

## Article

# Phylogenomic and Evolutionary Analyses Reveal Diversifications of SET-Domain Proteins in Fungi

Guoqing Ding <sup>†</sup>, Liqiu Shang <sup>†</sup>, Wenliang Zhou <sup>†</sup>, Siyi Lu, Zong Zhou, Xinyi Huang and Juan Li <sup>\*</sup>

State Key Laboratory for Conservation and Utilization of Bio-Resources in Yunnan, Yunnan University, Kunming 650091, China

<sup>\*</sup> Correspondence: juanli@ynu.edu.cn<sup>†</sup> These authors contributed equally to this work.

**Abstract:** In recent years, many publications have established histone lysine methylation as a central epigenetic modification in the regulation of chromatin and transcription. The histone lysine methyltransferases contain a conserved SET domain and are widely distributed in various organisms. However, a comprehensive study on the origin and diversification of the SET-domain-containing genes in fungi has not been conducted. In this study, a total of 3816 SET-domain-containing genes, which were identified and characterized using HmmsSearch from 229 whole genomes sequenced fungal species, were used to ascertain their evolution and diversification in fungi. Using the CLANS program, all the SET-domain-containing genes were grouped into three main clusters, and each cluster contains several groups. Domain organization analysis showed that genes belonging to the same group have similar sequence structures. In contrast, different groups process domain organizations or locations differently, suggesting the SET-domain-containing genes belonging to different groups may have obtained distinctive regulatory mechanisms during their evolution. These genes that conduct the histone methylations (such as H3K4me, H3K9me, H3K27me, H4K20me, H3K36me) are mainly grouped into Cluster 1 while the other genes grouped into Clusters 2 and 3 are still functionally undetermined. Our results also showed that numerous gene duplication and loss events have happened during the evolution of those fungal SET-domain-containing proteins. Our results provide novel insights into the roles of SET-domain genes in fungal evolution and pave a fundamental path to further understanding the epigenetic basis of gene regulation in fungi.

**Keywords:** histone methyltransferases; Ascomycota; gene duplication and loss; methylation; epigenetic modification; collinearity relationships



**Citation:** Ding, G.; Shang, L.; Zhou, W.; Lu, S.; Zhou, Z.; Huang, X.; Li, J. Phylogenomic and Evolutionary Analyses Reveal Diversifications of SET-Domain Proteins in Fungi. *J. Fungi* **2022**, *8*, 1159. <https://doi.org/10.3390/jof8111159>

Academic Editor: Anna Muszewska

Received: 22 September 2022

Accepted: 28 October 2022

Published: 2 November 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Histones are subject to numerous post-translational modifications, including acetylation, methylation, phosphorylation, ubiquitylation, sumoylation, and ADP-ribosylation [1]. In recent years, histone modifications, especially histone lysine methylation, have been demonstrated as a significant epigenetic modification in the regulation of gene expression and chromatin organization [2–4]. Histone lysine methylation is usually catalyzed by the histone lysine methyltransferases (HMTs) with an evolutionarily conserved motif named SET domain, with the exception of Dot1, which mediates the H3K79 methylation [5,6]. The SET domain stands for “Suppressor of variegation, Enhancer of zeste and Trithorax”, which are the three genes initially derived from *Drosophila melanogaster* [7,8]. At present, SET-domain-containing genes have been found in various organisms ranging from Prokaryotes to Eukaryotes [9]. The SET domain is mainly comprised of approximately 130–150 amino acid residues, which are capable of catalyzing mono-, di-, or tri-methylation of their lysine substrates, and lysine residues in different histones that can be methylated by the SET domain including K4 (K is the abbreviation for lysine), K9, K27, K36, and K79 in histone H3; K20 in histone H4; K59 in the globular domain of histone H4; and K26 in histone H1B [10].

The methylation of lysine residues in histone protein can serve as an epigenetic marker to regulate gene expressions at the post-translational level. Diverse biological mechanisms such as transcriptional activation and repression, enhancer function, mRNA splicing, and DNA replication have all been associated with SET-domain-containing genes [3,11,12].

In fungi, the modification of histone N-terminal tails is overall conserved. In general, the histone modification landscape usually generates as transcriptionally active or transcriptionally repressive. The cross-talk between various modification machineries formed interconnected positive and negative feedback loops that control gene regulation [13]. It has been frequently asserted that H3K9me is associated with the silent regions of both euchromatin and heterochromatin, and responsible for transcriptional repression [14–16], while H3K4me is related exclusively to actively transcribed genes [17]. In addition, H3K36me, which is catalyzed by SET2, is also associated with gene activity [18]. Moreover, trimethylation of histone H3 lysine 27 (H3K27me3) has been recognized as important for differentiation and development in fungi [19,20]. Aside from its relatively well-understood role in transcriptional repression, accumulating evidence suggests that H3K27 methylation has an important role in controlling the balance between maintenance and generation of novelty in fungal genomes [21]. Those histone lysine modifications have been extensively investigated to play vital roles in fungal development and various biological processes. For instance, in *Neurospora crassa*, a histone H3K9 methyltransferase Dim-5 was reported to be required for normal growth and full fertility of *N. crassa* [22]. NcSet2, the sole lysine histone methyltransferase responsible for the methylation of H3K36, is essential for normal growth and development in *N. crassa* [23]. ScSet4 has been characterized as a regulated chromatin factor required for controlling gene expression in response to stress in *Saccharomyces cerevisiae* [24]. MoSet1, a histone H3K4 methyltransferase in the phytopathogenic fungus *Magnaporthe oryzae*, has been confirmed to regulate global gene expression during infection-related morphogenesis [25]. FvSet1 was also found to play an important role in vegetative growth, virulence, and environmental stress responses in the plant pathogen *Fusarium verticillioides* [26]. All these studies provided strong evidence that SET-domain-containing methyltransferases play crucial roles in regulating a variety of biological processes in fungi, including growth, differentiation, development, metabolic processes, and environmental adaptation [27,28].

Although the importance of SET-domain containing methyltransferases in fungi has been confirmed, the SET-domain genes have only been identified in a limited number of representative fungal species, primarily in model filamentous fungi such as *N. crassa* and *M. oryzae* and in model yeasts *S. cerevisiae* and *Schizosaccharomyces pombe* [4,29]. Similarly, studies related to the evolution of SET-domain genes in fungi have been limited to a few fungal species [30]. Thus, the pattern of conservation and diversification of SET genes among fungal species remains largely unknown. Therefore, in this study, we conducted a large-scale phylogenomic survey of SET-domain genes from 229 fungal species, which cover all the phyla within the fungal kingdom. We aimed to identify the evolutionary dynamics and functional divergences of SET-domain genes among diverse fungal species. Our results provide a comprehensive understanding of the evolution of SET-domain genes in fungi.

## 2. Materials and Methods

### 2.1. Identification and Homologous Searches of SET-Domain Containing Genes in Fungi

To find all the putative SET-domain containing sequences from fungi, the HMM profile SET domain (PF00856; <http://pfam.xfam.org/family/PF00856>, accessed on 1 August 2022) was downloaded and used as a query with E values  $< 10^{-5}$  in the program and HMMSEARCH from the HMMER package (<http://hmmer.wustl.edu/>, accessed on 4 August 2022) to search the homologous proteins in 229 fungal genome sequences. All 229 fungal genomes were available from the Mycocosm of Joint Genome Institute (<https://mycocosm.jgi.doe.gov/mycocosm/home>, accessed on 27 July 2022), Fungal Genome Research (<http://fungalgenomes.org/>, accessed on 27 July 2022), and the NCBI

database. This is the first analysis of SET-domain-containing genes covering such a range of fungal species with different lifestyles.

### 2.2. Distribution, Structure Identification of SET-Domain Containing Genes in Fungi

A Java-based software program, CLANS, was used to elucidate the relationships between and within subfamilies of the Hmsearch results [31]. After the identifications of groups, the conserved functional domain structures of some representative sequences of each group were predicted using the InterProScan 5.0 (<http://www.ebi.ac.uk/interpro/>, accessed on 1 August 2022, Hinxton, UK) and MEME suite (<https://meme-suite.org/meme/>, accessed on 1 August 2022, Seattle, WA, USA) with default parameter settings. We combined both the two analysis results to figure out the domain structures of the SET-domain containing proteins in each cluster.

### 2.3. Phylogenetic Analysis of SET-Domain Containing Genes in Fungi

Before phylogenetic analysis, MUSCLE v3.7 was used to generate protein alignment with default settings [32]. Then, the aligned sequences for each group were analyzed using two different tree construction methods: Neighbor-joining (NJ) analysis and Maximum likelihood (ML) analysis. For NJ analysis [33], the data were analyzed using MEGA X (Hachioji, Japan) [34] using bootstrap analysis (BS) with 1000 replicates. For ML analysis, the best-fit models of protein evolution for each subfamily were first estimated using the program ProtTest [35], then the recommended models and parameters were used for ML analysis with FastTree [36].

### 2.4. Analysis of Family Size over Evolutionary Time

In order to identify the expansions and contractions over evolutionary time about the five groups (Group 2, 4, 5, 6, and 12), which conduct the histone methylations of H3K27me, H3K4me, H3K9me, H3K36me, H4K20me, respectively, we chosen 20 representative fungal species to compare their family size for each group using the program Computational Analysis of gene Family Evolution v5 that use phylogenomics and family sizes to estimate the timing of gene family evolution (CAFE5; <https://github.com/hahnlab/>, accessed on 1 August 2022, CAFE5, Gothenburg, Sweden) [37,38]. Ultrametric phylogenies were generated using Orthofinder v2.3.11 (Oxford, UK) [39] and TimeTree was generated by the r8s [40].

### 2.5. Collinearity Analysis of SET-Domain Genes in Specific Fungal Species

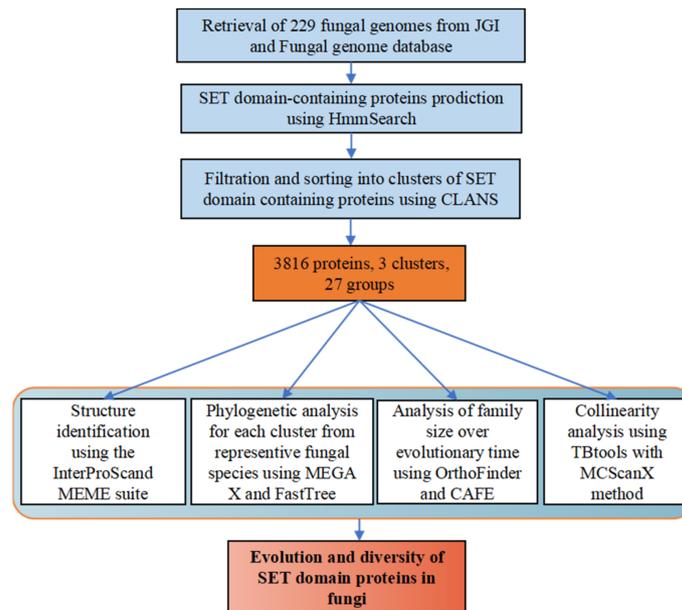
To understand the collinearity relationships of the SET-domain gene family, five extensively studied Ascomycota fungal species, including three Eurotiomycetes fungi *Aspergillus fumigatus*, *Aspergillus niger*, and *Penicillium chrysogenum*, and one Sordariomycetes fungus *N. crassa* and one yeast species from Taphrinomycotina (*S. pombe*), with their genomes chromosomal-level assembled, were investigated using the TBtools (Guangdong, China) with MCScanX (Athens, GA, USA) [41,42]. In addition, using TBtools [41], we also visualized the chromosomal distribution of the SET-domain genes in the *A. fumigatus* genome.

## 3. Results and Discussions

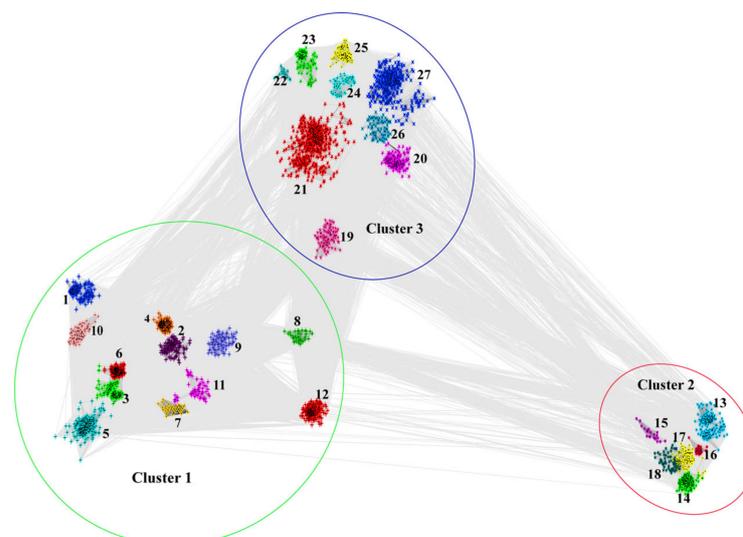
### 3.1. Distribution of SET-Domain-Containing Genes in Fungi

We analyzed the SET-domain-containing genes in the fungi according to the flow chart shown in Figure 1. Firstly, we downloaded the HMM profile of the SET-domain from <http://pfam.xfam.org/family/PF00856> and used this profile as a query model in the program HMMSEARCH. A total of 4172 SET-domain proteins were identified from 229 fungal species with different lifestyles with E values  $<10^{-5}$ . To elucidate the relationships among these SET-domain sequences, a clustering analysis that relied on the sequence similarities of proteins was conducted using the CLANS program [31]. CLANS is a Java-based software program that visualizes pairwise sequence similarities in either two-dimensional or three-dimensional space [31]. In cases where a sequence showed low sequence similarities with other sequences, it would be set as a separate point in the CLANS program. Thus, those sequences that showed

low sequence similarities with other sequences would be formed as the separated points in the CLANS program [31]. As seen in Figure S1, several sequences cannot group together with any other genes. For example, the *Jumonji C (JmjC)* genes identified from a few Pezizomycetes species, which have been identified to play roles in the modulation of histone methylation marks [43], showed low sequence similarities with each other and have been set as separated points in the CLANS program. Thus, those sequences, including JmjC-SET genes, are not considered in the following analysis because they cannot be well-grouped into any families. Finally, a total of 3816 SET-domain genes were used to conduct subsequent cluster analysis in this study, after removing the separated points in the CLANS program (Figure 2).



**Figure 1.** Pipeline schematic of the SET domain proteins in fungi. Predicted protein sets for a given species are fed into the pipeline. In the first step, the protein sets are searched with the SET-domain (PF00856) HMM profile. Next, candidate SET-domain genes are filtered and sorted into families by CLANS program. Then those identified 3816 SET-domain genes were subsequently analyzed with structure analyses, phylogenetic analysis, gene family size analysis, and collinearity analysis.



**Figure 2.** CLANS clustering of 3816 SET domain sequences obtained from 229 whole-genome sequenced fungal species. Three Clusters and 27 groups among the 3816 SET domain sequences were identified. Those 27 groups within these three Clusters were marked with different colors and shapes.

The taxonomic information and the identified number of SET-domain-containing genes from each species and group are presented in Table S1. Specifically, those 3816 SET-domain genes from 229 fungal species can mainly be classified into three distinct clusters based on amino acid sequence similarities (Cluster 1, 2, and 3, respectively), and 12, 6, and 9 groups can be observed in each of the three clusters (Figure 2). Remarkable variations in the number of SET-domain-containing genes were observed among different fungal species (Table S1). A considerable variation in numbers was found within the Ascomycota, Basidiomycota, Chytridiomycota, Mucoromycota, Zoopagomycota, and Microsporidia fungi, ranging from 0 to 43 (Table S1). The Basidiomycota fungus *Coprinopsis cinerea* contains 43 SET-domain sequences, whereas no specific SET-domain gene was identified from several Microsporidia fungi. The explanation for why those ancestral Microsporidia fungi did not contain any SET-domain genes may be associated with their genome size. As we can see in Table S1, all the genome sizes of Microsporidia fungal species were less than 10 Mb, suggesting that these species have only retained those essential functional genes to maintain their lives. In addition, scientists have found that the yeast *Cryptococcus neoformans* has lost the methylase gene between 50–150 mya, but the methylation in *C. neoformans* has persisted because an epigenome has been propagated for >50 million years through a process analogous to the Darwinian evolution of the genome [44,45]. Thus, although it is still hard to explain how the lysine methyltransferase activity could work in the fungi without the SET-domain gene, we supposed that there may be some uncovered mechanisms for those fungi to activate their lysine methyltransferase. Among the Pezizomycotina fungi of Ascomycota, the majority contain more than 10 SET-domain-encoding genes, and the Leotiomyces fungus *Acephala macrosclerotiorum* contains 41 SET-domain genes, which is the largest number identified in Ascomycota fungi. For the yeast-like fungi of Ascomycota fungi, the gene number of most of the Saccharomycotina fungi is less than 10. In contrast, the total number of the SET-domain genes in each Taphrinomycotina fungi is larger than 10, suggesting that there may have been a constraint on the number of SET-domain genes after the split of Saccharomycotina and Taphrinomycotina fungi. These variations imply that gene duplication and loss may have played significant roles during the evolution of the fungal SET-domain-containing genes.

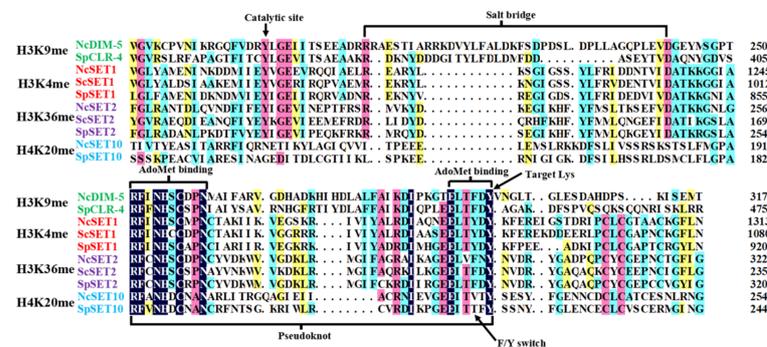
### 3.2. Conserved Motifs and Residues in SET Domain Proteins of Fungi

A protein sequence alignment of the SET domains from several representative histone lysine methyltransferases (HKMT) grouped according to their histone-lysine specificity showed that they contain conserved short conserved binding and catalytic motifs. According to the solved structures of the SET domain, we found that the SET-domain sequences contained the two most-conserved sequence motifs (RFINHXCPN and ELXFDY, see Figure 3) binding to S-adenosylmethionine (AdoMet) and the target lysine of the SET domain. Moreover, the residues of the catalytic site; an intra-molecular interacting salt bridge; and a F/Y switch controlling whether the product is a mono-, di-, or tri-methylated histone were also identified. Previous studies supposed that the two conserved motifs binding to AdoMet were brought together by the “pesudoknot” fold to form an active site in a location immediately next to the pocket where the methyl donor binds to the peptide-binding cleft. Most notably, the substrate-binding clefts are connected through a deep channel that runs through the core of the SET domain and permits the transfer of the methyl group from AdoMet to the -amino group of the lysine substrate.

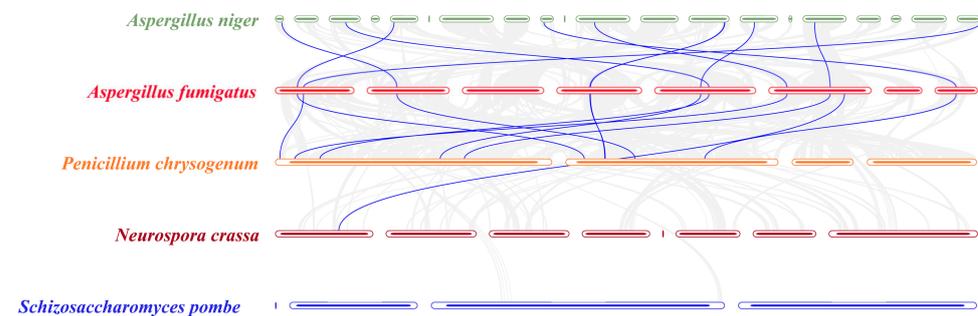
### 3.3. Collinearity Analysis of SET-Domain Genes in Specific Fungal Species

Subsequently, using a Toolkit for Biologists integrating various biological data-handling tools (TBtools) with the Multiple Collinearity Scan toolkit (MCScanX) method [41,42], we explored the collinearity relationships of the SET-domain genes among five chromosomal-level assembled Ascomycota species, including *A. fumigatus*, *Aspergillus niger*, *P. chrysogenum*, *N. crassa*, and *S. pombe* (Figure 4). Collinearity is a more specific form of synteny, which requires conserved gene order [42]. In our analysis, a total of 5895 gene pairs showed

collinearity relationships between *A. fumigatus* and *A. niger*. In addition, 5642 gene pairs showed collinearity relationships between *A. fumigatus* and *P. chrysogenum*. Moreover, 13 SET-domain genes from *A. fumigatus* have collinearity relationships with *A. niger*, and *P. chrysogenum* (Figure 4). The great collinearity relationships among *A. fumigatus*, *A. niger*, and *P. chrysogenum* are consistent with their closed phylogenetic relationships. However, although 455 gene pairs showed collinearity relationships between *A. fumigatus* and *N. crassa*, only one SET-domain gene pair was identified from these two species (Figure 4). Moreover, no SET-domain gene pairs were identified between *A. fumigatus* and *S. pombe*, which can be explained by their far phylogenetic relationships. Using *A. fumigatus* genome annotation information and TBtools [41], we also visualized the chromosomal distribution of the SET-domain-containing genes. Results from Figure S2 indicate that the 13 SET-domain genes derived from *A. fumigatus* are unevenly distributed on the eight chromosomes, and the number of genes on each chromosome is irrelevant to chromosome size. Moreover, no segmental duplication events occurred among all the identified SET-domain genes in *A. fumigatus* (Figure S2).



**Figure 3.** Alignment of several representative SET-domain-containing histone lysine methyltransferases (HKMT) grouped according to their histone-lysine specificity from *N. crassa* (Nc), *S. cerevisiae* (Sc), and *S. pombe* (Sp). Those HKMTs all belong to Cluster 1 in Figure 2 with NcDIM-5 and SpCLR-4 belonging to Group 5, NcSET1, ScSET1, and SpSET1 belonging to Group 4, NcSET2, ScSET2, and SpSET2 belonging to Group 6, and NcSET10 and SpSET1 belonging to Group 12. The residues of the AdoMet, target lysine, catalytic site, the structural pseudoknot, an intra-molecular interacting salt bridge, and a F/Y switch controlling whether the product is a mono-, di- or tri-methylated histone are indicated.



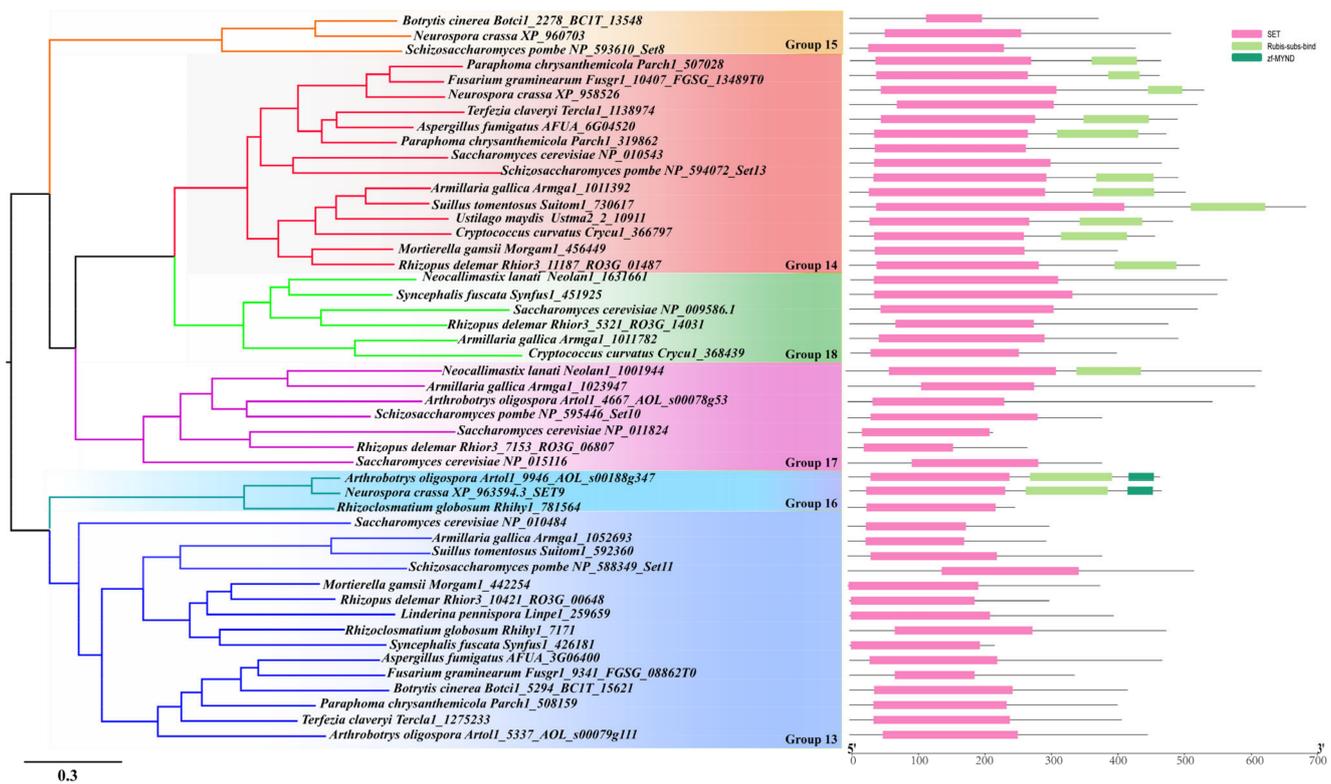
**Figure 4.** Collinearity analysis of the SET-domain genes from five species. A total of 13 SET-domain genes derived from *A. fumigatus* (AFUA\_1G03000, AFUA\_1G11090, AFUA\_2G08510, AFUA\_2G08775, AFUA\_2G10080, AFUA\_3G06400, AFUA\_3G06480, AFUA\_4G09180, AFUA\_5G06000, AFUA\_5G12710, AFUA\_6G04520, AFUA\_6G06335, and AFUA\_7G04410) to show the collinearity relationships with the orthologous from other four species. These 13 SET-domain genes were distributed into all the three clusters from Figure 2. The gray lines indicate gene blocks in *A. fumigatus* that are orthologous to the other genomes. The blue lines delineate the syntenic SET-domain gene pairs.

### 3.4. Characterization of Each Identified SET-Domain-Containing Gene in Fungi

To study the evolution of the SET-domain genes in fungi, we conducted a phylogenetic analysis based on the SET-domain genes identified from 20 representative fungal species according to their taxonomy classification. Phylogenetic trees were constructed based on the three distinct clusters of these SET-domain genes (Figures 5–7). To show the relationships with SET-domain genes identified in previous studies and attempt to make the names of the SET genes less confusing, we used a nomenclature system according to previous research on the SET domain proteins in fungi. In addition to the families reported before, we also found some novel groups identified here.



**Figure 5.** Phylogenetic analysis and domain organization of the Cluster 1 SET domain sequences from 20 representative 20 fungal species. The protein domains were labeled with different colors based on the domain type. Domain organization of SET-domain-containing proteins was detected by the InterProScan 5.0 (<http://www.ebi.ac.uk/interpro/>) and MEME suite (<https://meme-suite.org/meme/>) with default parameter settings. Phylogenetic analyses were performed using NJ with MEGA X and ML analysis with FastTree.



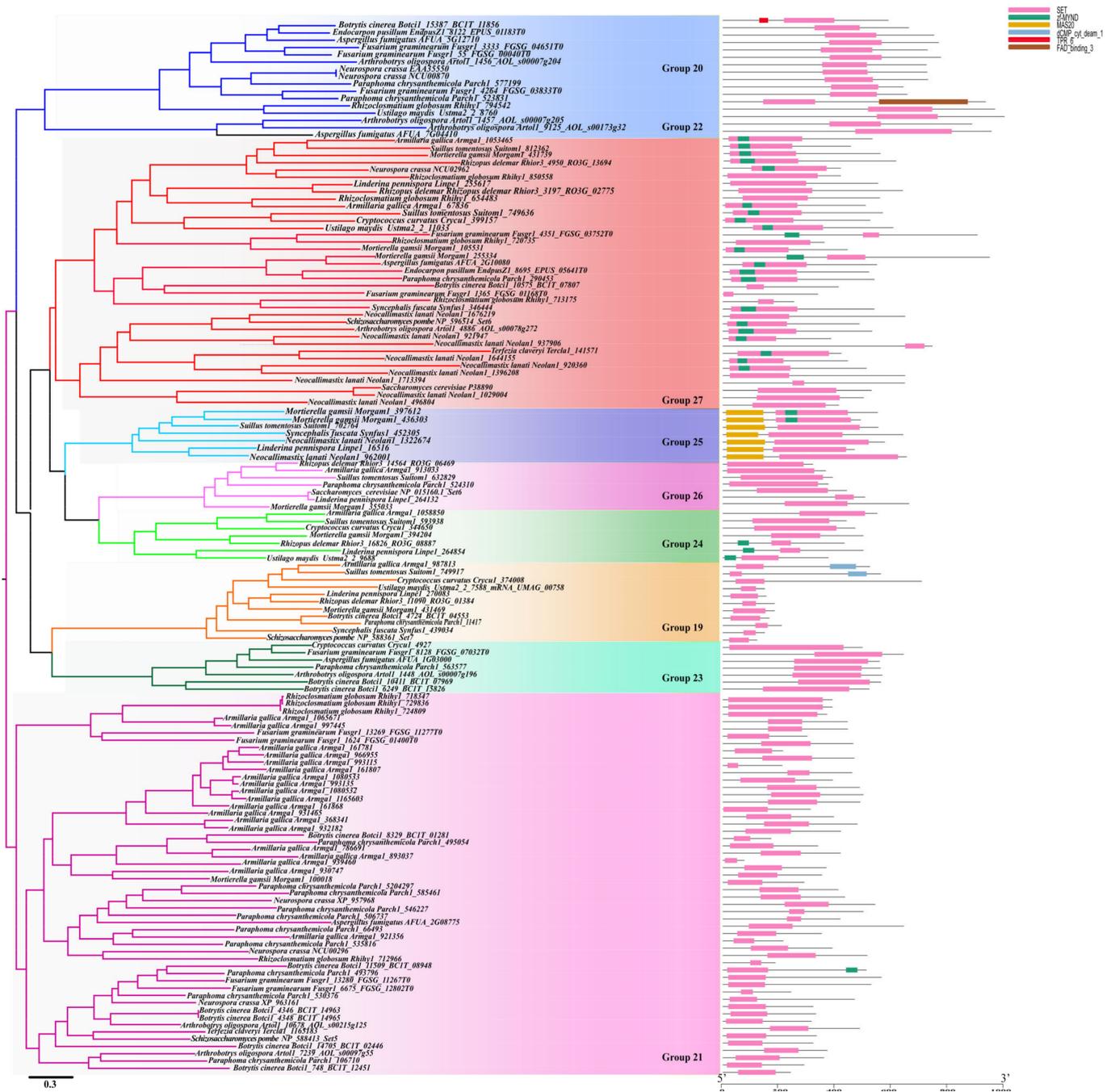
**Figure 6.** Phylogenetic analysis and domain organization of the Cluster 2 SET domain sequences from 20 representative 20 fungal species. The protein domains were labeled with different colors based on the domain type. Domain organization of SET-domain-containing proteins was detected by InterProScan 5.0 (<http://www.ebi.ac.uk/interpro/>) and MEME suite (<https://meme-suite.org/meme/>) with default parameter settings. Phylogenetic analyses were performed using NJ with MEGA X and ML analysis with FastTree.

### 3.4.1. Cluster 1

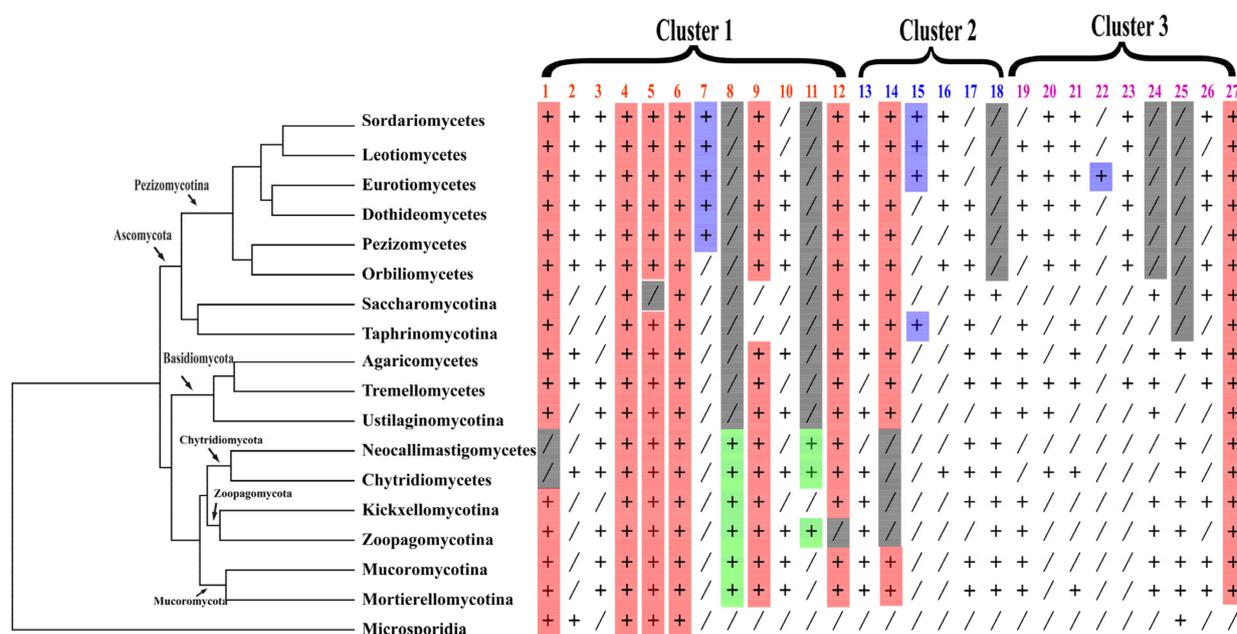
Cluster 1 was comprised of 12 distinct groups. The genes in this cluster are known to catalyze di- and tri-methylation of H3K27, H3K4, H3K9, H3K36, and H4K20 (Group 2, 4, 5, 6, and 12, respectively) (Figure 5). Phylogenetic analysis of Cluster 1 genes identified from 20 representative fungal species showed that these 12 groups clustered into distinct clades with each other (Figure 5). Also, the functional domain structures of each group showed that the genes within the same group shared similar motif features while different group genes had different motif elements.

#### The SET3/4 Family (Group 1)

This group of proteins was commonly present in most of the fungal lineages except for Chytridiomycota (Figure 8). Structure analysis showed that proteins belonging to this group contain the conserved SET domain and a conserved PHD (plant homeodomain) finger domain at their N-terminal (Figure 5). Previously identified *ScSet3* (*Sc-NP\_012430*), *ScSet4* (*Sc-NP\_012954*) from *S. cerevisiae*, *NcSet4* (*Nc-XP\_957466*) from *N. crassa*, and *SpSet3* (*Sp-NP\_594837*) from *S. pombe* were clustered in this group (Table 1 and Figure 5). A previous study suggested that the *ScSet3* and *NcSet4* were components of a conserved histone deacetylase complex (HDAC), which can bind histone H3 dimethylation at lysine 4 (H3K4me2) to mediate deacetylation of histones in 5' transcribed regions via their PHD finger [46]. In yeast, *ScSet3* is involved in the meiosis-specific repression of sporulation genes [47] and is required for cellular protection against oxidative stress [24]. *C. albicans* *Set3C* is required for efficient biofilm production and drug resistance [48].



**Figure 7.** Phylogenetic analysis and domain organization of the Cluster 3 SET domain sequences from 20 representative 20 fungal species. The protein domains were labeled with different colors based on the domain type. Domain organization of SET domain containing proteins were detected by the InterProScan 5.0 (<http://www.ebi.ac.uk/interpro/>) and MEME suite (<https://meme-suite.org/meme/>) with default parameter settings. Phylogenetic analyses were performed using NJ with MEGA X and ML analysis with FastTree.



**Figure 8.** Diversification of SET-domain genes in fungi. Distribution pattern of SET genes in different fungal taxonomic groups. Those 3816 SET-domain genes from 229 fungal species can mainly be classified into three distinct clusters (Cluster 1, 2, and 3, respectively), and twelve, six, and nine groups can be observed in each of the three clusters. One representative is elected for each order. “Plus” indicates the presence of SET genes, and “Slash” indicates the absence of SET genes.

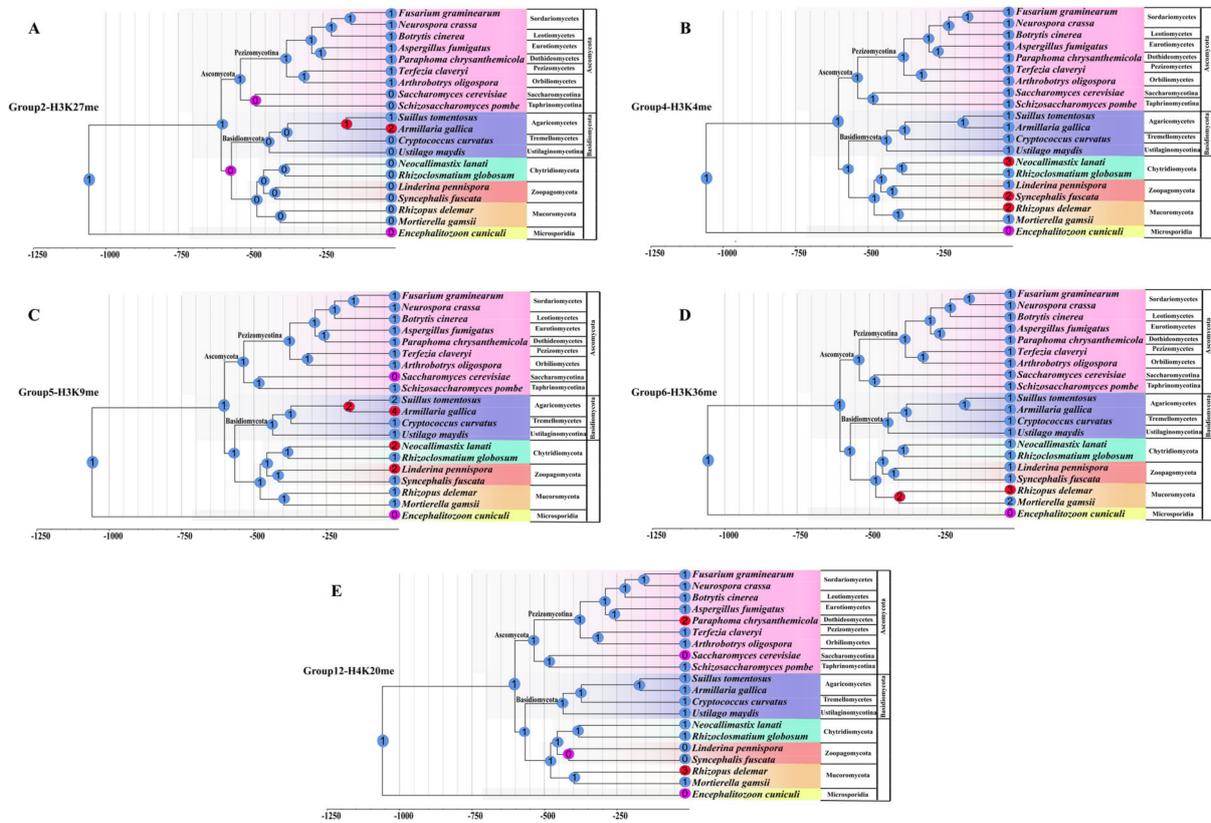
### The Enhancer of Zeste (EZ) Family (Group 2)

This group’s proteins were also named the Enhancer of Zeste (Ez) family because they are components of the Polycomb PRC2 complexes, which play as a transcriptional suppressor for genes. The structures of the E(z)-type group genes include a CXC domain and a SET domain, whereas the *Nc-XP\_965043* from *N. crassa* further processes another EZH2 domain (Figure 5 and Table 1). Earlier studies reported the absence of this group’s genes in the Saccharomycotina and the *S. pombe* genomes. However, in our analysis, we found this group’s genes from most of the Pezizomycotina fungi and identified several genes from Basidiomycota and the Chytridiomycetes and Microsporidia fungi. In contrast, no E(z)-type gene was found in Mucoromycota, Zoopagomycota, and the yeast-like fungi employed in our study (Table S1 and Figure 5). CAFE results from 20 fungal species showed that this group’s genes may have been lost before the split of Saccharomycotina and Taphrinomycotina fungi. Moreover, this group’s genes may have been lost by the ancestor of Basidiomycota but the Agaricomycetes fungi gained another gene copy during their evolution (Figure 9). Also, gene duplication has occurred in the fungus *Armillaria gallica*. Furthermore, we also found that several Eurotiomycetes fungi have lost the E(z)-type gene during their evolution (Figure 9). At present, the E(z) methylases have been reported to specifically mediate the H3K27me3 within Polycomb PRC2 complexes [9,49]. Thus, the Group 2 genes are predicted to account for the deposition of H3K27me3.

### The ASH1-Like Family (Group 3)

The genes in Group 3 were defined as an ASH1-like family because their two recognizable structural motifs (cysteine-rich peptide AWS, and SET domain) show similar motif structures to the ASH1 genes of *Drosophila melanogaster* but lack AThooks and other motifs found in the *Drosophila* ASH1 protein (Figure 5). In our study, this group’s proteins were not identified from the yeast-like fungi, the Agaricomycetes, Kickxellomycotina, and Microsporidia fungi (Table S1 and Figure 5). Interestingly, although this group’s genes identified from other fungal species showed a similar domain structure to those genes from the Pezizomycotina fungi, their sequence lengths were significantly different. How-

ever, although the epigenetic activator *ASH1* of *Drosophila* has been characterized as a multi-catalytic histone methyltransferase (HMTase) methylating H3K4, H3K9, H4K20, and H3K36 [50,51], the substrates and functions of the *ASH1*-related proteins in fungi remain unknown.



**Figure 9.** The evolution of five groups SET-domain genes analyzed with CAFE. Each tip of the tree represents a species. The numbers present on each node correspond to the number of SET-domain genes predicted by CAFE. Blue = No change; Red = Expansion; Purple = Contraction.

The SET1 Family (Group 4)

This group of genes commonly exists in all the fungal lineages examined in this study (Table S1 and Figure 5), suggesting their importance in fungi. Besides the conserved N-SET and SET domains at the C-terminus, this group’s proteins also contain a SET-associated (SET-assoc) domain and an RNA recognition motif (RRM) domain at their upstream regions (Figure 5). The RRM domain has been considered to affect the level of K4 methylation [27,52]. CAFE results showed that this group of genes conserved existed in the Ascomycota and Basidiomycota fungi with only one gene copy (Figure 9). However, in the Chytridiomycota fungus *Neoallimastix lanate*, the Zoopagomycotina fungus *Syncephalis fuscata*, and the Mucoromycota fungus *Rhizopus delemar*, this group genes may have suffered expansion (Figure 9). It has been confirmed that the *ScSet1* (*Sc-NP\_011987*) is responsible for the overall chromatin mono-, di-, and tri-methylation of histone H3-lysine 4 (H3K4) in *S. cerevisiae*. Thus, we predicted that this group’s genes might conduct the function of H3K4me in fungi.

The SU(Var)3-9 Family (Group 5)

*NcDim-5* (*Nc-EAA28243*) of *Neurospora crassa* and *SpClr4* (*Sp-NP\_595186*) of fission yeast *S. pombe*, which have been characterized as Su(var)3-9 family genes, are clustered in this group (Tables S1 and 1). Thus, group 5 is also named the Su(var)3-9 family. Previous studies showed that Su(var)3-9 family proteins have HKMT activity to conduct the methylation of H3K9 depending on their functional regions in pre-SET [14,53]. The pre-SET domain,

which is involved in the structural stability of the SET domain, contains nine invariant cysteine residues, coordinating three zinc ions [54]. Thus, the genes within this group may all have the potential functions to perform the methylation of H3K9. This group's genes contain a pre-SET motif located upstream of the SET domain (Figure 5). Moreover, several genes such as *SpClr4* (*Sp-NP\_595186*) from *S. pombe* and *Ao-AOL\_s00076g661* in this group also contain an additional chromodomain at their N-terminus. The Su(var)3-9 family genes in this group were found in the early diverging fungal lineages (Table S1). However, in the yeast-like fungi, it is interesting that this group's genes were totally lost in the Saccharomycotina fungi but still presented in the Taphrinomycotina fungi (Table S1). In Ascomycota, almost all filamentous fungi contain one copy of this group's genes. CAFE results showed that these group genes have been lost in the Saccharomycotina fungi, while gene duplication events have occurred in the Chytridiomycota fungus *Neoallimastix lanate*, the Zoopagomycotina fungus *Linderina pennisporea*, and the Agaricomycetes fungus *Suillus tomentosus* and *Armillaria gallica* (Figure 9).

#### The SET2 Family (Group 6)

The representative genes of the Group 6 include the methyltransferase *ScSET2* (*Sc-NP\_012367*) from *S. cerevisiae*, *SpSET2* (*Sp-NP\_594980*) from *S. pombe*, *NcSET2* (*Nc-XP\_957740*) from *N. crassa* (Table 1). Thus, this group is also named as SET2 family. This group's genes are commonly found in all fungal lineages, with only one gene identified from one of the Microsporidia fungus *Mitosporidium daphniae* while most of the other fungal species employed in this study contain only one gene copy (Table S1, Figures 5 and 8). It is interesting that our CAFE result showed that the gene expansion only occurred in the Mucoromycota lineage while the other fungal lineages conserved keep this group's genes, suggesting the importance of this group's genes (Figure 9). The structure motifs of this group's proteins show similarity to those of the ASH1-like family (Group 3) proteins, which both process the AWS (Associated With SET), and SET domains. However, the two domains in the ASH1-like family were located at the C-terminal, while those domains in the SET2 family were located at the N-terminal (Figure 5). In addition, this group's proteins also carry an SRI motif (Figure 5), which has been identified to mediate RNA polymerase II interaction [55,56]. In addition, several genes in this group contain another WW-peptide motif and Med26 motif (Figure 5). It is known that the SET2 methyltransferases in the yeasts *S. cerevisiae* and *S. pombe* are responsible for mono-, di-, and tri-methylation of H3K36 [55,57]. The orthologue SET2, which methylates H3K36 in *N. crassa*, is required for normal growth and expression of genes in both the asexual and sexual differentiation pathways [23].

#### The Group 7-11 Family

Group 7 consists of 31 genes from several fungi belonging to Pezizomycotina fungi except for the three Orbiliomycetes fungal species, suggesting this group's genes are Pezizomycotina-specific (Table S1). SET domain, which is located at the N-terminal is the only recognizable conserved structural motif among this group's genes. It is interesting that the gene derived from the fungus *Paraphoma chrysanthemicola* contains a DEAD and a helicase\_C domain at its C-terminal (Figure 5). Group 8 consists of 35 genes from several fungi belonging to Chytridiomycetes, Mucoromycota, and Zoopagomycota fungi (Table S1).

Group 9 genes were identified from most of the fungal lineages except for the yeast-like and the Microsporidia fungi. For Group 10, we obtained 85 genes from several fungal species. We only identified Group 11 genes from the Chytridiomycota and Zoopagomycotina fungi. For these four group genes, the SET domain is the only recognizable conserved structural motif among them. Up to now, the functions of these two groups' genes have not been identified (Figure 5).

### The SU(Var)4-20 Family (Group 12)

Group 12 genes were represented in most of the fungal lineages except for the Zoopagomycotina fungi. Intriguingly, we did not find the group 12 genes in most of the Saccharomycotina fungi except for *Yarrowia lipolytica*, which contains one gene in this group (Table S1, Figures 5 and 8). Thus, our CAFE results showed that this group gene was constrained in the Saccharomycotina fungi and Zoopagomycotina fungi, but this group's genes have expanded in the *Paraphoma chrysanthemicola* because two genes were identified in this fungus (Figure 9). Moreover, in the Mucoromycota fungus *R. delemar* third group, 12 genes were identified (Figure 9). Thus, gene expansion has occurred in this fungus. It is interesting that the putative SET9 homologs (*SpSet9*, *Sp-NP\_588078*) identified from the fission yeast *S. pombe* are present in this group (Table 1). This group's genes only carry only one SET domain located at the N-terminal (Figure 5). It's known that SpSet9 carries out mono-, di-, and trimethylation of H4K20 and is involved in the maintenance of heterochromatin in metazoans [56,58].

#### 3.4.2. Cluster 2

The genes grouped into Cluster 2 can be divided into six groups (Group 9–13) (Figure 6). Groups 13 and 14 commonly existed in most of the fungal lineages except for Zoopagomycotina and Microsporidia in Group 13 and the Chytridiomycota, Zoopagomycota fungi, and Microsporidia in Group 14 (Figure 8 and Table S1). For Group 13, the SET domain is the only domain identified while in Group 14, another RubiscoMS domain has been identified downstream of the SET domain. Within Group 13, *SpSet11* (*Sp-NP\_588349*) and *Sc-NP\_010484* were identified (Table 1). *SpSet13* (*Sp-NP\_594072*) and the other two genes (*Nc-XP\_958526* and *Sc-NP\_010543*) were clustered into Group 14 (Table 1). For Group 15, 22 genes were identified from several Pezizomycotina and Taphrinomycotina fungi. In this group, *N. crassa* (*Nc-XP\_960703*) and *SpSet8* (*Sp-NP\_593610*) were included in this group (Tables S1 and 1). As an Ascomycota-specific group, this group's genes may have emerged after the separation of Ascomycota and Basidiomycota fungi (Figure 8 and Table S1). Group 16 also mainly consisted of the genes from several Pezizomycotina fungi but also include several genes from Chytridiomycota and Mucoromycota fungi (Figure 8 and Table S1). Intriguingly, the genes from Pezizomycotina fungi contain a partial SET domain, a RubiscoMS, and a MYND-finger (myeloid, nervy, and DEAF-1 factor), whereas the genes in this group from the Chytridiomycota and Mucoromycota fungi only contain a SET domain (Figure 6), suggesting that the Pezizomycotina fungi have obtained another function during their evolution. MYND motif is a Zn-binding domain involved in H3K36- and H3K4-specific methylation [59,60]. However, although *NcSet9* (*Nc-XP\_963594*) from *N. crassa* is also included in this group, nothing is known about its activity or function yet (Table 1). For the genes in Group 17, two genes (*Sc-NP-011824* and *Sc-NP-015116*) from *S. cerevisiae* and the gene *SpSet10* (*Sp-NP-595446*) in *S. pombe* were identified (Table 1). It is interesting that no Group 18 genes were identified from the Pezizomycotina fungi, suggesting that this group's genes have been lost by the Pezizomycotina fungi.

#### 3.4.3. Cluster 3

In our study, Cluster 3 was divided into nine groups (Group 19–27) (Figure 7). *SpSet11* (*Sp-NP\_588361*) from *S. pombe* is clustered into group 19 (Table 1). Group 22 genes are only identified from the Eurotiomycetes fungi, suggesting this group's genes were Eurotiomycetes-specific. Moreover, none of the Group 24 genes were identified from the Pezizomycotina fungi, while no Group 25 genes were identified from the Ascomycota fungi (Figure 8 and Table S1), indicating that these genes have been lost by the Pezizomycotina and Ascomycota fungi, respectively. Domain structure analysis showed that those genes from Group 25 contain a domain named MAS20 upstream of the SET domain (Figure 7). The genes belonging to Group 21 also carry the MYND domain (Figure 7), suggesting their potential H3K36- and/or H3K4- methylation abilities, though their specific functions are still unknown. The *ScSet5* (*Sc-P38890*) from *S. cerevisiae* clustered into Group 21 and it

has been identified as an H4K5, K8, and K12 methyltransferase that functions with Set1 to promote repression at telomeres [61]. However, the exact functions of genes belonging to other groups remain unstudied. The gene *ScSET6* (*Sc-Np\_015160*) from *S. cerevisiae* belonged to Group 26. Group 21 is the biggest SET-domain-containing gene group in fungi. It contains 541 genes, and the gene numbers range from 0-16 within individual species (Table S1). Although so many genes are grouped into this Group, their functions are still unknown yet.

### 3.5. Different Families/Groups Have Distinct Evolutionary Histories

At present, the functions of most SET genes identified here have not been characterized, but undoubtedly not all the SET domain proteins are histone methyltransferases. At present, only five clades genes have known histone targets, suggesting that H3K4, H3K9, H3K27, and H3K36 methylation is likely an ancient ability. However, the gene that mediates H3K27 (Group 2) was found in Pezizomycotina fungi but was totally lost in the yeast-like fungi (Table S1 and Figure 8). It is interesting that the genes that mediate H3K9 (Group 5) were also not observed in the Saccharomycotina fungi, and not all the Basidiomycota fungi contain the genes (Table S1 and Figure 8). To study the evolutionary history of the SET-domain genes in fungi, we built phylogenetic trees at the species and gene levels. Using phylogenomic and family sizes to estimate the gene family evolution, the numbers present on each node correspond to the expansion and contraction numbers of orthologous genes predicted by CAFE. As seen in Figure S3, the CAFE analysis showed that there were various expansions and contractions among the fungal lineages (Figure S3). With the further characterization of the five SET domain groups that conduct the function of histone modification in fungi, we found that different gene duplication and loss events have occurred in the five gene groups along with the main evolutionary lineages of fungi, leading to the diversification of their functions in different fungi (Figures 4 and 8). However, for each group's proteins, too many expansions and contractions have not occurred, which means the fungal species retained a similar gene number of each SET-domain gene, suggesting their conservation and functional importance for fungi.

Our analysis showed that some SET-domain-containing gene groups are lineage-specific groups because this group's genes are only present in several specific lineages (Figures 4 and 8). For instance, the genes in Group 5 were all lost in Saccharomycotina fungi and the genes belonging to Chytridiomycota were not identified in Group 1 (Figure 8 and Table S1). Interestingly, some SET-domain genes were lost in a specific lineage while emerging in a few species of other lineages. For instance, the genes from Groups 7, 15, and 22 are only present in Pezizomycotina fungi, implying that these SET-domain genes have emerged during the evolution of Pezizomycotina lineages (Table S1 and Figure 8). The genes belonging to Groups 8 and 11 were only present in the Zoopagomycota, Chytridiomycota, and other early diverged fungal lineages but disappeared in the Ascomycota and Basidiomycota fungi (Table S1 and Figure 8), suggesting that they were lost by Ascomycota and Basidiomycota fungi. The functions of those genes only presented in specific groups are still unknown. However, they may be expected to modify histones or have very different biological functions during the evolution of fungal species. Thus, studying their functions will be an exciting task. However, although some SET-domain genes within Cluster 2 and Cluster 3 have identified functions, such as function as methylases of ribosomal proteins or elongation factor eEF1, the functions of most are unknown. This is especially the case with some genes that emerged recently and are only present in some lineages, or some genes that are only retained in a few ancestral fungal species.

In general, genes clustered into one clade might have similar functions, whereas novel functions might have formed among the genes clustered into separate clades. The genes within the same group showed highly conserved structured organizations, suggesting they have similar activities (Figures 5–7). Thus, although their potential activities have not yet been functionally identified through molecular experiments, we can propose their chemical activities according to the reported activity of homologous genes in the same

group. For example, NcDim-5 and SpClr4, which both conduct the methylation of H3K9 were clustered in Group 5 (Figure 5). Thus, we could speculate that the genes that are within this Group but distributed in other fungal species may all mediate the methylation of H3K9. However, more molecular experiments are still required to confirm their specific functions in the future. Interestingly, although the genes within a group have similar sequence and structure organizations, their functions may differ. For example, ScSet3C in *S. cerevisiae* functions as a repressor of meiosis-specific genes [47], mediated by the recruitment of Set3C and ScSet2 by long noncoding RNAs [62,63], the homologs of ScSet3 from *S. cerevisiae* was catalytically inactive due to the conserved NHS tripeptide has been changed to RRS [30], and the Set3C protein from *C. albicans* has been confirmed to play a role in biofilm production and drug resistance [48].

Genes within the same species but belonging to different groups may have obtained different functions. For example, in the model filamentous fungus *N. crassa*, 16 SET-domain genes were identified in our study, and several of them are known to have different functions (Table 1). Previous studies have confirmed that NcDim-5 is responsible for all H3K9me2/3 [46,64], and Ash1-like NcSET-3 may methylate H3K36 [4]. However, the other SET-domain genes in *N. crassa* were neither functionally identified nor activity confirmed.

Furthermore, the genes belonging to different SET groups may have similar functions. For example, the histone methyltransferases Set5 and Set1 share overlapping functions of gene silencing and telomere maintenance in the different fungal SET Groups. In conclusion, although not all the functions of the SET-domain-containing genes have been functionally characterized, those genes may play different roles in increasing adaptability or helping fungi survive in various ecological habitats.

**Table 1.** Distribution of the SET Domain Proteins in *N. crassa*, *S. cerevisiae* and *S. pombe*.

Cluster	Groups	Previous Names	Presence of Methylation	<i>Neurospora crassa</i>	<i>Saccharomyces cerevisiae</i>	<i>Schizosaccharomyces pombe</i>
Cluster 1	Group 1	SET3C family		XP_957466 (SET4)	NP_012430(SET4)/ NP_012954 (SET3)	NP_594837 (SET3)
	Group 2	Ez family	H3K27me	XP_965043 (SET7)		
	Group 3	Ash1 family		XP_964116 (SET3)		
	Group 4	SET1 family	H3K4me	XP_961572 (SET1)	NP_011987 (SET1)	NP_587812 (SET1)
	Group 5	SUV39 family	H3K9me	EAA28243 (DIM-5)		NP_595186 (Clr4)
	Group 6	SET2 family	H3K36me	XP_957740 (SET2)	NP_012367 (SET2)	NP_594980 (SET2)
	Group 7					
	Group 8					
	Group 9					
	Group 10					
	Group 11					
	Group 12	Suv4-20 family	H4K20me	EAA33797 (SET10)		NP_588078 (SET9)
Cluster 2	Group 13				NP_010484	NP_588349 (SET11)
	Group 14			XP_958526	NP_010543	NP_594072 (SET13)
	Group 15			XP_960703		NP_593610 (SET8)
	Group 16			XP_963594 (SET9)		
	Group 17				NP_011824/NP_015116	NP_595446 (SET10)
	Group 18				NP_009586	
Cluster 3	Group 19					NP_588361 (SET7)
	Group 20			EAA35550/NCU00870		
	Group 21			XP_957968(SET11)/ XP_963161/NCU_00296		NP_588413 (SET5)
	Group 22					
	Group 23					
	Group 24					
	Group 25					
Group 26					NP_015160(SET6)	
Group 27				NCU_002962	P38890(SET5)	NP_596514 (SET6)

#### 4. Conclusions

So far, only a tiny number of SET-domain genes in fungi have been characterized, and almost all studies have focused on the nonpathogenic model species. Thus, it would be challenging yet fascinating to study the functions and molecular mechanisms of SET-domain genes in those non-model species, especially the important plant and animal pathogens. Our analysis indicated the SET-domain containing genes in fungi had undergone a consid-

erable number of duplications, diversification, and loss events, resulting in broad functional diversities among various fungal species. The evolutionary dynamics help them obtain novel functions, leading them to help fungi adapt to various environments. We expect our study can provide a guideline for the functional research on SET-domain genes of fungi in future studies.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/jof8111159/s1>, Figure S1. CLANS clustering of 4171 SET domain sequences obtained from 229 whole genome sequenced fungal species. A Java-based software program, CLANS, was used to elucidate the relationships between and within subfamilies of the putative 4172 SET domain sequences from 229 fungal genomes. Figure S2. Chromosomal distribution of the SET-domain containing genes in *A. fumigatus* genome. 13 SET-domain genes are unevenly distributed on the 8 chromosomes of *A. fumigatus* genome, and the number of genes on each chromosome is irrelevant to chromosome size. Figure S3. Various expansions and contractions among the fungal lineages using CAFE analysis. The numbers present on each node correspond to the expansion and contraction numbers of orthologues genes predicted by CAFE. Table S1. Distribution of SET-domain genes in 229 fungal organisms. Numbers in brackets represent the deleted gene numbers which show ambiguously aligned regions around the three active catalytic residues.

**Author Contributions:** Conceptualization, G.D., L.S. and J.L.; methodology, J.L.; software, G.D., L.S. and W.Z.; formal analysis, L.S. and W.Z.; resources, L.S., S.L., Z.Z. and X.H.; data curation, G.D.; writing—original draft preparation, J.L.; writing—review and editing, J.L.; visualization, G.D.; funding acquisition, J.L. All authors have read and agreed to the published version of the manuscript.”

**Funding:** This research was jointly supported by the National Natural Science Foundation of China (approved nos. 32170198, 31970073, and 31760538).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** We are grateful to Jianping Xu of the Dept. Biology, McMaster University, for valuable comments and critical discussions. We also thank the anonymous reviewers for their valuable suggestions.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- Campos, E.I.; Reinberg, D. Histones: Annotating chromatin. *Annu. Rev. Genet.* **2009**, *43*, 559. [[CrossRef](#)] [[PubMed](#)]
- Carrozza, M.J.; Li, B.; Florens, L.; Suganuma, T.; Swanson, S.K.; Lee, K.K.; Shia, W.J.; Anderson, S.; Yates, J.; Washburn, M.P.; et al. Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. *Cell* **2005**, *123*, 581–592. [[CrossRef](#)] [[PubMed](#)]
- Briggs, S.D.; Bryk, M.; Strahl, B.D.; Cheung, W.L.; Davie, J.K.; Dent, S.Y.; Winston, F.; Allis, C.D. Histone H3 lysine 4 methylation is mediated by Set1 and required for cell growth and rDNA silencing in *Saccharomyces cerevisiae*. *Genes Dev.* **2001**, *15*, 3286–3295. [[CrossRef](#)]
- Freitag, M. Histone methylation by SET domain proteins in fungi. *Annu. Rev. Microbiol.* **2017**, *71*, 413–439. [[CrossRef](#)]
- Huck Hui, N.; Qin, F.; Hengbin, W.; Hediye, E.B.; Paul, T.; Yi, Z.; Kevin, S. Lysine methylation within the globular domain of histone H3 by Dot1 is important for telomeric silencing and Sir protein association. *Genes Dev.* **2002**, *16*, 1518–1527.
- Qin, F.; Hengbin, W.; Huck Hui, N.; Hediye, E.B.; Paul, T.; Kevin, S.; Yi, Z. Methylation of H3-lysine 79 is mediated by a new family of HMTases without a SET domain. *Curr. Biol.* **2002**, *12*, 1052–1058.
- Tschiersch, B.; Hofmann, A.; Krauss, V.; Dorn, R.; Korge, G.; Reuter, G. The protein encoded by the *Drosophila* position-effect variegation suppressor gene *Su(var)3–9* combines domains of antagonistic regulators of homeotic gene complexes. *EMBO J.* **1994**, *13*, 3822–3831. [[CrossRef](#)]
- Jones, R.S.; Gelbart, W.M. The *Drosophila* Polycomb-group gene Enhancer of zeste contains a region with sequence similarity to trithorax. *Mol. Cell. Biol.* **1993**, *13*, 6357. [[PubMed](#)]
- Dillon, S.C.; Zhang, X.; Trievel, R.C.; Cheng, X. The SET-domain protein superfamily: Protein lysine methyltransferases. *Genome Biol.* **2005**, *6*, 227. [[CrossRef](#)]
- Jenuwein, T.; Laible, G.; Dorn, R.; Reuter, G. SET domain proteins modulate chromatin domains in eu- and heterochromatin. *Cell. Mol. Life Sci. CMLS* **1997**, *54*, 80–93. [[CrossRef](#)]

11. Black, J.C.; Van Rechem, C.; Whetstine, J.R. Histone lysine methylation dynamics: Establishment, regulation, and biological impact. *Mol. Cell* **2012**, *48*, 491–507. [[CrossRef](#)] [[PubMed](#)]
12. Fahrenkrog, B. Histone modifications as regulators of life and death in *Saccharomyces cerevisiae*. *Microb. Cell* **2015**, *3*, 1–13. [[CrossRef](#)] [[PubMed](#)]
13. Zhang, T.; Cooper, S.; Brockdorff, N. The interplay of histone modifications—Writers that read. *EMBO Rep.* **2015**, *16*, 1467–1481. [[CrossRef](#)] [[PubMed](#)]
14. Rea, S.; Eisenhaber, F.; O’Carroll, D.; Strahl, B.D.; Sun, Z.W.; Schmid, M.; Opravil, S.; Mechtler, K.; Ponting, C.P.; Allis, C.D.; et al. Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature* **2000**, *406*, 593–599. [[CrossRef](#)]
15. Nakayama, J.; Rice, J.C.; Strahl, B.D.; Allis, C.D.; Grewal, S.I. Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly. *Science* **2001**, *292*, 110–113. [[CrossRef](#)]
16. Grewal, S.I.; Moazed, D. Heterochromatin and epigenetic control of gene expression. *Science* **2003**, *301*, 798–802. [[CrossRef](#)]
17. Sims, R.J., 3rd; Reinberg, D. Histone H3 Lys 4 methylation: Caught in a bind? *Genes Dev.* **2006**, *20*, 2779–2786. [[CrossRef](#)]
18. Bilokapic, S.; Halic, M. Nucleosome and ubiquitin position Set2 to methylate H3K36. *Nat. Commun.* **2019**, *10*, 3795. [[CrossRef](#)]
19. Jamieson, K.; Rountree, M.R.; Lewis, Z.A.; Stajich, J.E.; Selker, E.U. Regional control of histone H3 lysine 27 methylation in *Neurospora*. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 6027–6032. [[CrossRef](#)]
20. Connolly, L.R.; Smith, K.M.; Freitag, M. The *Fusarium graminearum* histone H3 K27 methyltransferase KMT6 regulates development and expression of secondary metabolite gene clusters. *PLoS Genet.* **2013**, *9*, e1003916. [[CrossRef](#)]
21. Ridenour, J.B.; Möller, M.; Freitag, M. Polycomb Repression without Bristles: Facultative Heterochromatin and Genome Stability in Fungi. *Genes* **2020**, *11*, 638. [[CrossRef](#)] [[PubMed](#)]
22. Tamaru, H.; Selker, E.U. A histone H3 methyltransferase controls DNA methylation in *Neurospora crassa*. *Nature* **2001**, *414*, 277–283. [[CrossRef](#)] [[PubMed](#)]
23. Adhvaryu, K.K.; Morris, S.A.; Strahl, B.D.; Selker, E.U. Methylation of histone H3 lysine 36 is required for normal development in *Neurospora crassa*. *Eukaryot. Cell* **2005**, *4*, 1455–1464. [[CrossRef](#)] [[PubMed](#)]
24. Tran, K.; Jethmalani, Y.; Jaiswal, D.; Green, E.M. Set4 is a chromatin-associated protein, promotes survival during oxidative stress, and regulates stress response genes in yeast. *J. Biol. Chem.* **2018**, *293*, 14429–14443. [[CrossRef](#)]
25. Pham, K.T.; Inoue, Y.; Vu, B.V.; Nguyen, H.H.; Nakayashiki, T.; Ikeda, K.; Nakayashiki, H. MoSET1 (histone H3K4 methyltransferase in *Magnaporthe oryzae*) regulates global gene expression during infection-related morphogenesis. *PLoS Genet.* **2015**, *11*, e1005385. [[CrossRef](#)]
26. Gu, Q.; Tahir, H.A.; Zhang, H.; Huang, H.; Ji, T.; Sun, X.; Wu, L.; Wu, H.; Gao, X. Involvement of FvSet1 in Fumonisin B1 biosynthesis, vegetative growth, fungal virulence, and environmental stress responses in *Fusarium verticillioides*. *Toxins* **2017**, *9*, 43. [[CrossRef](#)]
27. Fingerman, I.M.; Wu, C.L.; Wilson, B.D.; Briggs, S.D. Global loss of Set1-mediated H3 Lys4 trimethylation is associated with silencing defects in *Saccharomyces cerevisiae*. *J. Biol. Chem.* **2005**, *280*, 28761–28765. [[CrossRef](#)]
28. McDaniel, S.L.; Hepperla, A.J.; Huang, J.; Dronamraju, R.; Adams, A.T.; Kulkarni, V.G.; Davis, I.J.; Strahl, B.D. H3K36 Methylation Regulates Nutrient Stress Response in *Saccharomyces cerevisiae* by Enforcing Transcriptional Fidelity. *Cell Rep.* **2017**, *19*, 2371–2382. [[CrossRef](#)]
29. Allshire, R.C.; Ekwall, K. Epigenetic Regulation of Chromatin States in *Schizosaccharomyces pombe*. *Cold Spring Harb. Perspect. Biol.* **2015**, *7*, a018770. [[CrossRef](#)]
30. Veerappan, C.S.; Avramova, Z.; Moriyama, E.N. Evolution of SET-domain protein families in the unicellular and multicellular Ascomycota fungi. *BMC Evol. Biol.* **2008**, *8*, 190. [[CrossRef](#)]
31. Frickey, T.; Lupas, A. CLANS: A Java application for visualizing protein families based on pairwise similarity. *Bioinformatics* **2004**, *20*, 3702–3704. [[CrossRef](#)] [[PubMed](#)]
32. Edgar, R. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **2004**, *32*, 1792–1797. [[CrossRef](#)] [[PubMed](#)]
33. Tamura, K.; Nei, M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol. Biol. Evol.* **1993**, *10*, 512–526. [[PubMed](#)]
34. Kumar, S.; Stecher, G.; Li, M.; Knyaz, C.; Tamura, K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol. Biol. Evol.* **2018**, *35*, 1547–1549. [[CrossRef](#)]
35. Abascal, F.; Zardoya, R.; Posada, D. ProtTest: Selection of best-fit models of protein evolution. *Bioinformatics* **2005**, *21*, 2104–2105. [[CrossRef](#)]
36. Price, M.N.; Dehal, P.S.; Arkin, A.P. FastTree: Computing large minimum evolution trees with profiles instead of a distance matrix. *Mol. Biol. Evol.* **2009**, *26*, 1641–1650. [[CrossRef](#)]
37. De Bie, T.; Cristianini, N.; Demuth, J.P.; Hahn, M.W. CAFE: A computational tool for the study of gene family evolution. *Bioinformatics* **2006**, *22*, 1269–1271. [[CrossRef](#)]
38. Mendes, F.K.; Vanderpool, D.; Fulton, B.; Hahn, M.W. CAFE 5 models variation in evolutionary rates among gene families. *Bioinformatics* **2020**, *36*, 5516–5518. [[CrossRef](#)]
39. Emms, D.M.; Kelly, S. OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biol.* **2019**, *20*, 238. [[CrossRef](#)]

40. Sanderson, M.J. r8s: Inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics* **2003**, *19*, 301–302. [[CrossRef](#)]
41. Chen, C.; Chen, H.; Zhang, Y.; Thomas, H.R.; Frank, M.H.; He, Y.; Xia, R. TBtools: An Integrative Toolkit Developed for Interactive Analyses of Big Biological Data. *Mol. Plant* **2020**, *13*, 1194–1202. [[CrossRef](#)] [[PubMed](#)]
42. Wang, Y.; Tang, H.; Debarry, J.D.; Tan, X.; Li, J.; Wang, X.; Lee, T.H.; Jin, H.; Marler, B.; Guo, H.; et al. MScanX: A toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* **2012**, *40*, e49. [[CrossRef](#)] [[PubMed](#)]
43. Accari, S.L.; Fisher, P.R. Emerging Roles of JmjC Domain-Containing Proteins. *Int. Rev. Cell Mol. Biol.* **2015**, *319*, 165–220. [[CrossRef](#)]
44. Catania, S.; Dumesic, P.A.; Pimentel, H.; Nasif, A.; Stoddard, C.I.; Burke, J.E.; Diedrich, J.K.; Cook, S.; Shea, T.; Geinger, E.; et al. Evolutionary Persistence of DNA Methylation for Millions of Years after Ancient Loss of a De Novo Methyltransferase. *Cell* **2020**, *180*, 263–277.e220. [[CrossRef](#)]
45. Du, J.; Johnson, L.M.; Jacobsen, S.E.; Patel, D.J. DNA methylation pathways and their crosstalk with histone methylation. *Nat. Rev. Mol. Cell Biol.* **2015**, *16*, 519–532. [[CrossRef](#)] [[PubMed](#)]
46. Kim, T.; Buratowski, S. Dimethylation of H3K4 by Set1 recruits the Set3 histone deacetylase complex to 5' transcribed regions. *Cell* **2009**, *137*, 259–272. [[CrossRef](#)]
47. Pijnappel, W.W.; Schaft, D.; Roguev, A.; Shevchenko, A.; Tekotte, H.; Wilm, M.; Rigaut, G.; Séraphin, B.; Aasland, R.; Stewart, A.F. The *S. cerevisiae* SET3 complex includes two histone deacetylases, Hos2 and Hst1, and is a meiotic-specific repressor of the sporulation gene program. *Genes Dev.* **2001**, *15*, 2991–3004. [[CrossRef](#)]
48. Nobile, C.J.; Fox, E.P.; Hartooni, N.; Mitchell, K.F.; Hnisz, D.; Andes, D.R.; Kuchler, K.; Johnson, A.D. A histone deacetylase complex mediates biofilm dispersal and drug resistance in *Candida albicans*. *mBio* **2014**, *5*, e01201–e01214. [[CrossRef](#)]
49. Cao, R.; Wang, L.; Wang, H.; Xia, L.; Erdjument-Bromage, H.; Tempst, P.; Jones, R.S.; Zhang, Y. Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science* **2002**, *298*, 1039–1043. [[CrossRef](#)]
50. Beisel, C.; Imhof, A.; Greene, J.; Kremmer, E.; Sauer, F. Histone methylation by the *Drosophila* epigenetic transcriptional regulator Ash1. *Nature* **2002**, *419*, 857–862. [[CrossRef](#)]
51. Dorafshan, E.; Kahn, T.G.; Glotov, A.; Savitsky, M.; Walther, M.; Reuter, G.; Schwartz, Y.B. Ash1 counteracts Polycomb repression independent of histone H3 lysine 36 methylation. *EMBO Rep.* **2019**, *20*, e46762. [[CrossRef](#)] [[PubMed](#)]
52. Schlichter, A.; Cairns, B.R. Histone trimethylation by Set1 is coordinated by the RRM, autoinhibitory, and catalytic domains. *EMBO J.* **2005**, *24*, 1222–1231. [[CrossRef](#)] [[PubMed](#)]
53. Thon, G.; Hansen, K.R.; Altes, S.P.; Sidhu, D.; Singh, G.; Verhein-Hansen, J.; Bonaduce, M.J.; Klar, A.J. The Clr7 and Clr8 directionality factors and the Pcu4 cullin mediate heterochromatin formation in the fission yeast *Schizosaccharomyces pombe*. *Genetics* **2005**, *171*, 1583–1595. [[CrossRef](#)]
54. Zhang, X.; Yang, Z.; Khan, S.I.; Horton, J.R.; Tamaru, H.; Selker, E.U.; Cheng, X. Structural basis for the product specificity of histone lysine methyltransferases. *Mol. Cell* **2003**, *12*, 177–185. [[CrossRef](#)]
55. Li, J.; Moazed, D.; Gygi, S.P. Association of the histone methyltransferase Set2 with RNA polymerase II plays a role in transcription elongation. *J. Biol. Chem.* **2002**, *277*, 49383–49388. [[CrossRef](#)]
56. Du, L.L.; Nakamura, T.M.; Russell, P. Histone modification-dependent and -independent pathways for recruitment of checkpoint protein Crb2 to double-strand breaks. *Genes Dev.* **2006**, *20*, 1583–1596. [[CrossRef](#)]
57. Strahl, B.D.; Grant, P.A.; Briggs, S.D.; Sun, Z.W.; Bone, J.R.; Caldwell, J.A.; Mollah, S.; Cook, R.G.; Shabanowitz, J.; Hunt, D.F.; et al. Set2 is a nucleosomal histone H3-selective methyltransferase that mediates transcriptional repression. *Mol. Cell Biol.* **2002**, *22*, 1298–1306. [[CrossRef](#)]
58. Schotta, G.; Lachner, M.; Sarma, K.; Ebert, A.; Sengupta, R.; Reuter, G.; Reinberg, D.; Jenuwein, T. A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin. *Genes Dev.* **2004**, *18*, 1251–1262. [[CrossRef](#)]
59. Hamamoto, R.; Furukawa, Y.; Morita, M.; Iimura, Y.; Silva, F.P.; Li, M.; Yagyu, R.; Nakamura, Y. SMYD3 encodes a histone methyltransferase involved in the proliferation of cancer cells. *Nat. Cell Biol.* **2004**, *6*, 731–740. [[CrossRef](#)]
60. Brown, M.A.; Sims, R.J., 3rd; Gottlieb, P.D.; Tucker, P.W. Identification and characterization of Smyd2: A split SET/MYND domain-containing histone H3 lysine 36-specific methyltransferase that interacts with the Sin3 histone deacetylase complex. *Mol. Cancer* **2006**, *5*, 26. [[CrossRef](#)]
61. Green, E.M.; Mas, G.; Young, N.L.; Garcia, B.A.; Gozani, O. Methylation of H4 lysines 5, 8 and 12 by yeast Set5 calibrates chromatin stress responses. *Nat. Struct. Mol. Biol.* **2012**, *19*, 361–363. [[CrossRef](#)] [[PubMed](#)]
62. Kim, T.S.; Xu, Z.; Clauder-Münster, S.; Steinmetz, L.; Buratowski, S. Set3 HDAC Mediates Effects of Overlapping Noncoding Transcription on Gene Induction Kinetics. *Cell* **2012**, *150*, 1158–1169. [[CrossRef](#)] [[PubMed](#)]
63. Van Werven, F.J.; Neuert, G.; Hendrick, N.; Lardenois, A.; Buratowski, S.; van Oudenaarden, A.; Primig, M.; Amon, A. Transcription of two long noncoding RNAs mediates mating-type control of gametogenesis in budding yeast. *Cell* **2012**, *150*, 1170–1181. [[CrossRef](#)] [[PubMed](#)]
64. Tamaru, H.; Zhang, X.; McMillen, D.; Singh, P.B.; Nakayama, J.; Grewal, S.I.; Allis, C.D.; Cheng, X.; Selker, E.U. Trimethylated lysine 9 of histone H3 is a mark for DNA methylation in *Neurospora crassa*. *Nat. Genet.* **2003**, *34*, 75–79. [[CrossRef](#)]