The histone code of the fungal genus *Aspergillus* uncovered by evolutionary and proteomic analyses

Xin Zhang[1,2], Roberta Noberini[3], Tiziana Bonaldi[3,4], Jérôme Collemare[2,*,+], Michael F. Seidl[1,*,+]

Author affiliations: [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; [3]Department of Experimental Oncology, IEO, European Institute of Oncology IRCCS, Via Adamello 16, 20139 Milan, Italy; [4]Department of Oncology and Haematology-Oncology, University of Milano, Via Santa Sofia 9/1, 20122 Milano, Italy

[+]These authors contributed equally

[*]Correspondence: J. Collemare (telephone: (+31) 030 21 22 600; email: j.collemare@wi.knaw.nl) and M. F. Seidl (telephone: (+31) 030 253 36 94; email: m.f.seidl@uu.nl)

Abstract

Chemical modifications of DNA and histone proteins impact the organization of chromatin within the nucleus. Changes in these modifications, catalyzed by different chromatin-modifying enzymes, influence chromatin organization, which in turn is thought to impact the spatial and temporal regulation of gene expression. While combinations of different histone modifications, the histone code, have been studied in several model species, we know very little about histone modifications in the fungal genus *Aspergillus*, whose members are generally well-studied due to their importance as models in cell and molecular biology as well as their medical and biotechnological relevance. Here, we used phylogenetic analyses in 94 Aspergilli as well as other fungi to uncover the occurrence and evolutionary trajectories of enzymes and protein complexes with roles in chromatin modifications or regulation. We found that these enzymes and complexes are highly conserved in Aspergilli, pointing towards a complex repertoire of chromatin modifications. Nevertheless, we also observed few recent gene duplications or losses, highlighting *Aspergillus* species to further study the roles of specific chromatin modifications. SET7 (KMT6) and other components of PRC2 (Polycomb Repressive Complex 2), which is responsible for methylation on histone H3 at lysine 27 in many eukaryotes including fungi, are absent in Aspergilli as well as in closely related *Penicillium* species, suggesting that these lost the capacity for this histone modification. We corroborated our computational predictions by performing untargeted mass spectrometry analysis of histone post-translational modifications in *Aspergillus nidulans*. This systematic analysis will pave the way for future research into the complexity of the histone code and its functional implications on genome architecture and gene regulation in fungi.

Keywords

Aspergilli, chromatin, DNA methylation, histone modification, H3K27methylation, mass spectrometry

Data Summary

The predicted proteomes used in this study are publicly available at the JGI (Joint Genome Institute) MycoCosm repository (1); the species names and abbreviations are listed in Supplementary Table 1. 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 Supplementary Table 2. The fasta, trimmed alignment, and maximum-likelihood phylogenetic tree files can be found in Supplementary Data 1 and 2 deposited at Zenodo (10.5281/zenodo.6586562). The mass spectrometry results mentioned in Supplementary Table 6 are deposited in the PRIDE database with the dataset identifier PXD033478.

Impact Statement

Uncovering how fungi regulate gene expression and genome organization is an important step to understand how they control biological processes such as growth or biosynthesis of bioactive molecules. Despite the known importance of chromatin modifications in controlling a plethora of biological processes across eukaryotes, they remain particularly poorly understood in filamentous fungi, even in model organisms like Aspergilli. Our systematic evolutionary analysis provides a robust framework for the functional analysis of chromatin modifications in *Aspergillus* species and other fungi. Our results do not only implicate candidate enzymes to play a role in new chromatin modifications, but they also point at species that have experienced duplications or losses of genes encoding enzymes for specific chromatin modifications. We therefore expect that this work will set the stage for future

65 research into the complexity of the histone code and its functional implications on gene
66 regulation and genome organization in fungi.

67 **Introduction**

68 Inside eukaryotic cells, the membrane-bound nucleus facilitates the storage and replication of
69 the genetic material as well as the regulation of gene expression (2). Within the nucleus, DNA
70 is arranged as a chromatin fiber formed by repeated nucleosome units consisting of histone
71 protein octamers (two copies of each H2A, H2B, H3, and H4 histones) that are wrapped by
72 145-147 bp of DNA (3,4). The chromatin structure is regulated by chemical modifications such
73 as methylation of DNA on cytosines (5mC) and adenines (6mA), and different post-
74 translational modifications (PTMs) on the histone proteins, including methylation and
75 acetylation (5,6). The combination of these different modifications, the histone code,
76 influences the recruitment of diverse nucleosome-associated proteins (e.g., DNA binding
77 proteins), which can cause dynamic transitions between the condensed and transcriptionally
78 silent heterochromatin and the accessible and transcriptionally active euchromatin (7–9).
79 Chromatin dynamics is therefore key for organisms to adapt gene expression patterns in
80 response to a variety of developmental and environmental clues.

81

82 In the past decades, several enzymes involved in orchestrating chromatin dynamics
83 have been identified and functionally characterized in diverse eukaryotes (5,9,10). These
84 enzymes can be broadly classified into three distinct types: (i) writers that place modifications
85 on nucleosomes or linker DNA; (ii) erasers that remove these modifications; and (iii) readers
86 that first recognize specific modifications and subsequently recruit writers, erasers, or other
87 proteins to the correct genomic location (11). Chromatin modifiers that catalyze the same type
88 of chemical modification often harbor functionally and evolutionarily conserved catalytic
89 domains. For instance, DNA methylation (5mC) is catalyzed by different DNA
90 methyltransferases (DNMTs) which contain a DNA methyltransferase domain, while histone
91 mono-, di-, and tri-methylation are nearly exclusively catalyzed by proteins containing a SET
92 (Su (var) 3-9, E(z) and Trithorax) domain (10). Next to these catalytic domains, the enzymes
93 often contain accessory domains, which can contribute to the overall stability of the enzyme
94 or establish interactions with other proteins (12). Furthermore, these multi-domain proteins,

95 together with other accessory protein subunits, can form multi-protein complexes to find the
96 correct target region witin the genome and modify the exact amino acid residue, collectively
97 contributing to the specific functions of these complexes (13,14). For example, the histone
98 methyltransferase complex PRC2 (Polycomb Repressive Complex 2) is responsible for the
99 mono-, di-, and tri-methylation of histone protein H3 on the lysine at position 27 (abbreviated
100 as H3K27me1/2/3; the same abbreviation principle will be applied below for other histone
101 PTMs) (15–18). Evolutionary analyses have suggested that the key enzymes responsible for
102 chromatin modifications have evolved and siginificantly expanded early in eukaryotic evolution
103 (19). An analysis of 94 gene families with roles in chromatin modifications reveals that 87 out
104 of 94 gene families emerged more than one billion years ago, and 48 were already present in
105 the last eukaryotic common ancestor (19).
106
107 Among eukaryotes, the fungal kingdom is remarkably large and diverse, comprising
108 an estimated 1,5-5 million species with various morphologies, life cycles, and ecological
109 niches (20). The most well-studied model fungal species in respect of chromatin modifications
110 is the unicellular yeast *Saccharomyces cerevisiae*, in which 75 histone PTMs on 43 distinct
111 amino acid positions have been reported with histone methylation and acetylation being most
112 abundant (11). In addition to ancient chromatin modifiers, also lineage-specific innovations
113 such as duplications and losses were found in fungi. For example, Dim-2 and RID (Repeat-
114 Induced Point Mutation (RIP)-Defective) are considered to be fungal-specific DNMTs, while
115 DNMT3 that is typically found in animals and plants seems to be absent in fungi (21–23).
116 However, *S. cerevisiae* is not representative of the complete fungal kingdom, and detailed
117 studies on PTMs are needed for other, less studied fungal species.
118
119 The fungal genus *Aspergillus* comprises approximately 350 species, including
120 pathogens of humans, animals, and crops, food contaminants, as well as important cell
121 factories for industrial and medical applications (24). Chromatin dynamics influences gene
122 expression in several Aspergilli and regulates growth, sexual development, secondary

123 metabolite biosynthesis, and virulence (11,25–27). Thus far, very few well-known chromatin
124 modifications and their catalytic protein complexes have been identified and studied in few
125 *Aspergillus* species (11). For example, gene knockout of the H3K9 methyltransferase *Clr4*
126 homolog in *Aspergillus fumigatus* reduced histone H3K9me1/3 results in reduction of radial
127 growth and conidia production, as well as delayed conidiation (26). Similarly, CclA, a subunit
128 of the COMPASS (COMplex of Proteins ASsociated with SET1) complex, is essential for
129 H3K4me2/3 in *A. fumigatus*, and the deletion strain showed increased secondary metabolite
130 production and decreased growth (28). The SAGA (Spt-Ada-Gcn5 Acetyltransferase) complex,
131 which is responsible for histone acetylation, has been functionally characterized in *Aspergillus*
132 *nidulans* (29). Notably, this complex in *A. nidulans* lacks the Sgf11 and Sus1 subunits that
133 have been described in *S. cerevisiae*, suggesting that this complexe evolves in distinct
134 trajectories in different fungi (29). Besides these few examples, we so far lack a complete view
135 of the occurrence and evolution of chromatin modifiers in the *Aspergillus* genus. Here, we
136 report the phylogenetic analyses of core enzymes involved in the deposition and removal of
137 histone PTMs, and assess the conservation of all the subunits belonging to chromatin
138 regulator complexes in 94 diverse Aspergilli. We also performed proteomics analyses to
139 corroborate our computational predictions and reveal a broader spectrum of histone
140 modifications.

141 **Material and Methods**

142 **Acquisition of predicted proteomes for 109 fungal species**

143 The predicted proteomes of 109 fungal species were retrieved from JGI (Joint Genome
144 Institute) MycoCosm (https://mycocosm.jgi.doe.gov/mycocosm/home) (1) on 24th January
145 2020. These species comprise 94 different Aspergilli species, 13 Ascomycota species from
146 other genera, as well as two Basidiomycota species (*Coprinopsis cinerea* and *Dendrothele*
147 *bispora*) (Supplementary Table 1).

148

149 **Phylogenetic analyses of chromatin modifiers**

150 Phylogenetic trees for six major chromatin modifiers were generated using their conserved
151 catalytic domains. Catalytic domain HMMs (Hidden Markov Model) of DNA methyltransferase
152 domain (PF00145), histone methylation SET domain (PF00856), acetyltransferase domain
153 (PF00583), histone acetyltransferase MOZ/SAS FAMILY domain (PF01853), histone
154 deacetylase domain (PF00850), and histone demethylase Jumonji C (JmjC) domain
155 (PF02373) were downloaded from the Pfam database (https://pfam.xfam.org). Hmmsearch,
156 which is part of the HMMER package (v3.1b2) (30), with the Gathering Cut-Off threshold (-
157 cut_ga) was used to identify occurrences of these domains in the predicted proteomes of all
158 109 fungal species considered here. To filter out short and incomplete fragments, we applied
159 a cutoff calculated by 50% length coverage of the query sequence (domain) or the hit.
160 Subsequently, the identified domain sequences were aligned using MAFFT v7.271 (31).
161 TrimAl v1.2 (-gt 0.1) was used to remove positions in the alignments with gaps in more than
162 90% of the sequences (32). Maximum-likelihood phylogenetic trees were constructed using
163 IQ-TREE v1.6.10 (33,34) using ultrafast bootstrap and Shimodaira–Hasegawa approximate
164 likelihood ratio test (SH-aLRT) (35) with 1,000 pseudo-replicates, and standard model
165 selection to automatically determine the best-fit model. We predicted additional domains in
166 the analyzed proteins using Hmmsearch with all Pfam domain profiles that are present in the
167 Pfam database (v28.0) and visualized them along the phylogenetic trees using iTOL (36).
168 Lastly, additional TBLASTN searches were performed on the NCBI website

169 (https://blast.ncbi.nlm.nih.gov/Blast.cgi) and checked manually to confirm dubious absences
170 in the domain trees.
171
172 **Species tree construction**
173 To construct a species phylogeny of the here analyzed fungi, we used BUSCO v4.0.1
174 (Benchmarking Universal Single-Copy Ortholog assessment tool) (37) to identify the
175 occurrence of 758 fungal single-copy genes (Supplementary Table 2) within the 109 predicted
176 proteomes (Supplementary Table 1). The identified homologs of single-copy BUSCO genes
177 were subsequently concatenated into a single super-alignment. Then, the maximum-likelihood
178 species phylogeny was estimated using IQ-TREE v1.6.10 (33,34) by employing partitioned
179 model selection, which allows selecting a substitution model for each BUSCO gene. Branch
180 supports were obtained using ultrafast bootstrap and SH-aLRT (35) with 1,000 pseudo-
181 replicates separately.
182
183 **Chromatin modifier subunits homology search and gene tree construction**
184 We obtained previously studied chromatin modifier complexes from literature, and the
185 associated protein sequences were downloaded from NCBI (https://www.ncbi.nlm.nih.gov)
186 (Supplementary Table 4). For each subunit of the chromatin modifier complexes, homologs in
187 the analyzed 109 predicted proteomes were detected based on sequence similarity searches.
188 First, BLASTP (v2.2.31+) (38) searches were performed, and an e-value threshold of 10^{-2} was
189 applied. Multiple BLAST matches (High-Scoring Pairs: HSPs) of the same query were
190 concatenated if they were aligned with different non-overlapping regions. Only the non-
191 redundant region was retained when HSPs of the same query were overlapping. For these
192 two conditions, the highest e-value of the different matches was used to represent the merged
193 match. Only matches with the e-value lower than 10^{-5} and covering more than 50% of query
194 or subject were retained for the following analyses. If we observed less than ten hits for the
195 BLASTP search after filtering in Aspergilli, we used PSI-BLAST (Position-Specific Initiated
196 BLAST v2.2.31+) (38) with five iterations (-num_iterations 5) using the best Aspergilli hit from

197 the BLASTP as the query; if no Aspergilli match was retained, we used the same query as for
198 the BLASTP searches. The same filter criterion as for the BLASTP selection was applied to
199 the PSI-BLAST hits, and the results of these two homology search strategies were
200 subsequently jointed.
201
202 To account for apparent protein absences due to erroneous gene annotation, we
203 tested these potential losses by using TBLASTN (BLAST v2.2.31+) (38) to query protein
204 sequences against the genome assemblies. Matches with e-value lower than 10^{-5} together
205 with 4,500 nucleotides upstream and downstream were extracted, and Exonerate v2.2.0 (39)
206 was used to predict the gene structure (options: protein2genome, extensive search,
207 showtargetgff, showcigar, and showquerygff options were set as true, and ">%ti(%tcb
208 - %tce)\n%tcs\n" option was used to output the complete aligned coding sequence). Transeq
209 v6.6.0.0 (40) was used to translate the coding sequence into protein sequences. Together
210 with the protein sequences identified from BLASTP and PSI-BLAST, all protein sequences
211 from each subunit were retrieved to build a phylogenetic tree (MAFFT, TrimAl, and IQ-TREE
212 settings were the same as for the tree reconstruction using the catalytic domains above).
213
214 For each sequence included in the phylogenic trees, the domain organization was
215 identified and visualized as described above. When the domain organization of a *Aspergillius*
216 protein differed from its orthologues, the corresponding gene model was manually checked in
217 the MycoCosm genome browser (1) and corrected. The presence and absence patterns of 83
218 subunits from 15 different chromatin modifier complexes were visualized using the Seaborn
219 library in Python (41,42). Lastly, additional TBLASTN searches for dubious absences in the
220 outgroup species were performed on the NCBI website as absences could be caused by older
221 and more fragmented assemblies in the JGI database compared to NCBI.
222
223 **Histone enrichment and mass spectrometry**

224 Spores of 4- or 5-days old sporulating *A. nidulans* (AnWT pabaA1 veA1, kindly provided by
225 Prof. Joseph Strauss) growing on ME (Malt Extract) agar medium were harvested with 2 mL
226 sterile water and cultured in 50 mL ME liquid medium at 28 ℃, 200 rpm for 16 hours. The
227 sample was centrifuged at 8,000 g for 10 minutes to discard supernatant, and germinating
228 spores were resuspended in 25 mL 0.8 M NaCl and centrifuged again. Ten mL protoplasting
229 solution (20 mg/mL *Trichoderma* lysing enzyme and 5 mg/mL driselease in 0.8 M NaCl) were
230 used to resuspend spores, which were then incubated at 30 ℃, 100 rpm for 2-3 hours.
231 Protoplasts were retrieved by filtering through sterile Miracloth and were concentrated by
232 centrifugation at 3,000 rpm, 4 ℃ for 5 minutes. Histone proteins were extracted according to
233 an adapted histone protein extraction protocol (43). Briefly, 10^8 protoplasts were resuspended
234 in 1 mL Nuclei Isolation Buffer (1×PBS, 0.5 mM PMSF, 5 μM Leupeptin, 5 μM Aprotinin, and
235 5mM Na-butyrate) with 1% Triton X-100, followed by vigorous pipetting through a 200 μL
236 pipette several times. Then, centrifuged at 2,300 g for 15 minutes at 4 ℃, discarded the
237 supernatant, and resuspended the nuclear pellet in 200 μL Nuclei Isolation Buffer with 0.1%
238 SDS and 250 U Pierce™ Universal Nuclease and incubated at 37 ℃ for 30 minutes to digest
239 nucleic acids. The protein concentration was measured using the Bradford Protein Assay (44).
240
241 The samples were separated on a 17% SDS-PAGE gel and histone bands were
242 excised, chemically acylated with propionic anhydride and in-gel digested with trypsin,
243 followed by peptide N-terminal derivatization with phenyl isocyanate (PIC) (45). Chemical
244 acylation of lysines (which occurs on unmodified or mono-methylated residues) impairs trypsin
245 cleavage, resulting in proteolityc cleavage at arginine residues only. This treatment allows
246 obtaining histone peptides of proper length for MS analysis. The samples were then desalted
247 on handmade StageTips columns (46). Peptide mixtures were separated by reversed-phase
248 chromatography on an EASY-Spray column (Thermo Fisher Scientic), 25-cm long (inner
249 diameter 75 μm, PepMap C18, 2 μm particles), which was connected online to a Q Exactive

250 Plus instrument (Thermo Fisher Scientific) through an EASY-Spray™ Ion Source (Thermo
251 Fisher Scientific), as described (45).
252
253 The acquired RAW data were analyzed using the integrated MaxQuant software
254 v.1.6.10. The Uniprot UP000000560 database was used for identification of *A. nidulans*
255 histone peptides. Enzyme specificity was set to Arg-C. The estimated false discovery rate
256 (FDR) was set at a maximum of 1%. The mass tolerance was set to 6 ppm for precursor and
257 fragment ions. Two missed cleavages were allowed, and the minimum peptide length was set
258 to 4 amino acids. Min. score for modified peptides and min. delta score for modified peptides
259 were set to 1. Variable modifications included lysine propionylation, monomethylation +
260 propionylation, dimethylation, trimethylation and acetylation. N-terminal PIC labeling was set
261 as a fixed modification (45). To reduce the search time and the rate of false positives, with
262 increasing the number of variable modifications included in the database search, the raw data
263 were analyzed through multiple parallel MaxQuant jobs (47), setting different combinations of
264 variable modifications. Peptides identified by MaxQuant with Andromeda score higher than 50
265 and localization probability score higher than 0.75 were quantitated, either manually or by
266 using a version of the EpiProfile 2.0 software (48) adapted to the analysis of histones from *A.*
267 *nidulans*. Identifications and retention times were used to guide the manual quantification of
268 each modified peptide using QualBrowser version 2.0.7 (Thermo Fisher Scientific). Site
269 assignment was evaluated from MS2 spectra using QualBrowser and MaxQuant Viewer.
270 Extracted ion chromatograms (XICs) were constructed for each doubly charged precursor,
271 based on its *m/z* value, using a mass tolerance of 10 ppm. For each histone modified peptide,
272 the % relative abundance (%RA) was estimated by dividing the area under the curve (AUC)
273 of each modified peptide for the sum of the areas corresponding to all the observed forms of
274 that peptide (49). The AUC values are reported (Supplementary Table 6) and visualized using
275 GraphPad Prism.

276 **Results and discussion**

277 DNA methylation, histone methylation, and histone acetylation are among the most studied
278 chromatin modifications in animals, plants, and fungi (50–52). Protein complexes involved in
279 writing, reading, and erasing these modifications have been reported in few model fungal
280 species to date (16,22), but the evolution and conservation of these complexes in diverse
281 fungal species have not been largely systematically explored. Besides some research on DNA
282 methyltransferases (21,23) and SET domain-containing proteins (17) that have been
283 performed throughout the fungal kingdom, we still lack the evolutionary overview of other
284 chromatin modifiers and in depth analyses for important fungal groups. To start uncovering
285 the occurrence and evolutionary histories of chromatin modifiers involved in these major
286 modifications in the fungal genus *Aspergillus,* we first focused on six conserved catalytic
287 domains reported to be involved in histone modifications: DNA methyltransferase domain for
288 DNMTs, SET domain for histone methyltransferases (HMTs), acetyltransferase domain and
289 MOZ/SAS FAMILY domain for histone acetyltransferases (HATs), histone deacetylase
290 domain for histone deacetylases (HDACs), and JmjC domain for histone demethylases
291 (HDMTs). We identified homologs harboring these conserved domains in 109 predicted
292 proteomes; the completeness of the predicted proteomes was generally high as 97 had more
293 than 97% complete BUSCO predicted proteins (Supplementary Table 1). The 94 Aspergilli
294 species cover 18 recognized sections: *Polypaecilum*, *Cremei*, *Restricti*, *Aspergillus*, *Cervini*,
295 *Clavati*, *Fumigati*, *Candidi*, *Circumdati*, *Janorum*, *Terrei*, *Flavi*, *Ochraceorosei*, *Bispori*,
296 *Cavernicolarum*, *Usti*, *Nidulantes*, and *Nigri* (53–59), and 15 other species were used as
297 outgroups (Supplementary Figure 1). Two of the outgroup species are Basidiomycota and the
298 remaining species cover each class in the Ascomycota phylum, which provides a robust
299 framework for our evolutionary analyses within Aspergilli. The outgroup species include well-
300 studied model species, some of which are important for crop production, food fermentation,
301 and antibiotics production (60,61). We used the identified homologs for each catalytic domain
302 to determine their evolutionary history using maximum-likelihood phylogenies (Figure 1-4 and
303 Supplementary Figure 2 and 3).

304

305 **Aspergilli harbor conserved DNA methyltransferases for genome defense and**
306 **methylation maintenance**

307 The conserved DNA methyltransferase domain is the only domain known to catalyze the
308 addition of methyl groups to DNA specifically at cytosines (5mC) (21). Based on previous
309 domain architecture analyses, phylogeny reconstructions, and functional experiments, five
310 distinct protein families containing a DNA methyltransferase domain have been recognized in
311 fungi: DNMT1, DNMT2, and DNMT5 are commonly found in eukaryotes, while Dim-2 and RID
312 are considered to be fungal specific (21). Among these five families, DNMT2 seems to function
313 as a tRNA methyltransferase rather than DNA methyltransferase (62). In line with previous
314 observation (21,63), DNMT2 only occurs in our study in two Basidiomycota and one
315 Ascomycota (*Schizosaccharomyces pombe*), suggesting it has been lost in most Ascomycota.

316

317 Three (DNMT1, DNMT5, and Dim-2) of the four remaining families catalyze the
318 maintenance of DNA methylation pattern from parental DNA strand onto synthesized daughter
319 DNA strands, while Dim-2 is also able to catalyze *de novo* methylation (64,65). Based on our
320 species selection and phylogenetic analyses, we here found *DNMT1* and *Dim-2* to be absent
321 in all Aspergilli and only present in four or seven out of 15 outgroup species, respectively,
322 while *DNMT5* is found in most Aspergilli and outgroup species (Figure 1). Only four outgroup
323 species harboring *DNMT1*, the two Basidiomycota species, *Botrytis cinerea*, and *Plectania*
324 *melastoma*, which corroborates previous findings (21,66). *Dim-2* was identified only in seven
325 Ascomycota outgroup species, and matches in *B. cinerea* and *Zymoseptoria tritici* were
326 included after an additional TBLASTN search in NCBI (Supplementary Table 6). The original
327 *Dim-2* absence in *B. cinerea* by local Hmmsearch is caused by the older version and
328 fragmented assembly at JGI compared to NCBI. The *Z. tritici* IPO323 strain used in this study
329 carries a non-functional *Dim-2* as previously reported (67) and no significant hit was found
330 with Hmmsearch or BLASTP searches, but several fragmented hits with TBLASTN search in

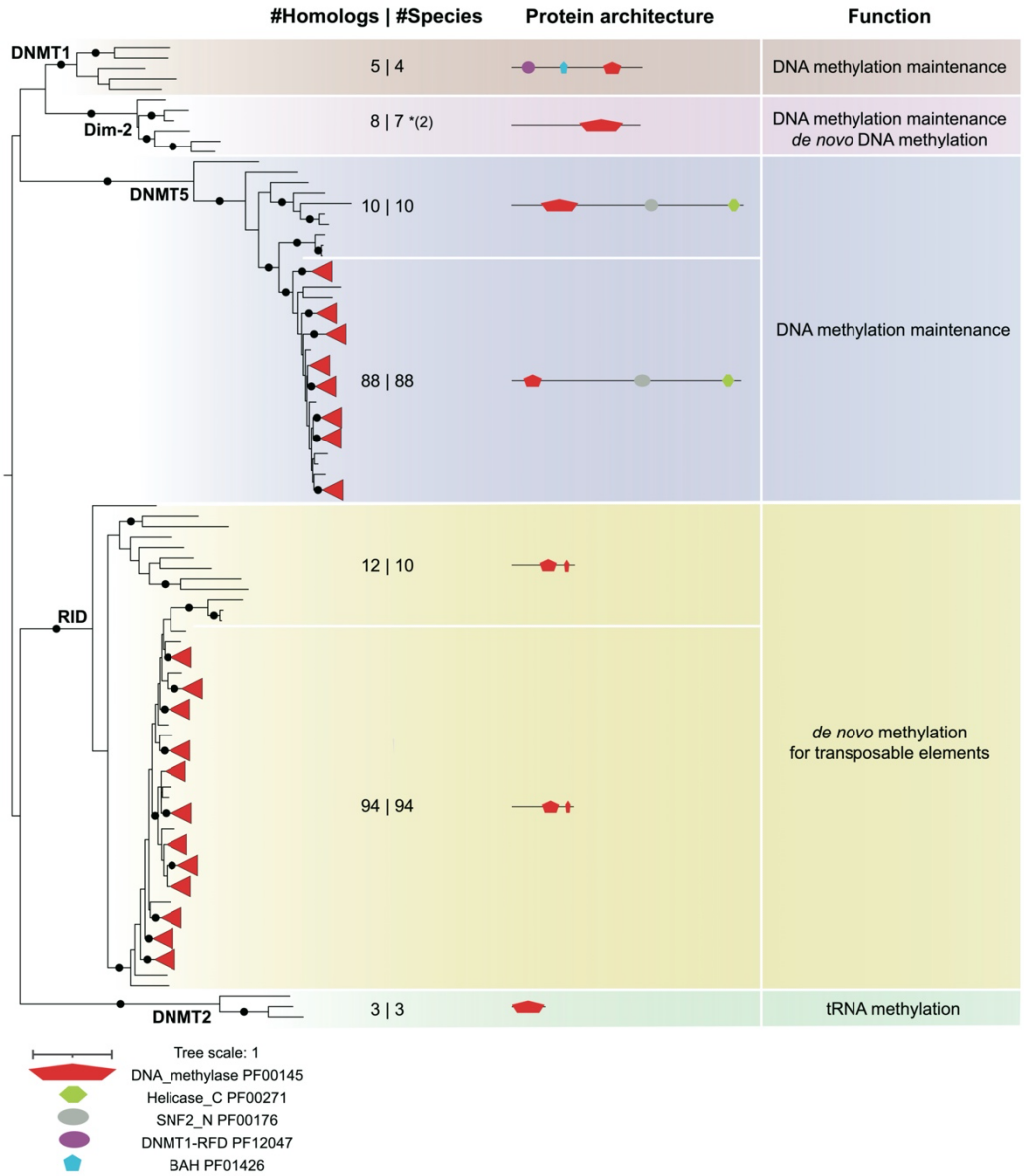

Figure 1. Conservation of DNA methyltransferase families in the *Aspergillus* genus and outgroup species. The maximum-likelihood phylogeny of DNA methyltransferase domain (PF00145) found in 94 Aspergilli and 15 outgroup species was determined using IQ-TREE (37,38). The black dot on the branch indicates ultrafast bootstrap values over 95 and SH-aLRT values over 80. Leaves from Aspergilli with trustworthy support are collapsed and shown as red triangles. The numbers of hits and species are separated by the pipe in the '#Homologs | #Species' column. The number in parentheses after * indicates the additional hits found by the TBLASTN search on NCBI. Additional conserved domains in DNA methyltransferase domain-containing proteins were identified using the PFAM database (https://pfam.xfam.org).

NCBI. This loss of *Dim-2* function in *Z. tritici* IPO323 is very recent because other isolates harbor a functional *Dim-2* gene (67). Its loss leads to increased activity of transposable elements (TEs) and decreased mutation rate, thus influencing genome defense and genome evolution (67). In contrast to the *DNMT1* and *Dim-2* losses in Aspergilli, *DNMT5* is identified

346 in 88 out of 94 Aspergilli and 10 out of 15 outgroup species (Figure 1), with a conserved
347 domain architecture consisting of a DNA methyltransferase domain, an SNF2 family domain
348 (PF00176), and a Helicase conserved C-terminal domain (PF00271), in which the latter two
349 domains are involved in transcription regulation (68). The five outgroup species without
350 DNMT5 are *S. cerevisiae*, *S pombe*, *Neurospora crassa*, *Fusarium graminearum*, and
351 *Cladonia grayi*, among which the two yeast species also lack the other two DNA methylation
352 maintaining enzymes, and thus accordingly are considered to be incapable of DNA
353 methylation (69–71). Five out of the six *Aspergillus* species that lost *DNMT5* belong to a single
354 clade within the *Nidulantes* section (*A. undulatus*, *A. olivicola*, *A. filifera*, *A. venezuelensis*, and
355 *A. stella-maris*), and the sixth one is *A. robustus* which is assigned to the *Circumdati* section
356 (Supplementary Figure 1). In the Basidiomycota fungus *Cryptococcus neoformans*, DNMT5
357 maintains DNA methylation using the same cofactors as DNMT1 (70,72). Thus, it is most likely
358 that DNMT5 is similarly involved in DNA methylation maintenance in Aspergilli as in *C.*
359 *neoformans*. Moreover, DNA methylation maintenance does not appear to be required for the
360 survival of fungi as shown with the absence of all three *DNMT1*, *DNMT5*, and *Dim-2* not only
361 within the *Nidulantes* section but also in *S. cerevisiae* and *S. pombe* (Figure 1). However, the
362 absence of these genes could have an impact on the regulation of certain genes like
363 secondary metabolite biosynthetic gene clusters and it would be interesting to compare
364 species within the *Nidulantes* section in more detail to identify how such loss may have
365 affected gene expression and genome organization.
366
367 RID is a more recently acquired single domain DNMT that specifically occurs in fungi
368 and is involved in a fungal-specific genome defense mechanism called RIP (Repeat-Induced
369 Point mutation) (21). RID is a putative *de novo* methyltransferase for TEs in both CG and non-
370 CG contexts relating to RIP during the sexual stage in *N. crassa*; however, the capacity of RID
371 to catalyze 5mC has not yet been demonstrated (73). Here, we identified RID in all analyzed
372 Aspergilli species as well as in ten outgroup species. A RID-like homolog knock-out mutant in
373 the model filamentous Ascomycota *Podospora anserina* shows blocked sexual development,

374 as gametes formation and fertilization can happen but dikaryotic cells cannot develop normally
375 (74). Interestingly, the knock-out mutant of *DmtA* (homolog of *RID* in *A. flavus*) shows altered
376 conidiation, sclerotial production, aflatoxin biosynthesis, and virulence (27), suggesting that
377 RID might have additional functions next to genome defense. Additionally, fungi lost the gene
378 for *de novo* DNA methyltransferase, *DNMT3,* around 50–150 million years ago (72,75), and
379 only *DNMT5* and *RID* were consistently found in Aspergilli (Figure 1). Thus, we hypothesize
380 that Aspergilli can only maintain DNA methylation *via* DNMT5 but do not spur genome-wide
381 *de novo* DNA methylation. This hypothesis seems to be corroborated as bisulfite sequencing
382 in *A. flavus* has uncovered very little genome-wide DNA methylation (76). However,
383 systematic analyses of DNA methylation patterns and experimental evidence for the roles of
384 DNMTs in Aspergilli are still largely missing.
385

386 **Aspergilli lack *SET7* (*KMT6*), but SET-domain proteins are abundant and other histone**
387 **methyltransferases are conserved**

388 Histone methylation is typically catalyzed by SET domain-containing proteins (17,77). We
389 identified a total of 1,548 SET domain-containing proteins and classified them into 15 groups
390 based on their domain architecture and phylogenetic support that contain species both from
391 Aspergilli as well as from the outgroup. Among these groups, 107, 111, 109, 102, and 108
392 sequences represent homologs of the known histone methyltransferases (HMTs) SET9
393 (KMT5), Dim-5 (KMT1), SET1 (KMT2), Ash1 (KMT2H), and SET2 (KMT3), respectively
394 (Figure 2). The remaining homologs were divided into nine groups that we broadly classified
395 as SET domain-containing proteins. Among them, eight groups contain a SET domain but lack
396 experimentally determined functions, and 117 sequences could be grouped as Ribosomal
397 lysine N-methyltransferases (Figure 2). Enzymes in this group contain a SET domain together
398 with a Rubisco LSMT substrate-binding domain (PF09273). While the homolog in plant
399 transfers a methyl group to the large subunit of the Rubisco holoenzyme complex (78,79), the

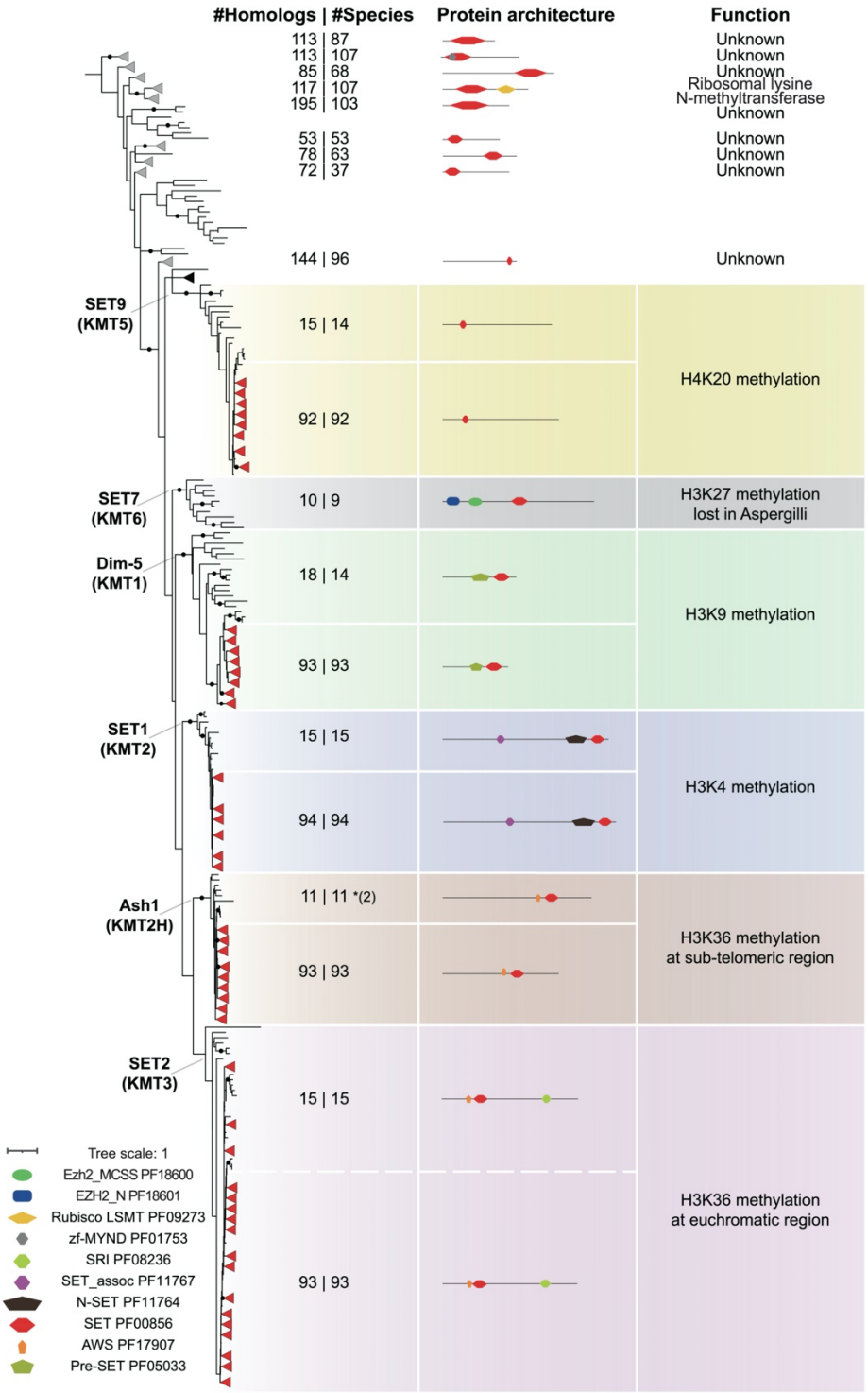

Figure 2. Conservation of histone methyltransferase families in the *Aspergillus* genus and outgroup species. The maximum-likelihood phylogeny of SET domains (PF00856) found in 94 Aspergilli and 15 outgroup species was determined using IQ-TREE (37,38). The black dot on the branch indicates ultrafast bootstrap values over 95 and SH-aLRT values over 80. Leaves from Aspergilli with trustworthy support are collapsed and shown as red triangles. Leaves with similar protein architecture but unknown functions covering both Aspergilli and outgroup species are collapsed and shown as grey triangles. The numbers of hits and species are separated by the pipe in the '#Homologs | #Species' column. The number in parentheses after * indicates the additional hits found by the TBLASTN search on NCBI. Additional conserved domains in SET domain-containing proteins were identified using the PFAM database (https://pfam.xfam.org).

411 homolog in *S. cerevisiae* has been characterized as a ribosomal lysine N-methyltransferase
412 (80).
413
414 The well-described HMTs, SET1, Dim-5, SET2, Ash1, and SET9, all contain distinct
415 accessory domains and catalyze methylation on lysine residues at different positions of H3 or
416 H4 tails (17). SET1 is composed of the SET domain, the N-SET domain (PF11764), as well
417 as the SET_assoc domain (PF11767), and is responsible for H3K4me1/2/3 in *S. cerevisiae*
418 (81). SET1 occurs in a single copy in all species (Figure 2), indicating that this enzyme and its
419 catalyzed PTM play fundamental roles in fungal biology. Dim-5 contains a SET domain and
420 Pre-SET domain (PF05033) and acts as an H3K9 methyltransferase inducing heterochromatin
421 formation notably in *N. crassa* (82). Dim-5 was found in all species included in this study
422 except *S. cerevisiae* and *A. calidoustus*, among which *S. cerevisiae* is reported to lack
423 H3K9me (82,83). How such a loss has impacted gene expression in *A. calidoustus* in
424 comparison to its closest relative *A. carlsbadensis* deserves further investigation. Both SET2
425 and Ash1 catalyze H3K36me in *N. crassa* and share the SET domain and the AWS domain
426 (PF17907) (Figure 2), yet SET2 harbor one additional SRI domain (PF08236) that interacts
427 with RNA polymerase II and links H3K36me with transcript elongation (84–86). In *N. crassa,*
428 Ash1 was shown to contribute to H3K36me at sub-telomeric regions, while SET2 acts primarily
429 at euchromatic regions (87,88). Considering the common origin of SET2 and Ash1 as well as
430 their similar activities, an ancient duplication followed by the loss or gain of the SRI domain is
431 likely to have contributed to their functional diversification (Figure 2). *SET2* was found in all
432 outgroup species, while *Ash1* was lost in both Basidiomycota and two yeasts. False absences
433 of *Ash1* in *B. cinerea* and *F. graminearum* were corrected as they could be retrieved by
434 TBLASTN search at NCBI (Supplementary Table 6). Both enzymes are conserved in all
435 *Aspergillus* species except in *A. carbonarius* from the *Nigri* section, suggesting that either
436 other SET domain-containing proteins exhibit functional redundancy or that H3K36
437 methylation is not fully required in this species. In *Fusarium fujikuroi*, deletion of *SET2* and
438 *Ash1* strongly affected vegetative growth and conidiation (88). SET9 homologs harbor only a

439 SET domain, and were shown to catalyze H4K20me in *S. pombe* (89). This group is well
440 conserved in most of the species except in *S. cerevisiae* (90) and in the two closely related
441 Aspergilli, *A. novoparasiticus* and *A. arachidicola* that both belong to section *Flavi* (Figure 2
442 and Supplementary Figure 1).
443
444 SET7 is the catalytic subunit of PRC2, and is highly conserved in mammals, plants,
445 and *N. crassa*, in which it has been well-characterized to catalyze methylation on H3K27, a
446 PTM that is deposited at facultative heterochromatin (16). Even though PRC2 is considered
447 to be a key chromatin modifier complex, its catalytic subunit *SET7* was lost in many fungal
448 species (17). *SET7* has previously been reported to be absent in *A. nidulans* and *A. fumigatus*
449 (15–17), and we here demonstrate that *SET7* is absent from all analyzed Aspergilli as well as
450 from *Penicillium* (Figure 2), suggesting that this enzyme was most likely lost at or before the
451 divergence of the last common ancestor of these two fungal genera. Thus, in line with previous
452 reports (17,25,91,92), our results indicate that methylation on H3K4, H3K9, H3K36, and
453 H4K20 are likely catalyzed by conserved enzymes in most Aspergilli, while H3K27 methylation
454 is absent. The loss of this methylation is usually reported to trigger serious phenotypes in
455 humans (94) and mice (95). Loss of H3K27me3 in fungi typically leads to gene expression
456 changes without severe growth defects (16, 93), and thus further investigating how Aspergilli
457 and other fungi are able to cope with such a loss will likely provide new insights on the histone
458 code and its evolution in fungi.
459
460 Importantly, the occurrence of a high number of uncharacterized SET domain-
461 containing proteins opens the possibility that these contribute to not yet characterized histone
462 modifications. Especially, four uncharacterized SET domain-containing groups with 144, 195,
463 113, and 113 members are conserved in Aspergilli and thus likely play important roles in
464 different cellular processes.
465
466 **Histone acetyltransferases are conserved in Aspergilli**

467 Histone acetyltransferases (HATs) have been classified into five families: MYST (MOZ,
468 Ybf2/Sas3, Sas2, and Tip60)-related HATs, Gcn5-related acetyltransferases (GNATs),
469 p300/CBP HATs, general transcription factor HATs, and nuclear hormone-related HATs (96–
470 98). We here only focused on the first two HAT families because these are known to be
471 involved in the regulation of histone proteins and appear largely conserved in fungi (96).The
472 MOZ/SAS FAMILY domain is the catalytic domain of MYST-related HATs (99). As previously
473 reported (100), the three characterized MYST-related HATs, Sas3, Sas2, and Esa1, belong
474 to distinct monophyletic clades, and these paralogues are nearly fully conserved in the here
475 analyzed fungi (Figure 3). MOZ/SAS FAMILY domain and zf-MYST domain (PF17772) are
476 present in Sas3, which can on its own acetylate histone H3 and H4, and H2A weakly *in vitro*
477 (101). Within the NuA3 (Nucleosomal Acetyltransferase of histone H3) complex, Sas3 histone
478 acetylation activity is *in vitro* restricted on H3K14 and to a less extent on H3K23, while in *S.*
479 *cerevisiae*, only H3K14ac can be detected (102). *Sas3* occurs in most of the species but is
480 lost in *S. pombe*, as previously observed (103,104), and is duplicated in *A. carbonarius* (Figure
481 3). In addition to the two domains found in Sas3, Sas2 and Esa1 carry a PHD_4 domain
482 (PF16866) or Tudor-knot domain (PF11717), respectively (Figure 3). *Esa*1 is present as a
483 single copy in all species included in this study, and its functions is known to contribute to the
484 acetylation of K5, K8, K12, K16 on histone H4, as well as on H2AZK14 in *S. cerevisiae* (105–
485 108). *Sas2* is similarly found in most of the species except for two Basidiomycota as well as
486 for *A. carbonarius*. Sas2 is involved in chromatin-mediated gene regulation and H4K16
487 acetylation in both *S. cerevisiae* and *Candida albicans* (105,109,110). Although Esa1 and
488 Sas2 catalyze H4K16ac, it was shown in *C. albicans* that they are differentially recruited at
489 different stages of development (105). Because both Esa1 and Sas2 are well-conserved in
490 the *Aspergillus* genus, we hypothesize that these two HATs may also have distinct roles during
491 development.
492
493 GNATs have a acetyltransferase domain as its catalytic core, and 2,604
494 acetyltransferase domain-containing proteins were found in our search (Supplementary

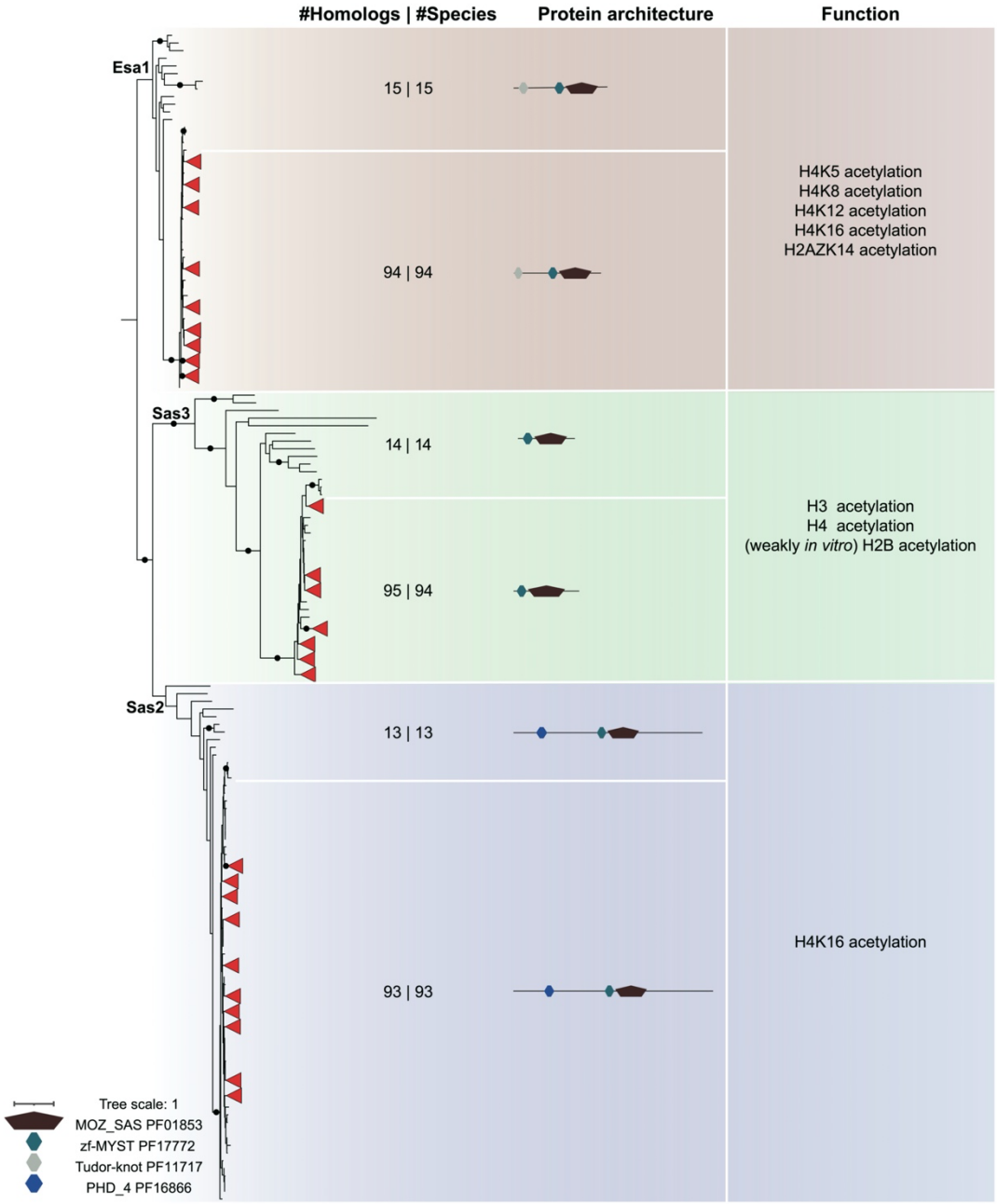

Figure 3. Conservation of MOZ/SAS FAMILY domain-containing histone acetyltransferases in the *Aspergillus* genus and outgroup species. The maximum-likelihood phylogeny of MOZ/SAS FAMILY domains (PF01853) found in 94 Aspergilli and 15 outgroup species was determined using IQ-TREE (37,38). The black dot on the branch indicates ultrafast bootstrap values over 95 and SH-aLRT values over 80. Leaves from Aspergilli with trustworthy support are collapsed and shown as red triangles. The numbers of hits and species are separated by the pipe in the '#Homologs | #Species' column. '*In vitro*' indicates the acetylation activity of this enzyme was only observed on the substrate *in vitro*. Additional conserved domains in MOZ/SAS FAMILY domain-containing proteins were identified using the PFAM database (https://pfam.xfam.org).

Figure 2). The genes for two well-characterized GNATs, *Gcn5* and *Elp3*, occur in a single copy in every species included in our study (Supplementary Figure 2). Gcn5 harbors a

508 Bromodomain (PF00439) in addition to the Acetyltransferase domain and is a crucial subunit
509 of distinct histone acetyltransferase complexes such as the HAT-A2 complex, the SAGA
510 complex, and the ADA complex in *S. cerevisiae* (96,98). All these complexes contain two
511 coactivator proteins, Ada2 and Ada3, that directly bind to Gcn5 (111) and the recruitment of
512 different other subunits determines the location where they catalyze acetylation (96). Elp3 is
513 characterized by the Acetyltransferase domain, Radical SAM (PF04055), and Radical SAM
514 C-terminal domain (PF16199), and has been shown to catalyze H3K14ac and H4K8ac in *S.*
515 *cerevisiae* (112–114). Elp3 also exhibits a dual enzymatic activity as it catalyzes
516 demethylation in mouse zygotes (115), which is due to the Radical SAM domain that catalyzes
517 the demethylation of methylated lysyl residues on histones (116).
518
519 Next to Gcn5 and Elp3, we have identified 2,386 additional acetyltransferase domain-
520 containing proteins. Because their functions remain unknown due to the lack of homology to
521 sequences with characterized functions, they are valuable candidates that may acetylate
522 histones or other proteins. Interestingly, it has been shown that Gcn5 also exhibits
523 succinyltransferase activity in humans (117) and can also function as histone
524 crotonyltransferase to regulate gene transcription *in vitro* (118). It is expected that other less
525 common histone modifications, i.e., histone lysine butyrylation, propionylation, or malonylation,
526 may also be catalyzed by histone acetyltransferases because these histone modifications all
527 use acyl-Coenzyme A to complete the reactions (119–121). Further research is needed to
528 determine whether these less common histone modifications are catalyzed by the known
529 histone acetyltransferases Gcn5 and Elp3, or as of yet unstudied HATs. In conclusion, all five
530 characterized fungal HATs described above are highly conserved in the *Aspergillus* genus,
531 showing their importance in fungal biology. Many other acetyl-transferase domain-containing
532 proteins are encoded in *Aspergillus* genomes and may be responsible for histone acetylation
533 or additional less common histone modifications like butyrylation, succinylation, and
534 malonylation.
535

536 **Duplication of *RpdA* histone deacetylase gene in the *Flavi* and *Janorum* sections**

537 Five families of histone deacetylases (HDACs) can be grouped into two classes based on
538 sequence similarity: class I (HOS1, HOS2, and Rpd3/RpdA) and class II (HdaA and HOS3)
539 (122,123). They all harbor the Histone deacetylase catalytic domain, and our phylogenetic
540 tree corroborates the HDAC classification into two distinct classes (Figure 4). *HdaA* is found
541 as a single copy in each species included in this study (Figure 4). It harbors the Histone
542 deacetylase domain and C-terminal Arb2 domain (PF09757) that serves as an anchor to target
543 centromeric heterochromatin region (124,125). HdaA deacetylates lysine in histones H3 and
544 H2B, but not H4 nor H2A in *S. cerevisiae* (126,127), and the Δ*hdaA* knock-out strains show
545 increased or reduced production of different secondary metabolites in *A. nidulans* (128), *A.*
546 *niger* (128), and *A. fumigatus* (129). By contrast, *HOS3* is found in only 80 *Aspergillus* species,
547 as it is lost in nine species (*A. nomiae*, *A. pseudocaelatus*, *A. tamarii*, *A. pseudotamarii*, *A.*
548 *caelatus*, *A. sergii*, *A. novoparasiticus*, *A. minisclerotigenes*, *A. arachidicola*) in the *Flavi*
549 section, three species (*A. desertorum*, *A. stercorarius*, and *A. multicolor*) in the *Nidulantes*
550 section, and two species (*A. heteromorphus* and *A. carbonarius*) in the *Nigri* section. *HOS3* is
551 well conserved in all outgroup species except *S. pombe* (Figure 4) (130). HOS3 has been
552 shown to catalyze deacetylation with distinct specificity for histone H2AK7, H2BK11, H3K14
553 and H3K23, as well as H4K5 and H4K8 in *S. cerevisiae* cell extracts (131). Notably, *HOS1*
554 was only detected in three outgroup species, *S. cerevisiae* and both Basidiomycota (Figure
555 4). The specificity of Hos1 in catalyzing deacetylation on histone proteins is not known yet
556 (132), but it can catalyze the deacetylation of the Smc3 subunit of Cohesin and influence the
557 regulation of chromosome segregation during mitosis in *S. cerevisiae* (133). *HOS2* is found in
558 15 outgroup species and most of Aspergilli except *A. oryzae* (Figure 4). It is known to
559 deacetylate lysine residues in H3 and H4 histone tails (134). The conservation of *HOS2* in
560 Aspergilli suggests that it plays a key role, which is supported by the deletion of *HOS2* in *A.*
561 *niger* where the deletion strains displays severe reduction in growth, sporulation, SM
562 biosynthesis, and stress resistance (128).

563

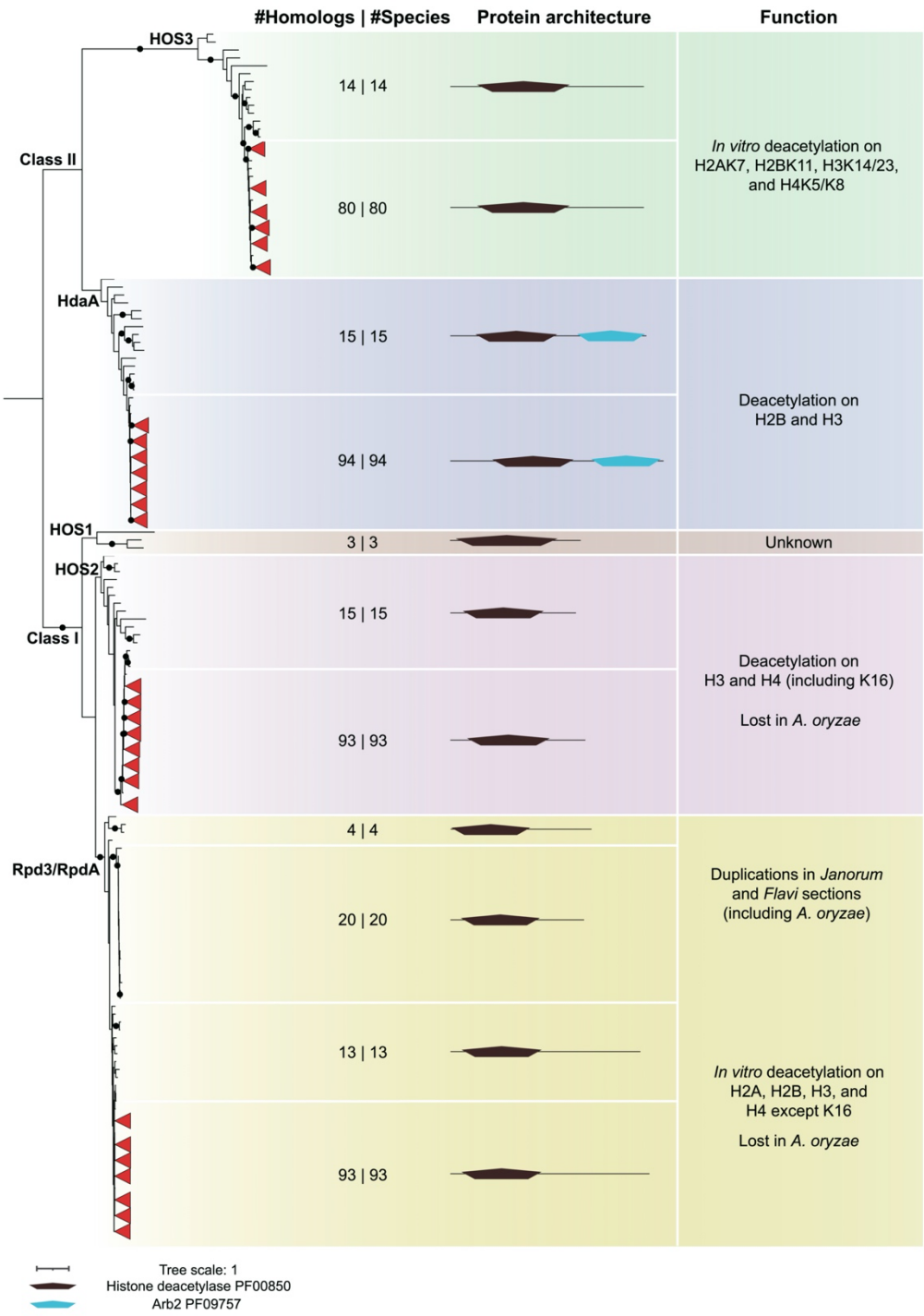

Figure 4. Conservation of histone deacetylase domain-containing proteins in the *Aspergillus* genus and outgroup species. The maximum-likelihood phylogeny of Histone deacetylase domains (PF00850) found in 94 Aspergilli and 15 outgroup species was determined using IQ-TREE (37,38). The black dot on the branch indicates ultrafast bootstrap values over 95 and SH-aLRT values over 80. Leaves from Aspergilli with trustworthy support are collapsed and shown as red triangles. The numbers of hits and species are separated by the pipe in the '#Homologs | #Species' column. '*In vitro*' indicates the deacetylation activity of this enzyme was only observed on the substrate *in vitro*. Additional conserved domains in histone deacetylase domain-containing proteins were identified using the PFAM database (https://pfam.xfam.org).

575 Within class I HDACs, the Rpd3/RpdA clade is divided into two sub-clades (Figure 4).
576 One is nearly fully conserved (except in *A. oryzae*) and comprises the well-characterized
577 Rpd3/RpdA HDACs, while the other one only comprises sequences from the *Flavi* and
578 *Janorum* sections (Figure 4). Our phylogenetic reconstruction suggests a duplication event
579 that has occurred in the ancestor of all Aspergilli giving rise to a paralog that was retained in
580 two sections only and could explain the *A. oryzae* loss in the conserved sub-clade. The two
581 paralogs from the same species vary in length as they share a conserved N-terminal catalytic
582 domain but have a different C-terminus (Figure 4). Follow-up studies are needed to investigate
583 the function of both paralogues in *Aspergillus* species. Rpd3 is a crucial deacetylase in
584 eukaryotes, as it is required for deacetylation in most locations on core histone proteins,
585 except H4K16 *in vitro* (135) that is catalyzed by HOS2 (134). *S. cerevisiae* lacking *Rpd3* is
586 sensitive to high osmolarity and shows compromised expression of osmotic stress genes
587 (136). Deletion of *RpdA* in *A. nidulans* is lethal (137), while only heterokaryon disruptants could
588 be obtained in *A. oryzae* (138), indicating its crucial role in the fungal genus *Aspergillus*. In
589 summary, class I and class II deacetylases are well conserved in Aspergilli, and duplication
590 events like the one observed for *RpdA* may confer new activities. While only one paralog was
591 retained in *A. oryzae*, the presence of two paralogs in sections *Flavi* and *Janorum* may provide
592 functional redundancy or diversification, which remains to be determined.
593
594 Lastly, we reconstructed the phylogeny of histone demethylases harboring the Jumonji
595 C (JmjC) domain (PF02373) (Supplementary Figure 3). Six groups are distinguished by their
596 domain distribution: KDM4A/JHDM3/JMJD2, KDM5/JARID, DMM-1, JMJD6/PKDM11,
597 KDM2/Epe1, and KDM6B/JMJD3, which also agrees with a previous study (91). The first five
598 groups occur in most of the outgroup and *Aspergillus* species, but the *KDM6B/JMJD3* is only
599 present in several outgroup species but not in any of the *Aspergillus* species. Certain JmjC
600 domain-containing proteins were shown to catalyze the demethylation of histone proteins, e.g.,
601 KDM5/JARID for H3K4me2/3 and KDM4A/JHDM3/JMJD2 for H3K9me2/3 and H3K36me2/3
602 (139). The histone demethylase activity has not been detected for DMM-1 (140) or Epe1 (141)

603 in *N. crassa*, while KDM2 can demethylate H3K36me1/2 in mouse (142) and JMJD6/PKDM11
604 is an arginine demethylase in *Arabidopsis* (143). Interestingly, the human KDM6B/JMJD3 is
605 responsible for the demethylation of H3K27me2/3 (144). As we did not find *SET7* and thus
606 conclude that H3K27me3 is absent, it would be tempting to speculate that H3K27 demethylase
607 should be absent in *Aspergillus* too. However, based on the fact that other enzymes, e.g.,
608 UTX can also catalyze the demethylation of H3K27 (144,145), we cannot unambiguously
609 correlate the absence of *KDMB6B/JMJD3* to *SET7* loss. More studies are needed to
610 investigate the regulation mechanism and interplay of proteins for H3K27 methylation.
611
612 **Presence-absence patterns of subunits that compose histone modifier complexes**
613 Histone PTMs are catalyzed by the conserved domains found in HMTs, HATs, and HDACs
614 (Figure 2-4, Supplementary Figure 2), but these chromatin modifying enzymes are often not
615 acting alone. Instead, they are part of protein complexes that comprise accessory subunits,
616 which are involved in the recognition and induction of specific histone modifications at the
617 correct locations (13,14). To further investigate the presence-absence patterns (PAPs) of
618 known chromatin modifier complexes in Aspergilli, we focused on 15 chromatin modifier
619 complexes with 83 subunits previously identified and studied in *S. cerevisiae* or *N. crassa*
620 (Supplementary Table 4). Forty-four subunits are present in all *Aspergillus* and outgroup
621 species. Most of the accessory subunits are either fully conserved or absent in Aspergilli
622 (Figure 5 and Supplementary Table 4). Ten subunits show PAPs in both Aspergilli and
623 outgroup species, while nine subunits appear specific to *S. cerevisiae* as they are not found
624 in any other species and are likely to play a specific role in this species. Seventeen subunits
625 show PAPs in the outgroup species, while eight are absent and nine are present in the
626 Aspergilli. Lastly, three subunits are present in all outgroup species but show PAPs in
627 Aspergilli.
628
629 Five histone methyltransferase complexes, COMPASS, DCDC (Dim-5/-7/-
630 9/CUL4/DDB1 Complex), SET2 Complex, PRC2, and Dot1 (disruptor of telomeric silencing 1)

631 (KMT4) were included in our analyses (Figure 5). Five COMPASS subunits (*SET1*, *CclA*,
632 *Swd1*, *Swd2*, and *Swd3*) are fully conserved, suggesting an important function for H3K4
633 methylation. Consistent with a PAP in most outgroup species and the loss in Aspergilli, Spp1p
634 is not essential for H3K4 methylation (146). As *Sdc1* shows a PAP in Aspergilli and outgroup
635 species, it is likely that this subunit also does not play a key role in H3K4 methylation. In the
636 DCDC complex, while Dim-7 and Dim-9 seem to be dispensable for H3K9 methylation, Dim-
637 5, Cul4, and Dim-8 are highly conserved, and we thus expect these enzymes to play key roles
638 in the activity of this complex (Figure 5 and Supplementary Table 4). However, it was shown
639 that Dim-7 and Dim-9 are responsible for heterochromatin recognition, which leads to
640 subsequent recruitment of the complete DCDC to induce H3K9 methylation in *N. crassa* (147).
641 Thus, it is very likely that other accessory subunits are involved in recruiting DCDC at specific
642 chromatin locations in Aspergilli. Furthermore, the unique absence of *Dim-8* in *A. indicus* may
643 suggest altered H3K9 methylation in this species. Consistent with the absence of *SET7*
644 (Figure 2), all components of the PRC2 complex are lost in Aspergilli (Figure 5). Lastly, the
645 interaction and crosstalk of multiple chromatin modifications in *Aspergillus* species should be
646 further explored. For example, heterochromatin formation and regulation are done by
647 H3K9me3, DNA methylation, and H3K27me3 in *N. crassa*. In this species, H3K9me3 is
648 recognized by an adapter protein, HP1, to recruit DNA methyltransferase Dim-2 for DNA
649 methylation (73), and cause the redistribution of H3K27me3 (15). In *Aspergillus* species,
650 although HP1 and the writer for H3K9me3, DCDC complex, are well conserved (Figure 5), the
651 inability for genome-wide *de novo* DNA methylation and H3K27me3 suggest different
652 mechanisms of heterochromatin formation and regulation in Aspergilli which remain to be
653 determined.
654
655 We included eight histone acetyltransferase complexes that differ by their distinct
656 catalytic subunits (Figure 5). The HAT-A2, ADA, SAGA, and Elongator complexes use an
657 acetyltransferase domain-containing protein as catalytic subunit (29,148). The SAS
658 (Something About Silencing), NuA4 (Nucleosomal Acetyltransferase of histone H4), and NuA3

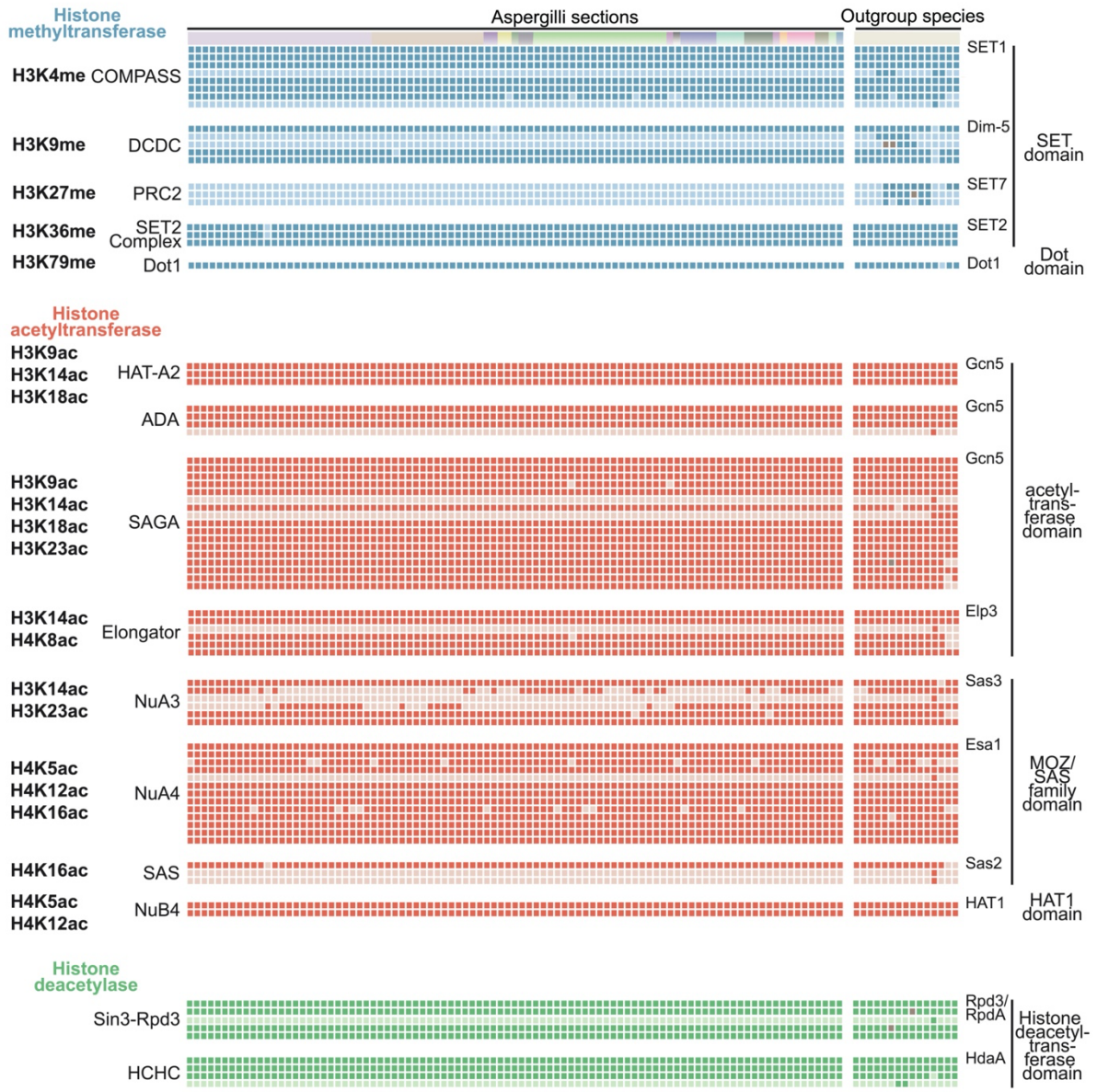

Figure 5. Presence-absence patterns of subunits that constitute characterized histone modification complexes. The top row with distinct color boxes indicates outgroup species and Aspergilli sections, the order of the species from right to left is the same as in the species tree from top to bottom (Supplementary Figure 1) without *Fusarium oxysporum* and *Aspergillus kawachii* since they were excluded in the later analysis. Each row represents a subunit in the histone modification complexes, the order agrees with the summary list (Supplementary Table 5). Each row shows the presence/absence of a specific complex subunit. For each complex, the first row represents the catalytic subunit. Each box in the matrix indicates the presence (dark color) or absence (light color) of a subunit (Supplementary Table 4). The brown boxes indicate additional hits found with TBLASTN search on NCBI (Supplementary Table 6).

complexes use a MOZ/SAS FAMILY domain-containing protein (149). The NuB4 (Nuclear Hat1p-containing type B histone acetyltransferase) complex uses the histone acetyltransferase HAT1 domain-containing protein (150). Gcn5, Ada2, and Ada3 are fully

674 conserved in all fungi and constitute the core enzymatic module of the HAT-A2, ADA, and
675 SAGA complexes. This catalytic module contributes to catalyzing histone acetylation, and
676 other subunits in these complexes guide them to specific locations and link them to the
677 transcriptional machinery (151). Besides the catalytic module, the ADA complex also harbors
678 the Ahc1 subunit that is found in *S. cerevisiae* only (148,152). Twelve additional subunits of
679 the SAGA complex are well conserved in Aspergilli and outgroup species (Figure 5). The
680 SAGA complex also contains a deubiquitylase (DUB) module, for which the composition
681 differs between Aspergilli and other species. *Sgf11* is unique to *S. cerevisiae* and *Sus1* was
682 likely lost in the ancestor of Pezizomycotina because we can only identify *Sus1* orthologs in
683 both Basidiomycota and the yeasts (Figure 5). The Elongator complex appears more variable
684 as four subunits are mostly conserved, but *Elp6* is specific to *S. cerevisiae*, and *Hpa2* and
685 *Hpa3* show PAPs in both Aspergilli and outgroup species.
686
687 The NuA4, SAS, and NuB4 complexes are well-conserved in Aspergilli as well as in
688 the outgroup species (Figure 5). These complexes appear slightly different in *S. cerevisiae*
689 with the presence of specific subunits (*Eaf5* in the NuA4 complex; *Sas4* and *Sas5* in the SAS
690 complex). Both SAS and NuA4 complexes can catalyze H4K16ac with opposite effects during
691 cell growth and morphogenesis in *C. albicans* (105). It would be interesting to assess whether
692 these two complexes exhibit similar activities in Aspergilli. For the NuA3 complex, two
693 accessory subunits are well conserved among Aspergilli while two others, *Pdp3*, and *Nto1*,
694 show different PAPs, and *Yng1* is specific to *S. cerevisiae*. *Pdp3* and *Nto1* seem to be lost in
695 a complementary manner, suggesting a possible functional redundancy (Figure 5). Finally,
696 two histone deacetylase complexes, Sin3-Rpd3 and HCHC, are fully conserved (Figure 5).
697 The Sin3-Rpd3 complex is slightly different in *S. cerevisiae* with the specific of the *Ume1*
698 subunit.
699 Overall, our analysis showed that the core catalytic modules of complexes involved in
700 histone PTMs are mostly conserved in all fungi, including Aspergilli, but variability in accessory
701 subunits could contribute to differences in the localization of histone PTMs along the genome.

702

703 **Unbiased mass spectrometry captures the diverse histone modifications in *A. nidulans***
704 **and verifies the absence of H3K27 methylation**

705 Our genomic and phylogenetic analyses have identified the presence of multiple chromatin
706 modifiers in Aspergilli, suggesting that these species have the ability to establish a wide variety
707 of histone modifications. To assess the presence of histone PTMs predicted in our
708 evolutionary analyses (Fig 2-5), we extracted histone proteins from *A. nidulans* and performed
709 unbiased mass spectrometry analyses (Figure 6A and 6B). H3 and H4 peptides with PTMs
710 corresponding to H3K4me, H3K9me, H3K36me, and H3K79me were detected (Figure 6C),
711 which we anticipated being catalyzed by SET1, Dim-5, Ash1 and SET2, and Dot1, respectively.
712 H4K20me, which is catalyzed by SET9 in *S. pombe* (89), was not detected (Figure 6C), which
713 is likely due to technical reasons because the unmodified form of histone H4 peptide 20-24
714 containing the K20 residue was not detected. For the acetylation on H3K9, H3K14, H3K18,
715 H3K23, H4K5, H4K8, H4K12, and H4K16, they were confidently detected (Figure 6C). We did
716 not observe any peptide that would indicate the presence of H3K27me in *A nidulans*, while we
717 did detect H3K27ac (Figure 6C and Supplementary Table 6). These results corroborate our
718 computational analysis that the PRC2 complex is absent in *A. nidulans*, as well as previous
719 reports on the absence of H3K27me3 in several Aspergilli (17,153,154). Thus, the mass
720 spectrometry results verify the occurrence of histone modifications predicted to be present
721 based on our comparative analyses in the fungal genus*Aspergillus*. Moreover, the differences
722 between the relative abundance of distinct histone modifications provide interesting
723 information to investigate these abundances and identify specific conditions under which these
724 modifications increase or decrease in abundance.

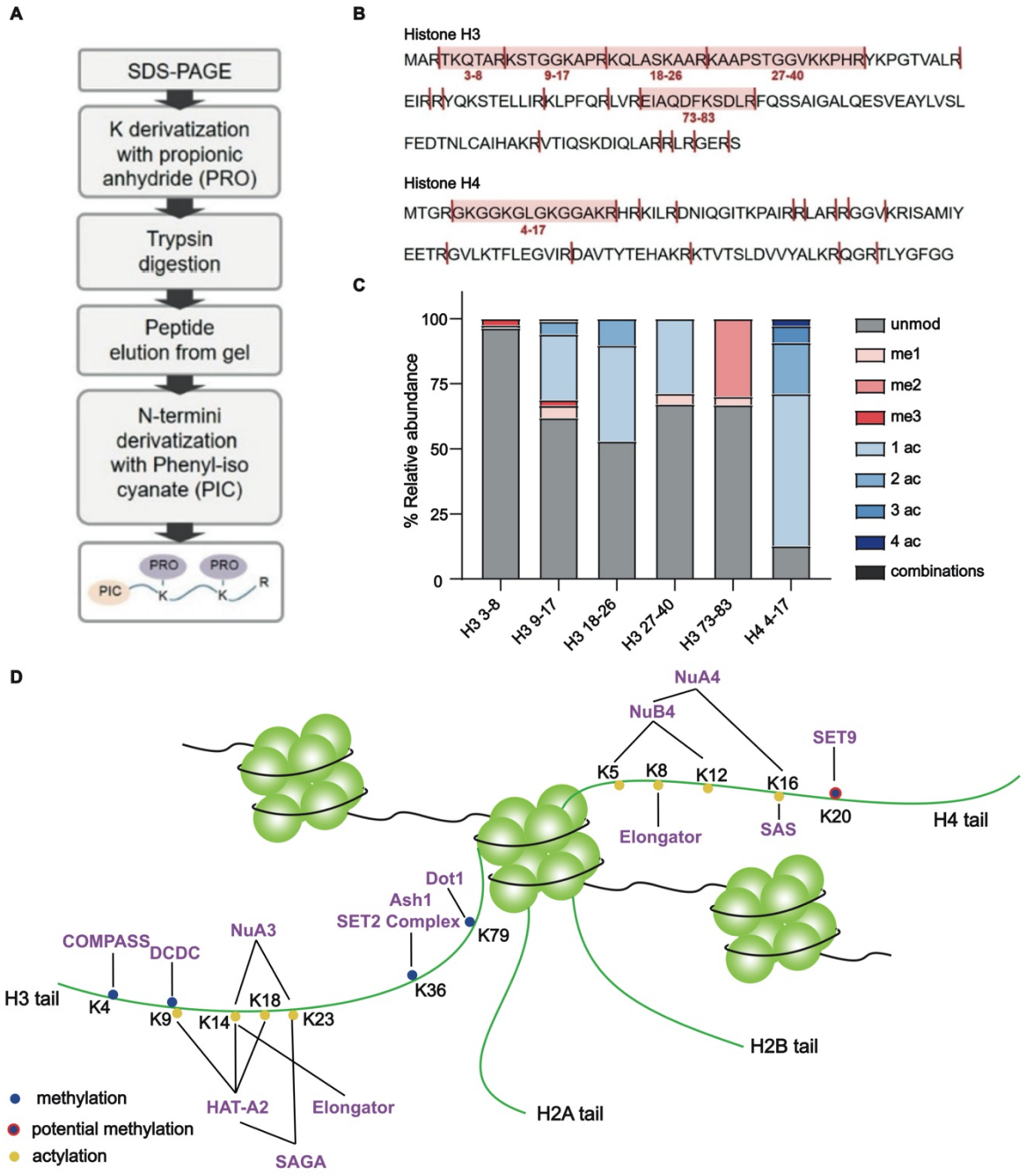

Figure 6. Mass spectrometry unlocks the histone code in Aspergilli. A. Schematic presentation of the in-gel digestion method used to process histones isolated in *A. nidulans*. After SDS-PAGE separation, gel bands corresponding to histone H3 and H4 are excised and chemically acylated with propionic anhydride (which derivatizes unmodified and mono-methylated lysines). Thanks to the derivatization, trypsin cutting at lysines is impaired, and histones are cleaved only at the C-terminus of arginine. Finally, digested peptides are N-terminally derivatized with phenyl isocyanate (PIC). **B.** *A. nidulans* histone H3 and H4 sequences. Cleavage sites after PRO-PIC digestion are shown. Peptides identified in this study are highlighted in light red and labeled using their starting-ending amino acids. **C.** Relative abundance of post-translational modifications detected on peptides of histone H3 and H4 from *A. nidulans* using unbiased mass spectrometry analysis. unmod: unmodified peptide; me1: monomethylation; me2: dimethylation; me3: trimethylation; ac: acetylation. The term “combinations” refers to multiple modifications occurring on the same peptide (Supplementary Table 6). **D.** Histone modifications are predicted by genomic

739 and phylogenetic analyses, and/or proved by mass spectrometry in *A. nidulans*. Histone
740 modifications verified by mass spectrometry are labeled as solid dots, while the ones that are only
741 predicted by phylogenic analysis are labeled as solid dots with a red outline.
742

743 **Conclusion**

744 Changes in chromatin organization play an important role in regulating the expression of
745 genes (130). The fungal genus *Aspergillus* is well known for its ability to thrive in a broad range
746 of different environmental niches and the huge potential of secondary metabolite biosynthesis
747 (11,155). However, a detailed and complete overview of occurrence and evolution of
748 chromatin modifiers in Aspergilli is a prerequisite to be able to better understand how these
749 fungi regulate gene expression. In the future, more research is needed to uncover how the
750 absence of *de novo* DNA methylation impacts genome evolution in Aspergilli. We demonstrate
751 that PRC2 is lost throughout Aspergilli, a finding consistent with the lack of H3K27me detection
752 by mass spectrometry. Although histone acetyltransferases and deacetylases are overall
753 conserved, our study also revealed specific duplication and loss events, which require further
754 investigation to understand their impact on chromatin and gene regulation in certain
755 *Aspergillus* species. Moreover, other less common histone modifications like histone lysine
756 succinylation, crotonylation, butyrylation, propionylation, and malonylation, are all likely
757 catalyzed by HATs and HDACs as they all use Coenzyme A (CoA) to complete the reactions
758 (119–121), but may require specific yet to be discovered subunits. In conclusion, our study
759 provides a detailed overview of the evolutionary routes of chromatin modifier complex in the
760 fungal genus *Aspergillus*, and therefore generates the necessary framework to identify targets
761 for functional studies to understand how chromatin regulates gene expression in Aspergilli.

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 the 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 and Joseph Strauss for providing *A. nidulans* strain. We thank Alessandro Vai for support with analysis of mass spectrometry data for histone PTMs' identification on *Aspergillus* core histone sequences.

Funding information

Xin Zhang is funded by the Chinese Scholarship Council (CSC) (201907720028). This work has been supported by EPIC-XS, project number 823839, funded by the Horizon 2020 programme of the European Union.

Authors and contribution

X.Z. formal analysis, investigation, methodology, visualization, writing original draft; R.N. formal analysis, investigation, methodology, visualization; T.B. funding acquisition, supervision, resources; 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.

782 **References**

783 1. Grigoriev I V, Nikitin R, Haridas S, Kuo A, Ohm R, Otillar R, et al. MycoCosm portal:
784 gearing up for 1000 fungal genomes. Nucleic acids research. 2014;42(D1):D699–704.
785 2. Labrador M, Corces VG. Setting the boundaries of chromatin domains and nuclear
786 organization. Cell. 2002;111(2):151–4.
787 3. Luger K, Mader AW, Richmond RK, Sargent DF, Richmond TJ. Crystal structure of
788 the nucleosome resolution core particle at 2.8 Å. Nature. 1997;389:251–60.
789 4. Luger K. Structure and dynamic behavior of nucleosomes. Current opinion in genetics
790 & development. 2003;13(2):127–35.
791 5. Jenuwein T, Allis CD. Translating the histone code. Science. 2001;293(5532):1074–
792 80.
793 6. Peterson CL, Laniel M-A. Histones and histone modifications. Current Biology.
794 2004;14(14):R546–51.
795 7. Santos-Rosa H, Schneider R, Bannister AJ, Sherriff J, Bernstein BE, Emre NCT, et al.
796 Active genes are tri-methylated at K4 of histone H3. Nature. 2002;419(6905):407–11.
797 8. Grewal SIS, Jia S. Heterochromatin revisited. Nature Reviews Genetics.
798 2007;8(1):35–46.
799 9. Berger SL. The complex language of chromatin regulation during transcription.
800 Nature. 2007;447(7143):407–12.
801 10. Xiao B, Wilson JR, Gamblin SJ. SET domains and histone methylation. Current
802 Opinion in Structural Biology. 2003;13(6):699–705.
803 11. Collemare J, Seidl MF. Chromatin-dependent regulation of secondary metabolite
804 biosynthesis in fungi: is the picture complete? FEMS Microbiology Reviews.
805 2019;43(6):591–607.
806 12. Helmlinger D, Papai G, Devys D, Tora L. What do the structures of GCN5-containing
807 complexes teach us about their function? Biochimica et Biophysica Acta - Gene
808 Regulatory Mechanisms. 2021;1864(2).
809 13. Hajheidari M, Koncz C, Bucher M. Chromatin evolution-key innovations underpinning

810 morphological complexity. Frontiers in Plant Science. 2019;10:1–12.
811 14. Gates LA, Foulds CE, O'Malley BW. Histone marks in the 'drivers seat': functional
812 roles in steering the transcription cycle. 2018;42(12):977–89.
813 15. Jamieson K, Wiles ET, McNaught KJ, Sidoli S, Leggett N, Shao Y, et al. Loss of HP1
814 causes depletion of H3K27me3 from facultative heterochromatin and gain of
815 H3K27me2 at constitutive heterochromatin. Genome Research. 2016;26(1):97–107.
816 16. Jamieson K, Rountree MR, Lewis ZA, Stajich JE, Selker EU. Regional control of
817 histone H3 lysine 27 methylation in *Neurospora*. Proceedings of the National
818 Academy of Sciences of the United States of America. 2013;110(15):6027–32.
819 17. Freitag M. Histone Methylation by SET Domain Proteins in Fungi. Annual Review of
820 Microbiology. 2017;71:413–39.
821 18. Kamei M, Ameri AJ, Ferraro AR, Bar-peled Y, Zhao F. IMITATION SWITCH is
822 required for normal chromatin structure and gene repression in PRC2 target domains.
823 2021;118(4).
824 19. Weiner AKM, Cerón-Romero MA, Yan Y, Katz LA. Phylogenomics of the epigenetic
825 toolkit reveals punctate retention of genes across eukaryotes. Genome Biology and
826 Evolution. 2020;1–39.
827 20. Choi J, Kim S-H. A genome tree of life for the fungi kingdom. Proceedings of the
828 National Academy of Sciences. 2017;114(35):9391–6.
829 21. Bewick AJ, Hofmeister BT, Powers RA, Mondo SJ, Grigoriev I V., James TY, et al.
830 Diversity of cytosine methylation across the fungal tree of life. Nature Ecology and
831 Evolution. 2019;3(3):479–90.
832 22. Kouzminova E, Selker EU. Dim-2 encodes a DNA methyltransferase responsible for
833 all known cytosine methylation in *Neurospora*. The EMBO journal. 2001;20(15):4309–
834 23.
835 23. Nai YS, Huang YC, Yen MR, Chen PY. Diversity of fungal DNA methyltransferases
836 and their association with DNA methylation patterns. Frontiers in Microbiology.
837 2021;11.

838 24. Samson RA, Visagie CM, Houbraken J, Hong SB, Hubka V, Klaassen CHW, et al.
839 Phylogeny, identification and nomenclature of the genus *Aspergillus*. Studies in
840 Mycology. 2014;78(1):141–73.
841 25. Cánovas D, Marcos AT, Gacek A, Ramos MS, Gutiérrez G, Reyes-Domínguez Y, et
842 al. The histone acetyltransferase GcnE (GCN5) plays a central role in the regulation
843 of *Aspergillus* asexual development. Genetics. 2014;197(4):1175–89.
844 26. Palmer JM, Perrin RM, Dagenais TRT, Keller NP. H3K9 methylation regulates growth
845 and development in *Aspergillus fumigatus*. Eukaryotic Cell. 2008;7(12):2052–60.
846 27. Yang K, Liang L, Ran F, Liu Y, Li Z, Lan H, et al. The DmtA methyltransferase
847 contributes to *Aspergillus flavus* conidiation, sclerotial production, aflatoxin
848 biosynthesis and virulence. Scientific Reports. 2016;6:1–13.
849 28. Palmer JM, Bok JW, Lee S, Dagenais TRT, Andes DR, Kontoyiannis DP, et al. Loss
850 of CclA, required for histone 3 lysine 4 methylation, decreases growth but increases
851 secondary metabolite production in *Aspergillus fumigatus*. PeerJ. 2013;2013(1):1–18.
852 29. Georgakopoulos P, Lockington RA, Kelly JM. The Spt-Ada-Gcn5 Acetyltransferase
853 (SAGA) Complex in *Aspergillus nidulans*. PLoS ONE. 2013;8(6).
854 30. Eddy SR. Accelerated profile HMM searches. PLoS computational biology.
855 2011;7(10):e1002195.
856 31. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7:
857 improvements in performance and usability. Molecular biology and evolution.
858 2013;30(4):772–80.
859 32. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated
860 alignment trimming in large-scale phylogenetic analyses. Bioinformatics.
861 2009;25(15):1972–3.
862 33. Chernomor O, Von Haeseler A, Minh BQ. Terrace aware data structure for
863 phylogenomic inference from supermatrices. Systematic biology. 2016;65(6):997–
864 1008.
865 34. Nguyen L-T, Schmidt HA, Von Haeseler A, Minh BQ. IQ-TREE: a fast and effective

866 stochastic algorithm for estimating maximum-likelihood phylogenies. Molecular
867 biology and evolution. 2015;32(1):268–74.
868 35. Anisimova M, Gil M, Dufayard J-F, Dessimoz C, Gascuel O. Survey of branch support
869 methods demonstrates accuracy, power, and robustness of fast likelihood-based
870 approximation schemes. Systematic biology. 2011;60(5):685–99.
871 36. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v4: recent updates and new
872 developments. Nucleic acids research. 2019;47(W1):W256–9.
873 37. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva E V, Zdobnov EM. BUSCO:
874 assessing genome assembly and annotation completeness with single-copy
875 orthologs. Bioinformatics. 2015;31(19):3210–2.
876 38. Boratyn GM, Camacho C, Cooper PS, Coulouris G, Fong A, Ma N, et al. BLAST: a
877 more efficient report with usability improvements. Nucleic acids research.
878 2013;41(W1):W29–33.
879 39. Slater GSC, Birney E. Automated generation of heuristics for biological sequence
880 comparison. BMC bioinformatics. 2005;6(1):31.
881 40. Rice P, Longden I, Bleasby A. EMBOSS: the European molecular biology open
882 software suite. Elsevier current trends; 2000.
883 41. Hunter JD. Matplotlib: A 2D graphics environment. Computing in science &
884 engineering. 2007;9(03):90–5.
885 42. Waskom ML. Seaborn: statistical data visualization. Journal of Open Source
886 Software. 2021;6(60):3021.
887 43. Noberini R, Restellini C, Savoia EO, Bonaldi T. Enrichment of histones from patient
888 samples for mass spectrometry-based analysis of post-translational modifications.
889 Methods. 2020;184:19–28.
890 44. He F. Bradford protein assay. Bio-protocol. 2011;45.
891 45. Noberini R, Savoia EO, Brandini S, Greco F, Marra F, Bertalot G, et al. Spatial epi-
892 proteomics enabled by histone post-translational modification analysis from low-
893 abundance clinical samples. Clinical epigenetics. 2021;13(1):1–16.

894 46. Prudhomme N, Gianetto-Hill C, Pastora R, Cheung W-F, Allen-Vercoe E, McLean
895 MD, et al. Quantitative proteomic profiling of shake flask versus bioreactor growth
896 reveals distinct responses of *Agrobacterium tumefaciens* for preparation in molecular
897 pharming. Canadian Journal of Microbiology. 2021;67(1):75–84.
898 47. Bremang M, Cuomo A, Agresta AM, Stugiewicz M, Spadotto V, Bonaldi T. Mass
899 spectrometry-based identification and characterisation of lysine and arginine
900 methylation in the human proteome. Molecular bioSystems. 2013;9(9):2231–47.
901 48. Yuan Z-F, Sidoli S, Marchione DM, Simithy J, Janssen KA, Szurgot MR, et al.
902 EpiProfile 2.0: a computational platform for processing epi-proteomics mass
903 spectrometry data. Journal of proteome research. 2018;17(7):2533–41.
904 49. Pesavento JJ, Mizzen CA, Kelleher NL. Quantitative analysis of modified proteins and
905 their positional isomers by tandem mass spectrometry: human histone H4. Analytical
906 chemistry. 2006;78(13):4271–80.
907 50. Wade PA, Pruss D, Wolffe AP. Histone acetylation: Chromatin in action. Trends in
908 Biochemical Sciences. 1997;22(4):128–32.
909 51. Binda O. On your histone mark, SET, methylate! Epigenetics. 2013;8(5):457–63.
910 52. Wolffe AP, Matzke MA. Epigenetics: Regulation through repression. Science.
911 1999;286(5439):481–6.
912 53. James TY, Stajich JE, Hittinger CT, Rokas A. Toward a Fully Resolved Fungal Tree of
913 Life. Annual Review of Microbiology. 2020;74(1):291–313.
914 54. Sun B Da, Houbraken J, Frisvad JC, Jiang XZ, Chen AJ, Samson RA. New species in
915 *Aspergillus* section *Usti* and an overview of *Aspergillus* section *Cavernicolarum*.
916 International Journal of Systematic and Evolutionary Microbiology. 2020;
917 55. Chen AJ, Frisvad JC, Sun BD, Varga J, Kocsubé S, Dijksterhuis J, et al. *Aspergillus*
918 section *Nidulantes* (formerly *Emericella*): Polyphasic taxonomy, chemistry and
919 biology. Studies in Mycology. 2016;84:1–118.
920 56. Steenwyk JL, Shen XX, Lind AL, Goldman GH, Rokas A. A robust phylogenomic time
921 tree for biotechnologically and medically important fungi in the genera *Aspergillus* and

922 *Penicillium*. mBio. 2019;10(4):1–25.
923 57. Vesth TC, Nybo JL, Theobald S, Frisvad JC, Larsen TO, Nielsen KF, et al.
924 Investigation of inter- and intraspecies variation through genome sequencing of
925 *Aspergillus* section *Nigri*. Nature Genetics. 2018;50(12):1688–95.
926 58. Kjærbølling I, Vesth T, Frisvad JC, Nybo JL, Theobald S, Kildgaard S, et al. A
927 comparative genomics study of 23 *Aspergillus* species from section *Flavi*. Nature
928 Communications. 2020;11(1).
929 59. Souza SC, Pereira VM, Moreira SI, Costa SS, Moreira GM, Mendes WO, et al.
930 *Aspergillus trisporus*: A new *Jani* section species from Brazilian soil. Current
931 Research in Environmental and Applied Mycology. 2019;9(1):175–86.
932 60. Grunstein M, Gasser SM. Epigenetics in *Saccharomyces cerevisiae*. Cold Spring
933 Harbor Perspectives in Biology. 2013;5(7).
934 61. Nielsen JC, Grijseels S, Prigent S, Ji B, Dainat J, Nielsen KF, et al. Global analysis of
935 biosynthetic gene clusters reveals vast potential of secondary metabolite production in
936 *Penicillium* species. Nature Microbiology. 2017;2(April).
937 62. Ko YH, So KK, Chun J, Kim DH. Distinct roles of two dna methyltransferases from
938 *cryphonectria parasitica* in fungal virulence, responses to hypovirus infection, and viral
939 clearance. mBio. 2021;12(1):1–16.
940 63. Jurkowski TP, Jeltsch A. On the evolutionary origin of eukaryotic DNA
941 methyltransferases and Dnmt2. PLoS ONE. 2011;6(11):1–9.
942 64. Bestor T. Structure of mammalian DNA methyltransferase as deduced from the
943 inferred amino acid sequence and direct studies of the protein. Portland Press Ltd.;
944 1988.
945 65. Okano M, Xie S, Li E. Cloning and characterization of a family of novel mammalian
946 DNA (cytosine-5) methyltransferases. Nature genetics. 1998;19(3):219–20.
947 66. Zhang Z, He C, Chen Y, Li B, Tian S. DNA methyltransferases regulate pathogenicity
948 of *Botrytis cinerea* to horticultural crops. Journal of Fungi. 2021;7(8).
949 67. Moller M, Habig M, Lorrain C, Feurtey A, Haueisen J, Fagundes WC, et al. Recent

950 loss of the Dim2 DNA methyltransferase decreases mutation rate in repeats and
951 changes evolutionary trajectory in a fungal pathogen. PLoS Genetics. 2021;17(3):1–
952 27.
953 68. Delmas V, Stokes DG, Perry RP. A mammalian DNA-binding protein that contains a
954 chromodomain and an SNF2/SWI2-like helicase domain. Proceedings of the National
955 Academy of Sciences. 1993;90(6):2414–8.
956 69. Capuano F, Mülleder M, Kok R, Blom HJ, Ralser M. Cytosine DNA methylation is
957 found in *Drosophila melanogaster* but absent in *Saccharomyces cerevisiae*,
958 *Schizosaccharomyces pombe*, and other yeast species. Analytical chemistry.
959 2014;86(8):3697–702.
960 70. Huff JT, Zilberman D. Dnmt1-independent CG methylation contributes to nucleosome
961 positioning in diverse eukaryotes. Cell. 2014;156(6):1286–97.
962 71. Proffitt JH, Davie JR, Swinton D, Hattman S. 5-Methylcytosine is not detectable in
963 *Saccharomyces cerevisiae* DNA. Molecular and cellular biology. 1984;4(5):985–8.
964 72. Catania S, Dumesic PA, Pimentel H, Nasif A, Stoddard CI, Burke JE, et al.
965 Evolutionary persistence of DNA methylation for millions of years after ancient loss of
966 a *de novo* methyltransferase. Cell. 2020;180(2):263-277.e20.
967 73. Rountree MR, Selker EU. DNA methylation and the formation of heterochromatin in
968 *Neurospora crassa*. Heredity. 2010;105(1):38–44.
969 74. Grognet P, Timpano H, Carlier F, Aït-Benkhali J, Berteaux-Lecellier V, Debuchy R, et
970 al. A RID-like putative cytosine methyltransferase homologue controls sexual
971 development in the fungus *Podospora anserina*. PLoS genetics.
972 2019;15(8):e1008086.
973 75. Madhani HD. Unbelievable but True: Epigenetics and Chromatin in Fungi. Trends in
974 Genetics. 2020;
975 76. Liu S-Y, Lin J-Q, Wu H-L, Wang C-C, Huang S-J, Luo Y-F, et al. Bisulfite sequencing
976 reveals that *Aspergillus flavus* holds a hollow in DNA methylation. PloS one.
977 2012;7(1):e30349.

978 77. Spellmon N, Holcomb J, Trescott L, Sirinupong N, Yang Z. Structure and function of
979 SET and MYND domain-containing proteins. International Journal of Molecular
980 Sciences. 2015;16(1):1406–8.
981 78. Raunser S, Magnani R, Huang Z, Houtz RL, Trievel RC, Penczek PA, et al. Rubisco
982 in complex with Rubisco large subunit methyltransferase. Proceedings of the National
983 Academy of Sciences of the United States of America. 2009;106(9):3160–5.
984 79. Trievel RC, Flynn EM, Houtz RL, Hurley JH. Mechanism of multiple lysine methylation
985 by the SET domain enzyme Rubisco LSMT. Nature Structural Biology.
986 2003;10(7):545–52.
987 80. Porras-Yakushi TR, Whitelegge JP, Clarke S. Yeast ribosomal/cytochrome c SET
988 domain methyltransferase subfamily: Identification of Rpl23ab methylation sites and
989 recognition motifs. Journal of Biological Chemistry. 2007;282(17):12368–76.
990 81. Roguev A, Schaft D, Shevchenko A, Pijnappel WWMP, Wilm M, Aasland R, et al. The
991 *Saccharomyces cerevisiae* Set1 complex includes an Ash2 homologue and
992 methylates histone 3 lysine 4. The EMBO journal. 2001;20(24):7137–48.
993 82. Zhang X, Tamaru H, Khan SI, Horton JR, Keefe LJ, Selker EU, et al. Structure of the
994 *Neurospora* SET Domain Protein DIM-5, a Histone H3 Lysine Methyltransferase. Cell.
995 2002;111:9.
996 83. Marina DB, Shankar S, Natarajan P, Finn KJ, Madhani HD. A conserved ncRNA-
997 binding protein recruits silencing factors to heterochromatin through an RNAi-
998 independent mechanism. Genes & development. 2013;27(17):1851–6.
999 84. Strahl BD, Grant PA, Briggs SD, Sun Z-W, Bone JR, Caldwell JA, et al. Set2 is a
1000 nucleosomal histone H3-selective methyltransferase that mediates transcriptional
1001 repression. Molecular and cellular biology. 2002;22(5):1298–306.
1002 85. Xiao T, Hall H, Kizer KO, Shibata Y, Hall MC, Borchers CH, et al. Phosphorylation of
1003 RNA polymerase II CTD regulates H3 methylation in yeast. Genes & development.
1004 2003;17(5):654–63.
1005 86. Li M, Phatnani HP, Guan Z, Sage H, Greenleaf AL, Zhou P. Solution structure of the

1006 Set2–Rpb1 interacting domain of human Set2 and its interaction with the
1007 hyperphosphorylated C-terminal domain of Rpb1. Proceedings of the National
1008 Academy of Sciences. 2005;102(49):17636–41.
1009 87. Bicocca VT, Ormsby T, Adhvaryu KK, Honda S, Selker EU. ASH1-catalyzed H3K36
1010 methylation drives gene repression and marks H3K27me2/3-competent chromatin.
1011 eLife. 2018;7:1–19.
1012 88. Janevska S, Baumann L, Sieber CMK, Münsterkötter M, Ulrich J, Kämper J, et al.
1013 Elucidation of the two H3K36me3 histone methyltransferases Set2 and Ash1 in
1014 *Fusarium fujikuroi* unravels their different chromosomal targets and a major impact of
1015 Ash1 on genome stability. Genetics. 2018;208:153–71.
1016 89. Wang Y, Reddy B, Thompson J, Wang H, Noma K, Yates III JR, et al. Regulation of
1017 Set9-mediated H4K20 methylation by a PWWP domain protein. Molecular cell.
1018 2009;33(4):428–37.
1019 90. Veerappan CS, Avramova Z, Moriyama EN. Evolution of SET-domain protein families
1020 in the unicellular and multicellular Ascomycota fungi. BMC Evolutionary Biology.
1021 2008;8(1):1–20.
1022 91. Gacek-Matthews A, Noble LM, Gruber C, Berger H, Sulyok M, Marcos AT, et al.
1023 KdmA, a histone H3 demethylase with bipartite function, differentially regulates
1024 primary and secondary metabolism in *Aspergillus nidulans*. Molecular Microbiology.
1025 2015;96(4):839–60.
1026 92. Bachleitner S, Sørensen JL, Gacek-Matthews A, Sulyok M, Studt L, Strauss J.
1027 Evidence of a demethylase-independent role for the H3K4-specific histone
1028 demethylases in *Aspergillus nidulans* and *Fusarium graminearum* secondary
1029 metabolism. Frontiers in Microbiology. 2019;10.
1030 93. Kramer HM, Seidl MF, Thomma BPHJ, Cook DE. Local rather than global H3K27me3
1031 dynamics associates with differential gene expression in *Verticillium dahliae*. mBio.
1032 2022 Feb;13(1):e0356621.
1033 94. Sneeringer CJ, Scott MP, Kuntz KW, Knutson SK, Pollock RM, Richon VM, et al.

1034 Coordinated activities of wild-type plus mutant EZH2 drive tumor-associated
1035 hypertrimethylation of lysine 27 on histone H3 (H3K27) in human B-cell lymphomas.
1036 Proceedings of the National Academy of Sciences. 2010;107(49):20980–5.
1037 95. Tao Y, Neppl RL, Huang Z-P, Chen J, Tang R-H, Cao R, et al. The histone
1038 methyltransferase Set7/9 promotes myoblast differentiation and myofibril assembly.
1039 Journal of Cell Biology. 2011;194(4):551–65.
1040 96. Carrozza MJ, Utley RT, Workman JL, Côté J. The diverse functions of histone
1041 acetyltransferase complexes. Trends in Genetics. 2003;19(6):321–9.
1042 97. Torchia J, Glass C, Rosenfeld MG. Co-activators and co-repressors in the integration
1043 of transcriptional responses. Current Opinion in Cell Biology. 1998;10(3):373–83.
1044 98. Sterner DE, Berger SL. Acetylation of histones and transcription-related factors.
1045 Microbiology and Molecular Biology Reviews. 2000;64(2):435–59.
1046 99. Roth SY, Denu JM, Allis CD. Histone acetyltransferases. Annual review of
1047 biochemistry. 2001;70(1):81–120.
1048 100. Kawahara T, Siegel TN, Ingram AK, Alsford S, Cross GAM, Horn D. Two essential
1049 MYST-family proteins display distinct roles in histone H4K10 acetylation and telomeric
1050 silencing in trypanosomes. Molecular microbiology. 2008;69(4):1054–68.
1051 101. Takechi S, Nakayama T. Sas3 is a histone acetyltransferase and requires a zinc
1052 finger motif. Biochemical and biophysical research communications.
1053 1999;266(2):405–10.
1054 102. Taverna SD, Ilin S, Rogers RS, Tanny JC, Lavender H, Li H, et al. Yng1 PHD Finger
1055 Binding to H3 Trimethylated at K4 Promotes NuA3 HAT Activity at K14 of H3 and
1056 Transcription at a Subset of Targeted ORFs. Molecular Cell. 2006;24(5):785–96.
1057 103. Gómez EB, Espinosa JM, Forsburg SL. *Schizosaccharomyces pombe* mst2+
1058 encodes a MYST family histone acetyltransferase that negatively regulates telomere
1059 silencing. Molecular and Cellular Biology. 2005;25(20):8887–903.
1060 104. Nugent RL, Johnsson A, Fleharty B, Gogol M, Xue-Franzén Y, Seidel C, et al.
1061 Expression profiling of *S. pombe* acetyltransferase mutants identifies redundant

1062 pathways of gene regulation. Bmc Genomics. 2010;11(1):59.
1063 105. Kim J, Park S, Lee JS. Epigenetic control of oxidative stresses by histone
1064 acetyltransferases in *Candida albicans*. Journal of Microbiology and Biotechnology.
1065 2018;28(2):181–9.
1066 106. Millar CB, Xu F, Zhang K, Grunstein M. Acetylation of H2AZ Lys 14 is associated with
1067 genome-wide gene activity in yeast. Genes and Development. 2006;20(6):711–22.
1068 107. Searle NE, Torres-Machorro AL, Pillus L. Chromatin regulation by the NuA4
1069 acetyltransferase complex is mediated by essential interactions between enhancer of
1070 polycomb (Epl1) and Esa1. Genetics. 2017;205(3):1125–37.
1071 108. Allard S, Utley RT, Savard J, Clarke A, Grant P, Brandl CJ, et al. NuA4, an essential
1072 transcription adaptor/histone H4 acetyltransferase complex containing Esa1p and the
1073 ATM-related cofactor Tra1p. EMBO Journal. 1999;18(18):5108–19.
1074 109. Wang X, Chang P, Ding J, Chen J. Distinct and redundant roles of the two MYST
1075 histone acetyltransferases Esa1 and Sas2 in cell growth and morphogenesis of
1076 *Candida albicans*. Eukaryotic Cell. 2013;12(3):438–49.
1077 110. Meijsing SH, Ehrenhofer-Murray AE. The silencing complex SAS-I links histone
1078 acetylation to the assembly of repressed chromatin by CAF-I and Asf1 in
1079 *Saccharomyces cerevisiae*. Genes and Development. 2001;15(23):3169–82.
1080 111. Balasubramanian R, Pray-Grant MG, Selleck W, Grant PA, Tan S. Role of the Ada2
1081 and Ada3 transcriptional coactivators in histone acetylation. Journal of Biological
1082 Chemistry. 2002;277(10):7989–95.
1083 112. Winkler GS, Kristjuhan A, Erdjument-Bromage H, Tempst P, Svejstrup JQ. Elongator
1084 is a histone H3 and H4 acetyltransferase important for normal histone acetylation
1085 levels *in vivo*. Proceedings of the National Academy of Sciences of the United States
1086 of America. 2002;99(6):3517–22.
1087 113. Krogan NJ, Greenblatt JF. Characterization of a Six-Subunit Holo-Elongator Complex
1088 Required for the Regulated Expression of a Group of Genes in *Saccharomyces*
1089 *cerevisiae*. Molecular and Cellular Biology. 2001;21(23):8203–12.

1090 114. Wittschieben B, Otero G, De Bizemont T, Fellows J, Erdjument-Bromage H, Ohba R,
1091 et al. A novel histone acetyltransferase is an integral subunit of elongating RNA
1092 polymerase II holoenzyme. Molecular Cell. 1999;4(1):123–8.
1093 115. Okada Y, Yamagata K, Hong K, Wakayama T, Zhang Y. A role for the elongator
1094 complex in zygotic paternal genome demethylation. Nature. 2010;463(7280):554–8.
1095 116. Booker SJ, Grove TL. Mechanistic and functional versatility of radical SAM enzymes.
1096 F1000 Biology Reports. 2010;2(1).
1097 117. Yuan Y, Yuan H, Yang G, Yun H, Zhao M, Liu Z, et al. IFN-α confers epigenetic
1098 regulation of HBV cccDNA minichromosome by modulating GCN5-mediated
1099 succinylation of histone H3K79 to clear HBV cccDNA. Clinical Epigenetics.
1100 2020;12(1):1–16.
1101 118. Kollenstart L, De Groot AJL, Janssen GMC, Cheng X, Vreeken K, Martino F, et al.
1102 Gcn5 and esa1 function as histone crotonyltransferases to regulate crotonylation-
1103 dependent transcription. Journal of Biological Chemistry. 2019;294(52):20122–34.
1104 119. Jones R. Identification and verification of lysine propionylation and butyrylation in
1105 yeast core histones using PTMap software. Bone. 2014;23(1):1–7.
1106 120. Liu B, Lin Y, Darwanto A, Song X, Xu G, Zhang K. Identification and characterization
1107 of propionylation at histone H3 lysine 23 in mammalian cells. Journal of Biological
1108 Chemistry. 2009;284(47):32288–95.
1109 121. Papanicolaou K, O'Rourke B, Foster DB. Metabolism leaves its mark on the
1110 powerhouse: recent progress in post-translational modifications of lysine in
1111 mitochondria. Frontiers in Physiology. 2014;5 JUL.
1112 122. Ekwall K. Genome-wide analysis of HDAC function. Trends in Genetics.
1113 2005;21(11):608–15.
1114 123. Bernstein BE, Tong JK, Schreiber SL. Genomewide studies of histone deacetylase
1115 function in yeast. Proceedings of the National Academy of Sciences of the United
1116 States of America. 2000;97(25):13708–13.
1117 124. Honda S, Bicocca VT, Gessaman JD, Rountree MR, Yokoyama A, Yu EY, et al. Dual

1118 chromatin recognition by the histone deacetylase complex HCHC is required for
1119 proper DNA methylation in *Neurospora crassa*. Proceedings of the National Academy
1120 of Sciences of the United States of America. 2016;113(41):E6135–44.
1121 125. Job G, Brugger C, Xu T, Lowe BR, Pfister Y, Qu C, et al. SHREC silences
1122 heterochromatin via distinct remodeling and deacetylation modules. Molecular cell.
1123 2016;62(2):207–21.
1124 126. Wu J, Suka N, Carlson M, Grunstein M. TUP1 utilizes histone H3/H2B-specific HDA1
1125 deacetylase to repress gene activity in yeast. Molecular Cell. 2001;7(1):117–26.
1126 127. Kurdistani SK, Grunstein M. Histone acetylation and deacetylation in yeast. Nature
1127 reviews Molecular cell biology. 2003;4(4):276–84.
1128 128. Li X, Pan L, Wang B, Pan L. The histone deacetylases HosA and HdaA affect the
1129 phenotype and transcriptomic and metabolic profiles of *Aspergillus niger*. Toxins.
1130 2019;11(9).
1131 129. Lee I, Oh JH, Keats Shwab E, Dagenais TRT, Andes D, Keller NP. HdaA, a class 2
1132 histone deacetylase of *Aspergillus fumigatus*, affects germination and secondary
1133 metabolite production. Fungal Genetics and Biology. 2009;46(10):782–90.
1134 130. Brosch G, Loidl P, Graessle S. Histone modifications and chromatin dynamics: A
1135 focus on filamentous fungi. FEMS Microbiology Reviews. 2008;32(3):409–39.
1136 131. Carmen AA, Griffin PR, Calaycay JR, Rundlett SE, Suka Y, Grunstein M. Yeast HOS3
1137 forms a novel trichostatin A-insensitive homodimer with intrinsic histone deacetylase
1138 activity. Proceedings of the National Academy of Sciences of the United States of
1139 America. 1999;96(22):12356–61.
1140 132. Pérez-Martínez ME, Benet M, Alepuz P, Tordera V. Nut1/Hos1 and Sas2/Rpd3
1141 control the H3 acetylation of two different sets of osmotic stress-induced genes.
1142 Epigenetics. 2020;15(3):251–71.
1143 133. Li S, Yue Z, Tanaka TU. Smc3 deacetylation by Hos1 facilitates efficient dissolution of
1144 sister chromatid cohesion during early anaphase. Molecular Cell. 2017;68(3):605-
1145 614.e4.

1146 134. Wang A, Kurdistani SK, Grunstein M. Requirement of Hos2 histone deacetylase for
1147 gene activity in yeast. Science. 2002;298(5597):1412–4.
1148 135. Suka N, Suka Y, Carmen AA, Wu J, Grunstein M. Highly specific antibodies
1149 determine histone acetylation site usage in yeast heterochromatin and euchromatin.
1150 Molecular cell. 2001;8(2):473–9.
1151 136. De Nadal E, Zapater M, Alepuz PM, Sumoy L, Mas G, Posas F. The MAPK Hog1
1152 recruits Rpd3 histone deacetylase to activate osmoresponsive genes. Nature.
1153 2004;427(6972):370–4.
1154 137. Tribus M, Bauer I, Galehr J, Rieser G, Trojer P, Brosch G, et al. A novel motif in
1155 fungal class 1 histone deacetylases is essential for growth and development of
1156 *Aspergillus*. Molecular biology of the cell. 2010;21(2):345–53.
1157 138. Kawauchi M, Nishiura M, Iwashita K. Fungus-specific sirtuin HstD coordinates
1158 secondary metabolism and development through control of LaeA. Eukaryotic Cell.
1159 2013;12(8):1087–96.
1160 139. Klose RJ, Yamane K, Bae Y, Zhang D, Erdjument-Bromage H, Tempst P, et al. The
1161 transcriptional repressor JHDM3A demethylates trimethyl histone H3 lysine 9 and
1162 lysine 36. Nature. 2006;442(7100):312–6.
1163 140. Tamaru H. Confining euchromatin/heterochromatin territory: Jumonji crosses the line.
1164 Genes and Development. 2010;24(14):1465–78.
1165 141. Isaac S, Walfridsson J, Zohar T, Lazar D, Kahan T, Ekwall K, et al. Interaction of epe1
1166 with the heterochromatin assembly pathway in *Schizosaccharomyces pombe*.
1167 Genetics. 2007;175(4):1549–60.
1168 142. Turberfield AH, Kondo T, Nakayama M, Koseki Y, King HW, Koseki H, et al. KDM2
1169 proteins constrain transcription from CpG island gene promoters independently of
1170 their histone demethylase activity. Nucleic acids research. 2019;47(17):9005–23.
1171 143. Qian S, Wang Y, Ma H, Zhang L. Expansion and functional divergence of Jumonji C-
1172 containing histone demethylases: significance of duplications in ancestral
1173 angiosperms and vertebrates. Plant Physiology. 2015;168(4):1321–37.

1174 144. Hong SH, Cho YW, Yu LR, Yu H, Veenstra TD, Ge K. Identification of JmjC domain-
1175 containing UTX and JMJD3 as histone H3 lysine 27 demethylases. Proceedings of
1176 the National Academy of Sciences of the United States of America.
1177 2007;104(47):18439–44.
1178 145. Agger K, Cloos PAC, Christensen J, Pasini D, Rose S, Rappsilber J, et al. UTX and
1179 JMJD3 are histone H3K27 demethylases involved in HOX gene regulation and
1180 development. Nature. 2007;449(7163):731–4.
1181 146. Dehé PM, Dichtl B, Schaft D, Roguev A, Pamblanco M, Lebrun R, et al. Protein
1182 interactions within the Set1 complex and their roles in the regulation of histone 3
1183 lysine 4 methylation. Journal of Biological Chemistry. 2006;281(46):35404–12.
1184 147. Lewis ZA, Adhvaryu KK, Honda S, Shiver AL, Knip M, Sack R, et al. DNA methylation
1185 and normal chromosome behavior in *Neurospora* depend on five components of a
1186 histone methyltransferase complex, DCDC. PLoS Genetics. 2010;6(11):1–11.
1187 148. Eberharter A, Sterner DE, Schieltz D, Hassan A, Yates JR, Berger SL, et al. The ADA
1188 complex is a distinct histone acetyltransferase complex in *Saccharomyces cerevisiae*.
1189 Molecular and Cellular Biology. 1999;19(10):6621–31.
1190 149. Doyon Y, Côté J. The highly conserved and multifunctional NuA4 HAT complex.
1191 Current Opinion in Genetics and Development. 2004;14(2):147–54.
1192 150. Verzijlbergen KF, van Welsem T, Sie D, Lenstra TL, Turner DJ, Holstege FCP, et al.
1193 A barcode screen for epigenetic regulators reveals a role for the NuB4/HAT-B histone
1194 acetyltransferase complex in histone turnover. PLoS Genetics. 2011;7(10):18–24.
1195 151. Sendra R, Tse C, Hansen JC. The yeast histone acetyltransferase A2 complex, but
1196 not free Gcn5p, binds stably to nucleosomal arrays. Journal of Biological Chemistry.
1197 2000;275(32):24928–34.
1198 152. Grant PA, Duggan L, Côté J, Roberts SM, Brownell JE, Candau R, et al. Yeast Gcn5
1199 functions in two multisubunit complexes to acetylate nucleosomal histones:
1200 Characterization of an ada complex and the saga (spt/ada) complex. Genes and
1201 Development. 1997;11(13):1640–50.

1202 153. Erlendson AA, Friedman S, Freitag M. A matter of scale and dimensions: chromatin of
1203 chromosome landmarks in the fungi. Microbiology spectrum. 2017;5(4):4–5.
1204 154. Buscaino A. Chromatin-mediated regulation of genome plasticity in human fungal
1205 pathogens. Genes. 2019;10(11).
1206 155. Steenwyk JL, Mead ME, Knowles SL, Raja HA, Roberts CD, Bader O, et al. Variation
1207 among biosynthetic gene clusters, secondary metabolite profiles, and cards of
1208 virulence across *Aspergillus* species. Genetics. 2020;216(2):481–97.
1209

1210 ## Figure captions

1211 **Figure 1. Conservation of DNA methyltransferase families in the *Aspergillus* genus and**
1212 **outgroup species.** The maximum-likelihood phylogeny of DNA methyltransferase domain
1213 (PF00145) found in 94 Aspergilli and 15 outgroup species was determined using IQ-TREE (37,38).
1214 The black dot on the branch indicates ultrafast bootstrap values over 95 and SH-aLRT values over
1215 80. Leaves from Aspergilli with trustworthy support are collapsed and shown as red triangles. The
1216 numbers of hits and species are separated by the pipe in the '#Homologs | #Species' column. The
1217 number in parentheses after * indicates the additional hits found by the TBLASTN search on NCBI.
1218 Additional conserved domains in DNA methyltransferase domain-containing proteins were
1219 identified using the PFAM database (https://pfam.xfam.org).
1220
1221 **Figure 2. Conservation of histone methyltransferase families in the *Aspergillus* genus and**
1222 **outgroup species.** The maximum-likelihood phylogeny of SET domains (PF00856) found in 94
1223 Aspergilli and 15 outgroup species was determined using IQ-TREE (37,38). The black dot on the
1224 branch indicates ultrafast bootstrap values over 95 and SH-aLRT values over 80. Leaves from
1225 Aspergilli with trustworthy support are collapsed and shown as red triangles. Leaves with similar
1226 protein architecture but unknown functions covering both Aspergilli and outgroup species are
1227 collapsed and shown as grey triangles. The numbers of hits and species are separated by the pipe
1228 in the '#Homologs | #Species' column. The number in parentheses after * indicates the additional
1229 hits found by the TBLASTN search on NCBI. Additional conserved domains in SET domain-
1230 containing proteins were identified using the PFAM database (https://pfam.xfam.org).
1231
1232 **Figure 3. Conservation of MOZ/SAS FAMILY domain-containing histone acetyltransferases**
1233 **in the *Aspergillus* genus and outgroup species.** The maximum-likelihood phylogeny of
1234 MOZ/SAS FAMILY domains (PF01853) found in 94 Aspergilli and 15 outgroup species was
1235 determined using IQ-TREE (37,38). The black dot on the branch indicates ultrafast bootstrap
1236 values over 95 and SH-aLRT values over 80. Leaves from Aspergilli with trustworthy support are
1237 collapsed and shown as red triangles. The numbers of hits and species are separated by the pipe
1238 in the '#Homologs | #Species' column. '*In vitro*' indicates the acetylation activity of this enzyme
1239 was only observed on the substrate *in vitro*. Additional conserved domains in MOZ/SAS FAMILY
1240 domain-containing proteins were identified using the PFAM database (https://pfam.xfam.org).
1241
1242 **Figure 4. Conservation of histone deacetylase domain-containing proteins in the**
1243 ***Aspergillus* genus and outgroup species.** The maximum-likelihood phylogeny of Histone
1244 deacetylase domains (PF00850) found in 94 Aspergilli and 15 outgroup species was determined
1245 using IQ-TREE (37,38). The black dot on the branch indicates ultrafast bootstrap values over 95
1246 and SH-aLRT values over 80. Leaves from Aspergilli with trustworthy support are collapsed and
1247 shown as red triangles. The numbers of hits and species are separated by the pipe in the
1248 '#Homologs | #Species' column. '*In vitro*' indicates the deacetylation activity of this enzyme was
1249 only observed on the substrate *in vitro*. Additional conserved domains in histone deacetylase
1250 domain-containing proteins were identified using the PFAM database (https://pfam.xfam.org).
1251
1252 **Figure 5. Presence-absence patterns of subunits that constitute characterized histone**
1253 **modification complexes.** The top row with distinct color boxes indicates outgroup species and
1254 Aspergilli sections, the order of the species from right to left is the same as in the species tree from
1255 top to bottom (Supplementary Figure 1) without *Fusarium oxysporum* and *Aspergillus kawachii*
1256 since they were excluded in the later analysis. Each row represents a subunit in the histone
1257 modification complexes, the order agrees with the summary list (Supplementary Table 4). Each
1258 row shows the presence/absence of a specific complex subunit. For each complex, the first row

1259 represents the catalytic subunit. Each box in the matrix indicates the presence (dark color) or
1260 absence (light color) of a subunit (Supplementary Table 4). The brown boxes indicate additional
1261 hits found with TBLASTN search on NCBI (Supplementary Table 5).
1262
1263 **Figure 6. Mass spectrometry unlocks the histone code in Aspergilli. A.** Schematic
1264 presentation of the in-gel digestion method used to process histones isolated in *A. nidulans*. After
1265 SDS-PAGE separation, gel bands corresponding to histone H3 and H4 are excised and chemically
1266 acylated with propionic anhydride (which derivatizes unmodified and mono-methylated lysines).
1267 Thanks to the derivatization, trypsin cutting at lysines is impaired, and histones are cleaved only
1268 at the C-terminus of arginine. Finally, digested peptides are N-terminally derivatized with phenyl
1269 isocyanate (PIC). **B.** *A. nidulans* histone H3 and H4 sequences. Cleavage sites after PRO-PIC
1270 digestion are shown. Peptides identified in this study are highlighted in light red and labeled using
1271 their starting-ending amino acids. **C.** Relative abundance of post-translational modifications
1272 detected on peptides of histone H3 and H4 from *A. nidulans* using unbiased mass spectrometry
1273 analysis. unmod: unmodified peptide; me1: monomethylation; me2: dimethylation; me3:
1274 trimethylation; ac: acetylation. The term “combinations” refers to multiple modifications occurring
1275 on the same peptide (Supplementary Table 6). **D.** Histone modifications are predicted by genomic
1276 and phylogenetic analyses, and/or proved by mass spectrometry in *A. nidulans*. Histone
1277 modifications verified by mass spectrometry are labeled as solid dots, while the ones that are only
1278 predicted by phylogenic analysis are labeled as solid dots with a red outline.

1279 **Supplementary materials**

1280 **Supplementary Figure 1. A robust species phylogeny for the fungal genus *Aspergillus*.** The
1281 species phylogeny was based on a maximum-likelihood phylogeny reconstructed by IQ-TREE with
1282 partitioned analysis using 758 BUSCO (Benchmarking Universal Single-Copy Orthologs) genes.
1283 The dot on the branch indicates ultrafast bootstrap values over 95 and SH-aLRT bootstraps values
1284 over 80. Species from the same section are labeled by background color with a section name. Two
1285 species, *Fusarium oxysporum* and *Aspergillus kawachii*, are included in the species tree, but we
1286 don't use these two in the following analyses.
1287
1288 **Supplementary Figure 2. Two known histone acyltransferase-domain containing protein**
1289 **groups are conserved in the *Aspergillus* genus and outgroup species together with many**
1290 **unknown proteins.** Maximum likelihood phylogeny of Histone acyltransferase domains (PF00583)
1291 found in 94 Aspergilli and 15 outgroup species were determined using IQ-TREE (Chernomor et al.
1292 2016; Nguyen et al. 2015). The black dot on the branch indicates ultrafast bootstrap values over
1293 95 and SH-aLRT bootstraps values over 80. Additional conserved domains in histone
1294 acyltransferase domain-containing proteins were identified using the PFAM database
1295 (https://pfam.xfam.org).
1296
1297 **Supplementary Figure 3. Five histone demethylase-domain containing protein groups are**
1298 **conserved in *Aspergillus* genus and outgroup species while KDM6B/JMJD3 is lost.**
1299 Maximum likelihood phylogeny of Jumonji C (JmjC) domain (PF02373) found in 94 Aspergilli and
1300 15 outgroup species were determined using IQ-TREE (Chernomor et al. 2016; Nguyen et al. 2015).
1301 The black dot on the branch indicates ultrafast bootstrap values over 95 and SH-aLRT bootstraps
1302 values over 80. Additional conserved domains in JmjC domain-containing proteins were identified
1303 using the PFAM database (https://pfam.xfam.org).
1304
1305 **Supplementary Table 1. List of species included in this study.**
1306
1307 **Supplementary Table 2. List of BUSCO genes used to build the species tree.**
1308
1309 **Supplementary Table 3. List of sequence names in the domain trees.**
1310
1311 **Supplementary Table 4. List of histone modifier complexes reported in the literature.**
1312
1313 **Supplementary Table 5. Identification of missing chromatin modifiers in the outgroup**
1314 **species using TBLASTN search**
1315
1316 **Supplementary Table 6. Mass spectrometry result for *A. nidulans*.**
1317
1318 **Supplementary Data 1. Fasta, trimmed alignment, and maximum-likelihood phylogenetic**
1319 **tree files associated with Figure 1-4 and Supplementary Figure 2-3.**
1320
1321 **Supplementary Data 2. Fasta, trimmed alignment, and maximum-likelihood phylogenetic**
1322 **tree files associated with Figure 5.**