frontiers in*

Microbiology, *9*, 2768.
Hane, J. K., Rouxel, T., Howlett, B. J., Kema, G. H. J., Goodwin, S. B. & Oliver, R. P. (2011). A novel mode of chromosomal evolution peculiar to filamentous Ascomycete fungi. *Genome Biology*, *12*(5).
Heaps, H. S. (1978). Information retrieval: Computational and theoretical aspects. *Academic Press*.
Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q. & Vinh, L. S. (2017). UFBoot2: Improving the ultrafast bootstrap approximation. *BioRxiv*, *35*(2), 518–522.
Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, *9*(03), 90–95.
Inglis, D. O., Binkley, J., Skrzypek, M. S., Arnaud, M. B., Cerqueira, G. C., Shah, P., Wymore, F., Wortman, J. R. & Sherlock, G. (2013). Comprehensive annotation of secondary metabolite biosynthetic genes and gene clusters of *Aspergillus nidulans*, *A. fumigatus*, *A. niger* and *A. oryzae*. *BMC Microbiology*, *13*, 1–23.
Junier, T. & Zdobnov, E. M. (2010). The Newick utilities: high-throughput phylogenetic tree processing in the UNIX shell. *Bioinformatics*, *26*(13), 1669–1670.
Katoh, K. & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution*, *30*(4), 772–780.
Keller, N. P. (2019). Fungal secondary metabolism: regulation, function and drug discovery. *Nature Reviews Microbiology*, *17*(3), 167–180.
Kessler, S. C. & Chooi, Y.-H. (2022). Out for a RiPP: challenges and advances in genome mining of ribosomal peptides from fungi. *Natural Product Reports*, *39*(2), 222–230.
Kjærbølling, I., Vesth, T. C., Frisvad, J. C., Nybo, J. L., Theobald, S., Kuo, A., Bowyer, P., Matsuda, Y., Mondo, S., Lyhne, E. K., Kogle, M. E., Clum, A., Lipzen, A., Salamov, A., Ngan, C. Y., Daum, C., Chiniquy, J., Barry, K., LaButti, K., … Andersen, M. R. (2018). Linking secondary metabolites to gene clusters through genome sequencing of six diverse *Aspergillus* species. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(4), E753–E761.
Kjærbølling, I., Vesth, T., Frisvad, J. C., Nybo, J. L., Theobald, S., Kildgaard, S., Petersen, T. I., Kuo, A., Sato, A., Lyhne, E. K., Kogle, M. E., Wiebenga, A., Kun, R. S., Lubbers, R. J. M., Mäkelä, M. R., Barry, K., Chovatia, M., Clum, A., Daum, C., … Andersen, M. R. (2020). A comparative genomics study of 23 *Aspergillus* species from section *Flavi*. *Nature Communications*, *11*(1).
Klejnstrup, M. L., Frandsen, R. J. N., Holm, D. K., Nielsen, M. T., Mortensen, U. H., Larsen, T. O. & Nielsen, J. B. (2012). Genetics of polyketide metabolism in *Aspergillus nidulans*. *Metabolites*, *2*(1), 100–133.
Lawrence, M., Daujat, S. & Schneider, R. (2016). Lateral Thinking: How Histone Modifications Regulate Gene Expression. *Trends in Genetics*, *32*(1), 42–56.
Lee, I., Oh, J. H., Keats Shwab, E., Dagenais, T. R. T., Andes, D. & Keller, N. P. (2009). HdaA, a class 2 histone deacetylase of *Aspergillus fumigatus*, affects germination and secondary metabolite production. *Fungal Genetics and Biology*, *46*(10), 782–790.
Li, X., Pan, L., Wang, B. & Pan, L. (2019). The histone deacetylases HosA and HdaA affect the phenotype and transcriptomic and metabolic profiles of *Aspergillus niger*. *Toxins*, *11*(9).
Lind, A. L., Wisecaver, J. H., Lameiras, C., Wiemann, P., Palmer, J. M., Keller, N. P., Rodrigues, F., Goldman, G. H. & Rokas, A. (2017). Drivers of genetic diversity in secondary metabolic gene clusters within a fungal species. *PLoS Biology*, *15*(11), 1–26.
Liu, D., Hunt, M. & Tsai, I. J. (2018). Inferring synteny between genome assemblies: a systematic evaluation. *BMC Bioinformatics*, *19*, 1–13.
Lovell, J. T., Sreedasyam, A., Schranz, M. E., Wilson, M., Carlson, J. W., Harkess, A., Emms, D., Goodstein, D. M. & Schmutz, J. (2022). GENESPACE tracks regions of interest and gene copy number variation across multiple genomes. *Elife*, *11*, e78526.
Macheleidt, J., Mattern, D. J., Fischer, J., Netzker, T., Weber, J., Schroeckh, V., Valiante, V. & Brakhage, A. A. (2016). Regulation and role of fungal secondary metabolites. *Annual Review of Genetics*, *50*, 371–392.
Mosunova, O., Navarro-Muñoz, J. C. & Collemare, J. (2021). The biosynthesis of fungal secondary metabolites: From fundamentals to biotechnological applications. In *Encyclopedia of mycology* (pp. 458–476). Elsevier.
Nakao, Y., Kanamori, T., Itoh, T., Kodama, Y., Rainieri, S., Nakamura, N., Shimonaga, T., Hattori, M. & Ashikari, T. (2009). Genome sequence of the lager brewing yeast, an interspecies hybrid. *DNA Research*, *16*(2), 115–129.
Newman, D. J. & Cragg, G. M. (2016). Natural products as sources of new drugs from 1981 to 2014. *Journal of Natural Products*, *79*(3), 629–661.
Nguyen, L.-T., Schmidt, H. A., Von Haeseler, A. & Minh, B. Q. (2015). IQ-TREE: a fast and effective

stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, *32*(1), 268–274.

Nielsen, K. F., Mogensen, J. M., Johansen, M., Larsen, T. O. & Frisvad, J. C. (2009). Review of secondary metabolites and mycotoxins from the *Aspergillus niger* group. *Analytical and Bioanalytical Chemistry*, *395*, 1225–1242.

Nierman, W. C., Pain, A., Anderson, M. J., Wortman, J. R., Kim, H. S., Arroyo, J., Berriman, M., Abe, K., Archer, D. B., Bermejo, C., Bennett, J., Bowyer, P., Chen, D., Collins, M., Coulsen, R., Davies, R., Dyer, P. S., Farman, M., Fedorova, N., … Denning, D. W. (2005). Genomic sequence of the pathogenic and allergenic filamentous fungus *Aspergillus fumigatus*. *Nature*, *438*(7071), 1151–1156.

Palmer, J. M. & Keller, N. P. (2010). Secondary metabolism in fungi: Does chromosomal location matter? *Current Opinion in Microbiology*, *13*(4), 431–436.

Perrin, R. M., Fedorova, N. D., Jin, W. B., Cramer, R. A., Wortman, J. R., Kim, H. S., Nierman, W. C. & Keller, N. P. (2007). Transcriptional regulation of chemical diversity in *Aspergillus fumigatus* by LaeA. *PLoS Pathogens*, *3*(4), 508–517.

Price, M. N., Dehal, P. S. & Arkin, A. P. (2009). FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Molecular Biology and Evolution*, *26*(7), 1641–1650.

Rasmussen, J. L. N. (2017). *Genomic Diversity in the Genus of* Aspergillus.

Rice, J. A. (2003). Mathematical statistics and data analysis. *China Machine Press*.

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V & Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, *31*(19), 3210–3212.

Steenwyk, J. L., Balamurugan, C., Raja, H. A., Gonçalves, C., Li, N., Martin, F., Berman, J., Oberlies, N. H., Gibbons, J. G. & Goldman, G. H. (2024). Phylogenomics reveals extensive misidentification of fungal strains from the genus *Aspergillus*. *Microbiology Spectrum*, *12*(4), e03980-23.

Steenwyk, J. L., Buida III, T. J., Li, Y., Shen, X.-X. & Rokas, A. (2020). ClipKIT: a multiple sequence alignment trimming software for accurate phylogenomic inference. *PLoS Biology*, *18*(12), e3001007.

Steenwyk, J. L., Knowles, S., Bastos, R. W., Balamurugan, C., Rinker, D., Mead, M. E., Roberts, C. D., Raja, H. A., Li, Y. & Colabardini, A. C. (2023). Evolutionary origin, population diversity, and diagnostics for a cryptic hybrid pathogen. *BioRxiv*.

Steenwyk, J. L., Shen, X. X., Lind, A. L., Goldman, G. H. & Rokas, A. (2019). A robust phylogenomic time tree for biotechnologically and medically important fungi in the genera *Aspergillus* and *Penicillium*. *MBio*, *10*(4), 1–25.

Studt, L., Janevska, S., Arndt, B., Boedi, S., Sulyok, M., Humpf, H.-U., Tudzynski, B. & Strauss, J. (2017). Lack of the COMPASS component Ccl1 reduces H3K4 trimethylation levels and affects transcription of secondary metabolite genes in two plant–pathogenic *Fusarium* species. *Frontiers in Microbiology*, *7*, 2144.

Team, R. D. C. (2010). R: A language and environment for statistical computing. *(No Title)*.

Theobald, S. (2018). *Genus-level studies of gene dynamics for the* Aspergillus *genus*.

Theobald, S., Vesth, T. C., Geib, E., Nybo, J. L., Frisvad, J. C., Larsen, T. O., Kuo, A., LaButti, K., Lyhne, E. K. & Kjærbølling, I. (2024). Genomic Analysis of *Aspergillus* Section *Terrei* Reveals a High Potential in Secondary Metabolite Production and Plant Biomass Degradation. *Journal of Fungi*, *10*(7), 507.

Theobald, S., Vesth, T. C., Rendsvig, J. K., Nielsen, K. F., Riley, R., de Abreu, L. M., Salamov, A., Frisvad, J. C., Larsen, T. O., Andersen, M. R. & Hoof, J. B. (2018). Uncovering secondary metabolite evolution and biosynthesis using gene cluster networks and genetic dereplication. *Scientific Reports*, *8*(1), 1–12.

Vesth, T. C., Nybo, J. L., Theobald, S., Frisvad, J. C., Larsen, T. O., Nielsen, K. F., Hoof, J. B., Brandl, J., Salamov, A., Riley, R., Gladden, J. M., Phatale, P., Nielsen, M. T., Lyhne, E. K., Kogle, M. E., Strasser, K., McDonnell, E., Barry, K., Clum, A., … Andersen, M. R. (2018). Investigation of inter- and intraspecies variation through genome sequencing of *Aspergillus* section *Nigri*. *Nature Genetics*, *50*(12), 1688–1695.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W. & Bright, J. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*, *17*(3), 261–272.

Visagie, C. M., Hirooka, Y., Tanney, J., Whitfield, E., Mwange, K., Meijer, M., Amend, A. S., Seifert, K. A. & Samson, R. A. (2014). Aspergillus, Penicillium and Talaromyces isolated from house dust samples collected around the world. *Studies in Mycology*, *78*, 63–139.

Visagie, C. M., Yilmaz, N., Kocsubé, S., Frisvad, J. C., Hubka, V., Samson, R. A. & Houbraken, J. (2023). A review of recently introduced Aspergillus, Penicillium, Talaromyces and other Eurotiales species. *Studies in Mycology*.

Wang, Y., Tang, H., Debarry, J. D., Tan, X., Li, J., Wang, X., Lee, T. H., Jin, H., Marler, B., Guo, H., Kissinger, J. C. & Paterson, A. H. (2012). MCScanX: A toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Research*, *40*(7), 1–14.

Waskom, M. L. (2021). Seaborn: statistical data visualization. *Journal of Open Source Software*, *6*(60), 3021.

Wiemann, P. & Keller, N. P. (2014). Strategies for mining fungal natural products. *Journal of Industrial Microbiology and Biotechnology*, *41*(2), 301–313.

Wiemann, P., Sieber, C. M. K., Von Bargen, K. W., Studt, L., Niehaus, E.-M., Espino, J. J., Huss, K., Michielse, C. B., Albermann, S. & Wagner, D. (2013). Deciphering the cryptic genome: genome-wide analyses of the rice pathogen *Fusarium fujikuroi* reveal complex regulation of secondary metabolism and novel metabolites. *PLoS Pathogens*, *9*(6), e1003475.

Wiemann, P., Willmann, A., Straeten, M., Kleigrewe, K., Beyer, M., Humpf, H. & Tudzynski, B. (2009). Biosynthesis of the red pigment bikaverin in *Fusarium fujikuroi*: genes, their function and regulation. *Molecular Microbiology*, *72*(4), 931–946.

Xu, F., Li, X., Ren, H., Zeng, R., Wang, Z., Hu, H., Bao, J. & Que, Y. (2022). The first telomere-to-telomere chromosome-level genome assembly of *Stagonospora tainanensis* causing sugarcane leaf blight. *Journal of Fungi*, *8*(10), 1088.

Yang, K., Tian, J. & Keller, N. P. (2022). Post-translational modifications drive secondary metabolite biosynthesis in *Aspergillus*: a review. *Environmental Microbiology*.

Zdobnov, E. M., Kuznetsov, D., Tegenfeldt, F., Manni, M., Berkeley, M. & Kriventseva, E. V. (2021). OrthoDB in 2020: evolutionary and functional annotations of orthologs. *Nucleic Acids Research*, *49*(D1), D389–D393.

Zhang, Noberini, R., Bonaldi, T., Collemare, J. & Seidl, M. (2022). The histone code of the fungal genus Aspergillus uncovered by evolutionary and proteomic analyses. *Microbial Genomics*, *8*.

Zhang, Noberini, R., Vai, A., Bonaldi, T., Seidl, M. & Collemare, J. (2023). Detection and quantification of the histone code in the fungal genus Aspergillus. *Fungal Genetics and Biology*, *167*, 103800.

Zhang, Y., Parmigiani, G. & Johnson, W. E. (2020). ComBat-seq: batch effect adjustment for RNA-seq count data. *NAR Genomics and Bioinformatics*, *2*(3), lqaa078.