Mycena genomes resolve the evolution of fungal bioluminescence

Huei-Mien Ke[1], Hsin-Han Lee[1], Chan-Yi Ivy Lin[1,2], Yu-Ching Liu[1], Min R. Lu[1,3], Jo-Wei Allison Hsieh[3,4], Chiung-Chih Chang[1,5], Pei-Hsuan Wu[6], Meiyeh Jade Lu[1], Jeng-Yi Li[1], Gaus Shang[7], Rita Jui-Hsien Lu[2,8], László G. Nagy[9], Pao-Yang Chen[3,4], Hsiao-Wei Kao[5] and Isheng Jason Tsai[1,3]

[1] Biodiversity Research Center, Academia Sinica, Taipei, Taiwan

[2] Department of Molecular, Cellular and Developmental Biology, Yale University, New Haven, Connecticut 06520, USA

[3] Genome and Systems Biology Degree Program, Academia Sinica and National Taiwan University, Taipei, Taiwan

[4] Institute of Plant and Microbial Biology, Academia Sinica, Taipei, Taiwan

[5] Department of Life Sciences, National Chung Hsing University, Taichung, Taiwan

[6] Master Program for Plant Medicine and Good Agricultural Practice, National Chung Hsing University, Taiwan

[7] Department of Biotechnology, Ming Chuan University, Taiwan

[8] Department of Medicine, Washington University in St. Louis, St. Louis, MO 63110, USA

[9] Synthetic and Systems Biology Unit, Biological Research Centre, 6726 Szeged, Hungary

Correspondence: Huei-Mien Ke cotton@sinica.edu.tw and Isheng Jason Tsai ijtsai@sinica.edu.tw

Abstract

Mushroom-forming fungi in the order Agaricales represent an independent origin of bioluminescence in the tree of life, yet the diversity, evolutionary history, and timing of the origin of fungal luciferases remain elusive. We sequenced the genomes and transcriptomes of five bonnet mushroom species (*Mycena* spp.), a diverse lineage comprising the majority of bioluminescent fungi. We show that bioluminescence evolved in the common ancestor of *Mycena* spp. and the marasmioid clade of Agaricales and was maintained through at least 160 million years of evolution. We revealed *Mycena* exhibit two-speed genomes and resolve how the luciferase cluster was derived by duplication and translocation, frequently rearranged and lost in most *Mycena* species, but conserved in the *Armillaria* lineage. Luciferase cluster members were co-expressed across developmental stages, with highest expression in fruiting body caps and stipes, suggesting fruiting-related adaptive functions. Our results contribute to understanding a *de novo* origin of bioluminescence and the corresponding gene cluster in a diverse group of enigmatic fungal species.

The genus *Mycena* (Pers.) Roussel, comprises approximately 600 small mushroom species widely distributed around the world[1]. Also known as bonnet mushrooms, *Mycena* species are usually characterised by a bell-shaped cap, a thin stem (**Fig. 1a**), and a gilled or porioid hymenophore[2]. *Mycena* also have a diversity of life history strategies; while many species are saprotrophic, they can be pathogens as well as mycorrhizal[3]. Despite its vast diversity of lifestyles and phenotypes, there are many outstanding questions concerning the basic biology, ecology and genomics of this genus. One particular fascinating trait is bioluminescence, which more than half of the 81 recorded luminescent fungi belong to *Mycena*[4]. Yet, bioluminescence occurs only in a small percentage of this genus, suggesting an intricate loss/gain history and potential convergence within the genus.

Fungal light emission involves two main steps. First, a luciferin precursor of hispidin is hydroxylated by hispidin-3-hydroxylase (H3H) into 3-hydroxyhispidin (luciferin)[5]. Oxygen is then added to the luciferin by luciferase, producing a high energy intermediate which is subsequently decomposed, yielding light emission. Previously, Kotlobady *et al.* have identified the fungal luciferase, which is physically adjacent to these enzymes and forms a gene cluster containing luciferase, hispidin synthase and H3H[6]. This cluster was found to be conserved across bioluminescent fungi of three lineages: *Armillaria*, mycenoid and *Omphalotus*[6]. Phylogeny reconstruction suggested that luciferase originated in early Agaricales. *Armillaria* and *Omphalotus* belong to the marasmioid clade, whereas *Mycena* was recently found to be sister of the marasmioid clade[7]. Recent genome sequencing efforts in the marasmioid clade revealed diverse genomic and life history traits, including genome expansion and pathogenicity in *Armillaria* spp.[8], novel wood decay strategies[9] or fruiting body development[10]. Genomes of two *Mycena* species were sequenced[6], however, the fragmented assemblies (N50 5.8–16.7 kb) impeded comparative genomic analyses of features such as synteny[11]. These resources provide a substrate for studies of genome evolution and of bioluminescence in fungi, however, several key questions are still unknown from an evolutionary perspective. Here, we set out to identify novel genes involved in bioluminescence, understand how the luciferase cluster was lost or retained, and whether there are levels of variation in this cluster across these lineages.

To gain insights into the evolution of fungal bioluminescence and the ecology of mycenoid, we sequenced the genomes of four bioluminescent (*Mycena chlorophos*, *M. kentingensis*, *M. sanguinolenta* and *M. venus*) and one non-bioluminescent (*M. indigotica*) species. We conducted comparative genomics with representative genomes of all bioluminescent fungal clades, putting particular emphasis on genome-wide

synteny to investigate the evolutionary dynamics of the luciferase gene cluster through hundreds of millions of years. The variability in genome sizes among *Mycena* is likely associated with the differential expansion of repeats in the genomes, potentially due to the differential control on repeat activity by DNA methylation. The transcriptome of bioluminescent mycelium correctly revealed the luciferase cluster and co-expression analyses identified further genes that may be relevant to bioluminescence and development. Based on comparative analyses from fifteen available genomes of bioluminescent fungi, we reconstructed and presented a model for evolution of fungal bioluminescence.

Results

Assemblies and annotations of five *Mycena* species

We sequenced the genomes of the bioluminescent fungi *Mycena chlorophos*, *M. kentingensis*, *M. sanguinolenta* and *M. venus,* as well as the non-bioluminescent *M. indigotica* (**Fig. 1a**). These species were chosen for their phylogenetic positions (Supplementary Fig. 1) and because they displayed different bioluminescence intensities. An assembly was produced for each species using a combination of Illumina and Oxford Nanopore reads (Supplementary Table 1). The nuclear genomes were 50.9–167.2 Mb with two amongst the largest for the Agaricales. The assemblies consisted of 30–155 contigs with N50 4.1–17.8 Mb (Supplementary Table 2), which were comparable to representative fungal reference assemblies and allowed for synteny comparisons[11]. Stretches of TTAGGG hexamers were identified at the end of scaffolds, indicating telomeric repeats commonly found in Agaricales[12,13]. The largest scaffolds in *M. indigotica* and *M. kentingensis* were telomere-to-telomere, indicating gapless chromosomes. Mitochondrial genomes in these species were assembled into single circular contigs of 88.3–133 kb (Supplementary Fig. 2).

Using a combination of reference fungal protein homology support and mycelium transcriptome sequencing (Supplementary Table 3), 13,940–26,334 protein encoding genes were predicted in the *Mycena* genomes using MAKER2[14] pipeline, and were 92.1–94.8% complete (Supplementary Table 2) based on BUSCO[15] analysis. Orthology inference using Orthofinder[16,17] placed these genes models and those of 37 other basidiomycetes genomes (Supplementary Table 4) into 22,244 orthologous groups (OGs; Supplementary Table 5). Of these OGs, 44.3% contained at least one orthologue from other basidiomycete, while 15–29% of the proteomes were found exclusively in each *Mycena* species (Supplementary Table 6). The genome sizes were positively correlated with proteome sizes, with the largest (*M. sanguinolenta*) and smallest (*M.*

chlorophos) varying two- and three-fold, respectively. Interestingly, the mitochondrial genomes were larger in species with smaller genomes, and this was due to various genes having gained an extensive number of introns (Supplementary Fig. 2). We assessed orthologous group evolution by analysing OG distribution dynamics along a time-calibrated phylogeny using CAFÉ[18]. Gene family changes were comparable to those of other branches of Agaricales. A total of 589 orthologous groups were expanded at the origin of the mycenoid lineage (Supplementary Fig. 3). Analysis of gene ontology terms showed that these genes were enriched in iron ion binding, transferase activity, DNA-binding transcription factor activity, and monooxygenase and dioxygenase activities, suggesting that these genes were necessary for mycenoid species adapting to different environments (Supplementary Table 7).

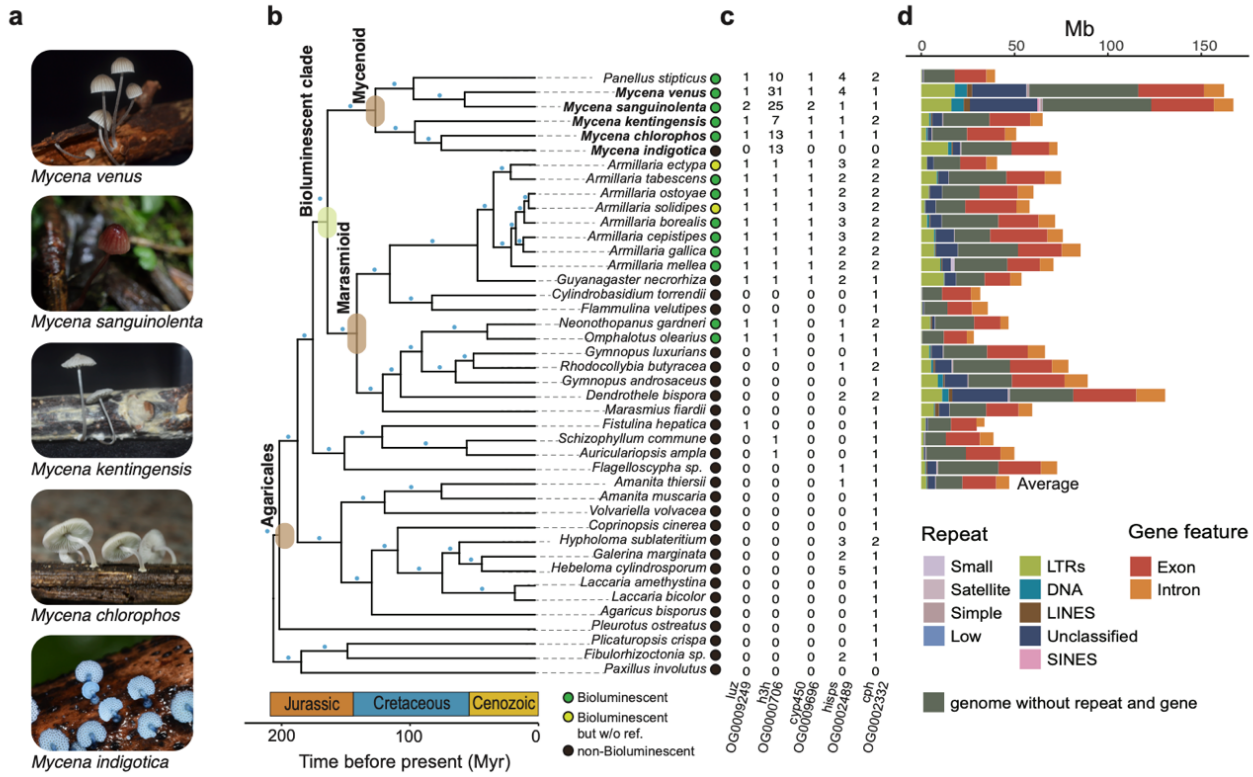


Fig. 1| Phylogenomic analysis of *Mycena* and related fungi. a, The five species sequenced in this study. **b**, Species trees inferred from a concatenated supermatrix of the gene alignments using the 360 single-copy orthogroups. X-axis denotes divergence time estimates. Blue dot on a branch indicates a bootstrap value > 90. **c**, Gene copy number in the orthologous groups (OG) associated with luciferin biosynthesis pathway including luciferase (*luz*), hispidin-3-hydroxylase (*h3h*), hispidin synthase (*hisps*), cytochrome P450 (*cyp450*) and caffeylpyruvate hydrolase (*cph*). **d**, Genome sizes for 42 species broken down by repeat types and gene features.

Averaged content in the genomes of 14 outgroup species are indicated as one bar. Repeats including transposable elements (TEs): long terminal repeats (LTRs), long interspersed nuclear elements (LINES), short interspersed nuclear elements (SINEs), DNA transposons (DNA), and other types of repeats: small RNA (Small), simple repeats (Simple), and low complexity repeats (Low).

A single origin of bioluminescent fungi in the ancestor of *Mycena* and the marasmioid clade

Phylogenomic analyses based on single-copy orthologue sets have placed *Mycena* sister to the marasmioid clade, including *Armillaria* and *Omphalotus*, which are the other two lineages in which bioluminescent species have been identified. This species phylogeny was recovered in both maximum likelihood analysis[19] of a concatenated supermatrix of single-copy gene alignments (**Fig. 1b**) and coalescent-based analysis using the 360 gene trees[20] (Supplementary Fig. 4). In our four bioluminescent *Mycena* species, we identified genes involved in luciferin biosynthesis and their orthologues across species (**Fig. 1c**). Phylogenies of individual gene families were congruent with the species tree (Supplementary Fig. 5 (a–e). In contrast to a previous report[6], our results suggest that the functional luciferase originated in the last common ancestor of mycenoid and marasmioid clade rather than in that of the Agaricales (**Fig. 1b**). The ancestral luciferase was initially duplicated in the last common ancestor of this large group and Schizophyllaceae and subsequently acquired bioluminescence ability after speciation (**Fig. 1c**; Supplementary Fig. 5a). Using MCMCtree[21] with three fossil calibrations, we estimated the age of mycenoid most recent common ancestor to be 105–147 million years ago (Mya) in the Cretaceous (**Fig. 1b**). This is consistent with recent estimates (78–110[7] and mean 125[1] Mya) and overlaps with the initial rise and diversification of angiosperms[22], suggesting that they are ecologically associated with fungi acting as saprotrophs or mycorrhizal partners[3]. Finally, the age of bioluminescence which was also the age of functional luciferase in fungi was estimated to originate around 160 million years ago during the late Jurassic (**Fig. 1b**).

Interplay between transposable elements and DNA methylation in *Mycena*

Similar to other fungal genomes[23,24], much of the variation in the *Mycena* nuclear genome can be explained by repetitive DNA content (Supplementary Table 8). Only 11.7% of the smallest genome (*M. chlorophos*) was repeats, which is in stark contrast to the 39.0% and 35.7% in *M. sanguinolenta* and *M. venus*, respectively. The majority of transposable elements in *Mycena* were long terminal repeats (LTRs) retrotransposons (60–85%), followed by DNA transposable elements (11%–24%) (**Fig. 1d** and Supplementary Table 8). Interestingly, the larger genomes of *M. sanguinolenta* and *M.*

venus contained the lowest proportion of LTRs (24.9 and 31.1%, respectively), but highest proportion of unclassified repeats (55.4 and 50.3%, respectively) (Supplementary Table 8). 16.6–36.5% of the unclassified repeat families shared 53.8–60.5% nucleotide identity with known transposable elements, suggesting they were degenerated copies which we defined as relic TEs (Supplementary Table 9). **Fig. 2a** shows that the largest assembled chromosome of *M. indigotica* exhibits high protein-coding gene content and low transposable element density at scaffold centres, which is typical of fungal chromosomes[25,26]. Such observations were consistent across large *Mycena* scaffolds (typically >1 Mb), suggesting that our assemblies were robust enough to capture evolutionary dynamics across chromosomes.

We detected DNA methylation levels of 5-methylcytosine (5mC) across the five *Mycena* assemblies using Nanopore long reads. CG sites were found either highly (mCG level >60%) or weakly methylated (<15%) in gene body, displaying a bimodal distribution (Supplementary Fig. 6). Such a bimodal distribution has also been observed in plants, animals, and other fungi, including *Tuber melanosporum* and *Pseudogymnoascus destructans*[27-32]. Within *Mycena*, the CG methylation in genes (5.4–10.5%) was much lower than that in repeats—i.e., TEs and unclassified repeats (11.6–84.5%) (**Fig. 2b**; Supplementary Table 10)—suggesting that DNA methylation may have a specific effect on repeats. Except for DNA transposons in *M. kentingensis*, LTR retrotransposons had the highest CG methylation levels of all types of transposable elements (**Fig. 2b**). Furthermore, CG methylation in relic TEs was clearly lower than that in classic TEs (Supplementary Table 10). Among the *Mycena* species, we found that *M. sanguinolenta* and *M. venus* with larger genomes and higher repeat content had lower levels of methylation in the repeats, and the repeat methylation was much higher in *M. indigotica*, *M. chlorophos*, and *M. kentingensis,* which have smaller genomes (**Fig. 2c**). The same pattern was also observed in genes, though they had fewer changes in their methylation level than did repeats. Our results indicate that the variant composition of repeats is differentially mediated by DNA methylation among closely-related *Mycena* species. Hence, genome expansion in *Mycena* was likely a result associated with transposable element proliferation and the accumulation of relic TEs, which yielded reduced methylation in active copies; this is also observed in some plants, e.g., *Arabis alpine*[33] and *Manihot esculenta*[34].

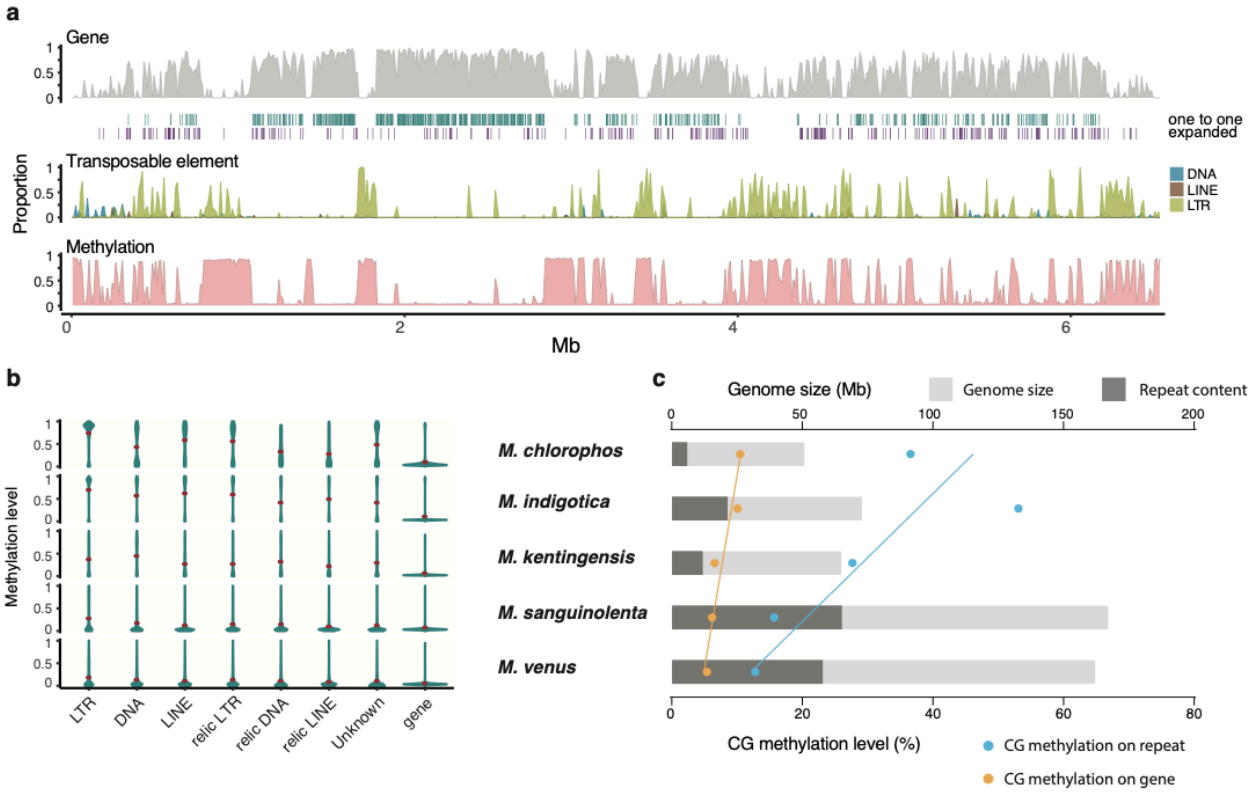


Fig. 2| Distribution of *Mycena* genome features. a, *M. indigotica* chromosome one. For every non-overlapping 10-kb window, the distributions from top to bottom are: (1) Gene density (percentage of nucleotides coverage). Stripes under this window denotes gene positions. Green stripe: gene containing single-copy orthologue with *M. chlorophos*. Purple stripe: gene expansion in individual species identified by CAFÉ. (2) Density of transposable elements (TEs), including LTRs, LINES, and DNA. (3) Average methylation level called from CpG sites per window. The high methylation window generally clustered in high TE regions with low gene density. **b**, The methylation level in genes and different types of repeats. **c**, The relationships among genome size, number of repeats and CG methylation levels in *Mycena*.

Core/dispensable genome partitioning based on synteny

We compared the patterns of 4,452 single-copy orthologue pairs between assemblies of *Mycena indigotica* and *Armillaria ectypa* (Supplementary Fig. 7). The majority of scaffolds between the two species were assigned to one-one relationship unambiguously, providing strong evidence that macro-synteny has been conserved between the marasmioid and mycenoid clades. There was no evidence of whole genome duplication events. Such chromosome-level synteny is also conserved in the common ancestor of Agaricales (Supplementary Fig. 8). The *M. indigotica* scaffolds typically exhibit high orthologous gene density in the centres of scaffolds **(Fig. 2a)**, and multi-genome comparison showed that synteny conservation was lost across the *Mycena* assemblies. Based on differential synteny exhibited across genomes, we further partitioned the scaffolds into core and dispensable regions (Methods). **Fig. 3a** shows that the core regions were typically at the scaffold centres of the *Mycena* species. In contrast, the dispensable regions can extend to a few mega-bases, even to entire scaffolds, as was the case for the largest (12.0 Mb) assembled scaffold of *M. venus* (**Fig. 3a**). These dispensable regions are highly enriched in repeats; they have a 3–8-fold higher methylation level and are overrepresented in expanded and contracted OGs compared to the core regions (**Fig 3b**; Supplementary Table 11; two-proportions z-test, $P < 2.2E-16$). Expansions and contractions were 3.5–6.7 and 2–3.5-fold higher in the dispensable regions, respectively. The observation that half the relative sequence coverage was independent to other scaffolds suggests that some of these dispensable scaffolds have aneuploidy (**Fig. 3c**). Two large scaffolds (>1Mb) displaying half the relative depth of coverage were present in *M. venus* (scaffold01; 12 Mb) and *M. sanguinolenta* (scaffold12; 4.5 Mb), both of which have larger genome sizes compared to other species. Interestingly, these scaffolds exhibit higher repeat content with reduced methylation level, as well as expanded proteins which resembles the lineage-specific chromosomes observed in some fungi[35]. For example, scaffold12 of *M. sanguinolenta* contains 32 copies of expanded ARD domain (PF03079) encoded by 31 proteins, which is involved in the methionine salvage pathway modulating cell division[36] (**Fig. 3d**). Together, these results suggest that some *Mycena* genomes possessed two-speed genomes similar to filamentous pathogenic fungi where genomes are compartmentalised into gene-dense and gene-sparse regions enriched in transposable elements[37]. Differential gain and loss of genes in the dispensable regions may have important implications for *Mycena*.

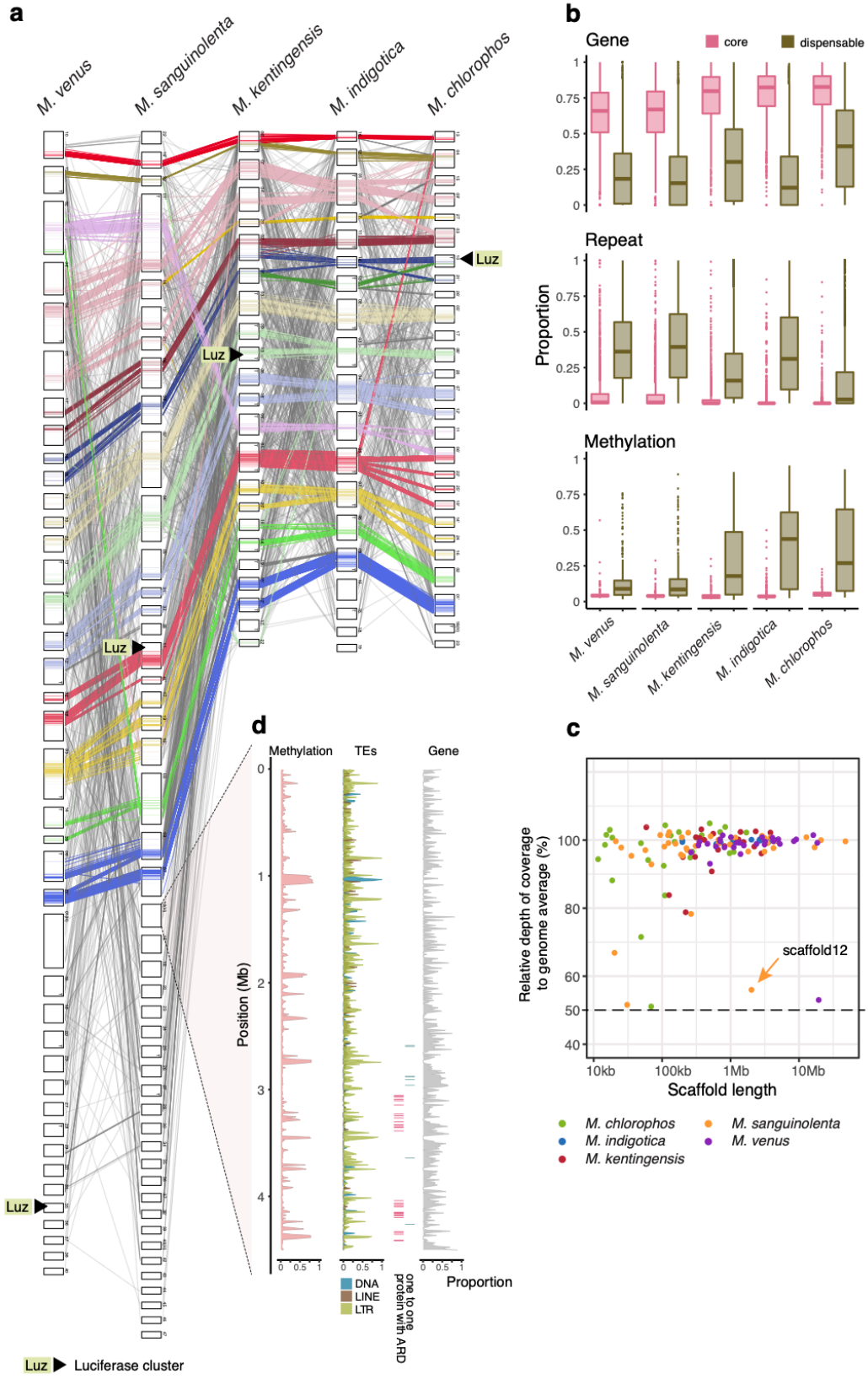


Fig. 3| The core and dispensable regions of scaffolds and anchoring for the four *Mycena* species scaffolds to *M. indigotica* reference genomes. a, Schematic representation of the inter-scaffold relationship between species. The links between scaffolds denoted the pairwise single-copy orthologues with closest related species. The core (coloured region in box) and dispensable regions in each scaffold were defined by the consecutive genes with less than 90 percent genetic distance to the adjacent one. The black triangle denotes the luciferase cluster. **b**, The gene density, repeat density and methylation level in core or dispensable regions. The gene and repeat densities are calculated from the non-overlapping 10-kb window located in the core or dispensable regions. The methylation level was calculated from the mean CG methylation level in core or dispensable regions. **c**, Sequence coverage was measured by the average number of reads that aligned to known reference bases per 10-kb non-overlapping windows with mosdepth[38]. Sequence coverage across scaffolds were normalized by the median coverage. The windows with more than 25% TEs were omitted. The coverage of each scaffold is normalized by the median coverage among the scaffolds longer than N90 in five species, except the Msan.scaffold36, which has 4-fold higher coverage than median coverage. **d,** Scaffold12 of *M. sanguinolenta* with expansion of proteins with the ARD domain. For every non-overlapping 10-kb window, the distribution from right to left is: (1) Density (percentage of nucleotides coverage) of gene. Green stripe denotes a gene containing single-copy orthologue with *M. venus*. Red stripe denotes the proteins with the ARD domain. (2) Density of transposable elements (TEs) including LTRs, LINES and DNA. (3) Average methylation level called from CpG sites per window.

Evolutionary dynamics of luciferase clusters

One of the previous challenges in addressing the evolution of fungal bioluminescence was that the bioluminescent species were scattered across the mycenoid and marasmioid clades, and this could only be explained by frequent loss of luciferase clusters. We used all the highly contiguous assemblies across the bioluminescent lineages in hand to investigate synteny around the luciferase cluster (**Fig. 4**). The genes surrounding the luciferase cluster among the eight *Armillaria* species generally shared the same order, but collinearity was partially lost in *G. necrorhiza* (**Fig. 4**). We reconstructed the ancestral upstream and downstream gene synteny of the *Armillaria* luciferase cluster and found that each belonged to different core regions of the same chromosome, suggesting that these regions were previously rearranged—including the luciferase cluster—and were subsequently retained (Supplementary Fig. 10).

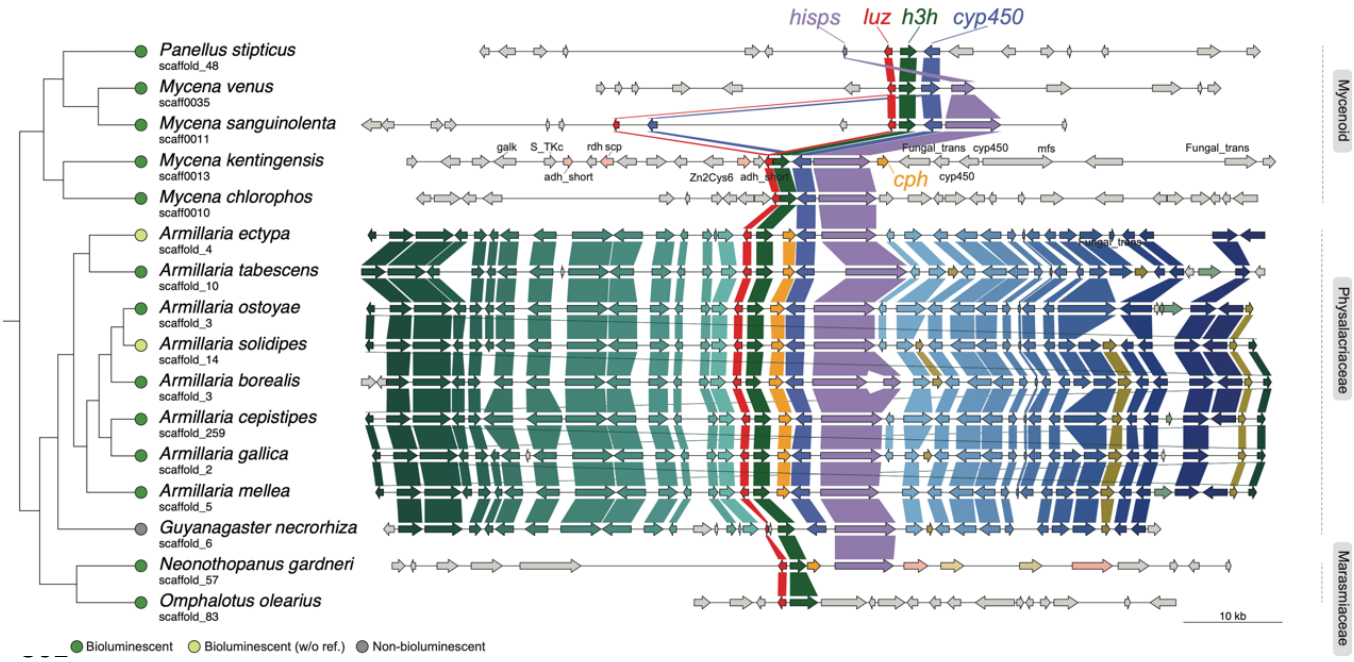


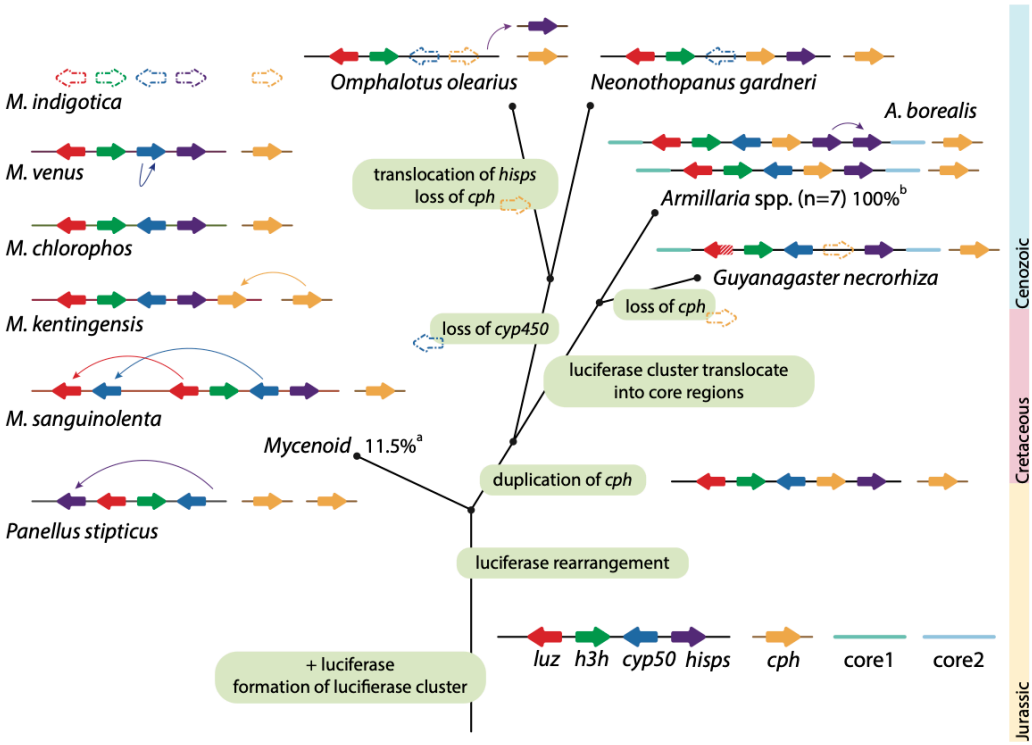
Fig. 4| Synteny around the luciferase cluster among bioluminescent fungi. The orthologous groups (OGs) shared by at least two species were labelled with the same colour, regardless of their orientation. The *cph* gene in some species was located in other scaffolds (Supplementary Fig. 9)

In contrast, the *Mycena* luciferase cluster was identified in the dispensable regions of different linkage groups (**Fig. 3a**), suggesting that the location of the cluster was extensively rearranged since the common ancestor of this lineage. A physical linkage was only maintained in the luciferase cluster, but no synteny was found in adjacent genes (**Fig. 4**). Variations were common amongst the markup of the luciferase cluster. The majority of the *Mycena* luciferase clusters included luciferase (*luz*), hispidin-3-hydroxylase (*h3h*), cytochrome P450 (*cyp450*), and hispidin synthase (*hisps*) (**Fig. 4**). caffeylpyruvate hydrolase (*cph*) was located in different scaffolds in four of the five *Mycena* species (Supplementary Fig. 9). In *M. sanguinolenta*, *luz* and *cyp450* were duplicated adjacent to the luciferase cluster (**Fig. 4**). Losses were observed in different positions of the phylogeny. The non-bioluminescent *M. indigotica* lost the entire luciferase cluster, but an *h3h* homologues were found in other regions of the genome, while *Guyanagaster necrorhiza* has a partial luciferase[6] and three other enzymes (**Fig. 4**), suggesting an independent loss of luciferase function alone was enough for it to lose its bioluminescence. Selection analysis of genes in the luciferase cluster revealed that the majority of conserved sites exhibit either no or strong purifying selection, with only 7–28 sites under episodic selection (Supplementary Fig. 11). Together, these results indicate limited but conserved roles of bioluminescence in the species that have retained

bioluminescence.

These observations lead us to propose a most plausible evolutionary scenario for luciferase cluster evolution across all available bioluminescent fungi genomes (**Fig. 5)**. The earliest ancestral state we could infer was a luciferase cluster having consisted of *luz*, *h3h*, *cyp450*, and *hisps*, with *cph*—involved in oxyluciferin recycling[5,6]—also present in the same chromosome. Such a combination was present in the 14 of the 15 bioluminescent species used in this study. The luciferase cluster likely originally formed at the dispensable region. Our model presents two contrasting scenarios by which the luciferase cluster was retained. First, the ancestor of the family Physalacriaceae had its luciferase cluster translocated into the core region. This scenario is based on the uniform observation that all members of *Armillaria* retained the luciferase cluster in their genome, and other members of Physalacriaceae had extended synteny in adjacent genes. The cases of extended synteny, however, were derived from different genomic regions in Agaricales, indicating rearrangements (Supplementary Fig. 10). Second, the luciferase cluster in bioluminescent *Mycena* fungi was conserved, despite being located in a highly rearranged genome partition (**Fig 3a**). Such rearrangements led to higher tendency of loss of luciferase cluster in the mycenoid lineage compared to *Armillaria* species. Additional evidence of rearrangement is that four out of five bioluminescent *Mycena* fungi had at least one gene rearrangement or duplication event compared to their ancestor. Interestingly, we found that *cph* was independently translocated adjacent to the luciferase cluster in both *M. kentingensis* and the ancestor of the marasmioid clade (Supplementary Fig. 12); it was presumably favoured and maintained here by natural selection[39].

Fig. 5| Evolutionary scenario for luciferase cluster evolution. The formation of the luciferase cluster originated at the dispensable region of the last common ancestor and was susceptible to translocate to different genomic locations through rearrangement. In the ancestor of marasmioid, *cph* was duplicated and translocated into the luciferase cluster. Before the ancestor of the Physalacriaceae family emerged, the luciferase cluster was translocated into the core region and have since kept its synteny in the *Armillaria* lineage. In the most recent common ancestor of *Mycena* species, the luciferase cluster was located in the dispensable region and have since been susceptible to further rearrangement. Arrow box indicates gene. The dashed arrow box denotes the loss of gene. Fishhook arrow denotes translocation event. [a] Percentage of bioluminescent fungi found in the mycenoid lineage[40]. [b] Percentage of bioluminescent fungi found in *Armillaria* lineage[41].



Expression profile of luciferase cluster and identification of conserved genes involved in fungal bioluminescence

Fungal bioluminescence is believed to have ecological roles, such as attracting insects and regulated by circadian rhythms[42]; however, the repertoire of genes involved in bioluminescence are still unknown. We carried out transcriptome profiling between mycelia with different bioluminescent intensities in four *Mycena* species, and identified genes that were either differentially expressed or positively correlated with bioluminescent intensities (Methods). There were 29 OGs found to contain upregulated gene members in all four *Mycena* species (**Fig. 1c and Fig. 6a**), including *luz*, *h3h*, and *hisps,* consistent with that bioluminescence expression depends on the expression of these three genes in the luciferase cluster. In particular, *luz* expression was significantly different between two tissues with relative high and low bioluminescence in *M. kentingensis* (log fold change (logFC) 3.0; adjusted $P< 0.001$) and *M. chlorophos* (logFC 4.7; adjusted $P<0.001$); there was also a significant correlation between bioluminescent intensity and expression level in *M. sanguinolenta* (Pearson's correlation coefficient (PCC) 0.82; $P<0.005$) and *M. venus* (PCC 0.86; $P<0.005$ Supplementary Table 12). In *M. chlorophos*, however, its *cyp450* and *h3h* in the luciferase cluster were not differentially expressed, and four distant homologues of *h3h* were found to be upregulated (Supplementary Fig. 5b). Although a second copy of *luz* and *cyp450* were found in *M. sanguinolenta*, they showed much lower expression (2 and 3 transcripts per million (TPM), respectively) than those in the cluster (282 and 138

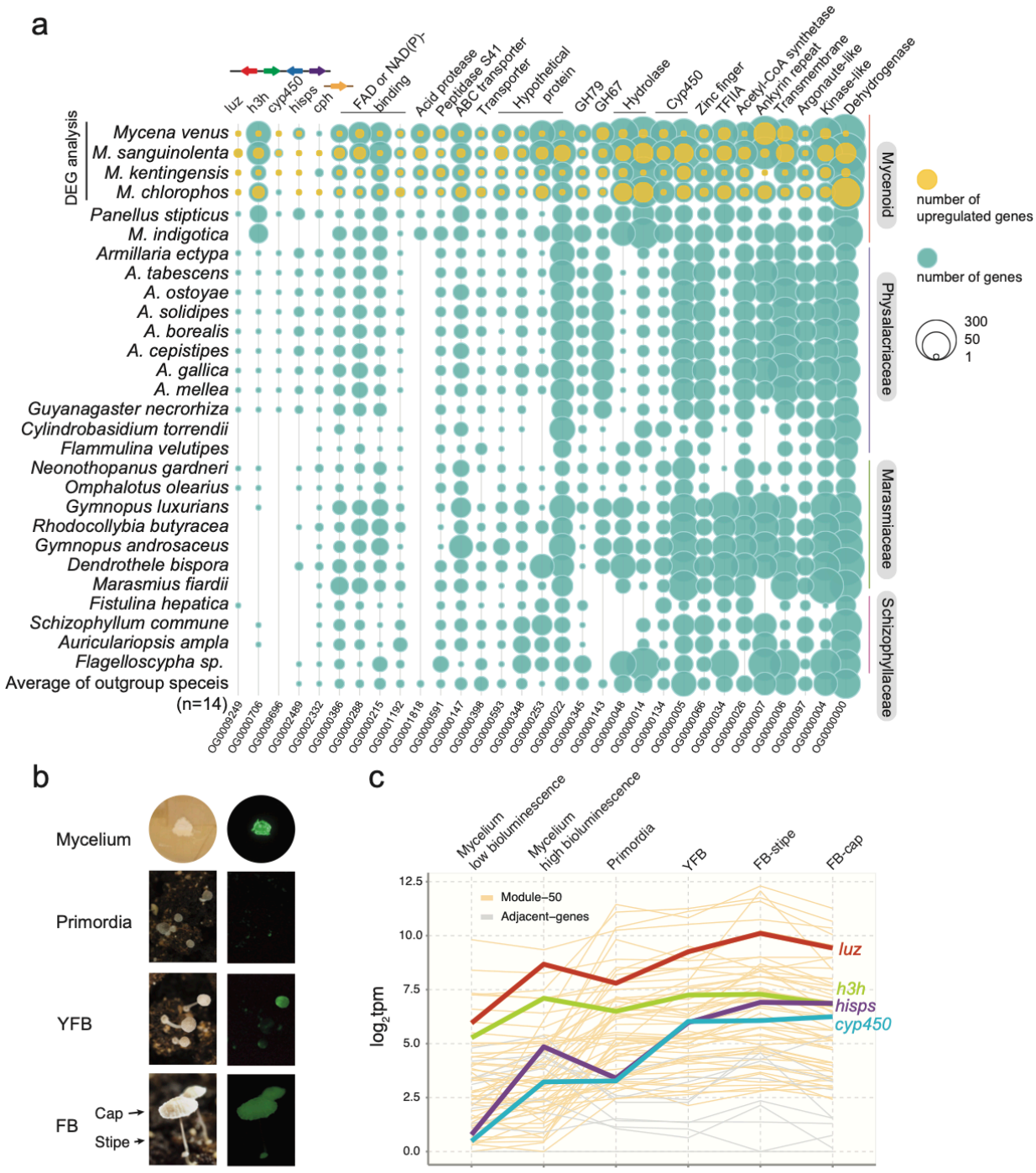
TPM, respectively). The remaining OGs upregulated in mycelia showing higher bioluminescence included genes involved in ABC transporters and Aceyl-CoA synthetases which also showed a predicted function in metabolic adaptations to bioluminescence in firefly and glowworm[43,44]. (**Fig. 6a**; Supplementary Table 13). In particular, four OGs were annotated as FAD or NAD(P)-binding domain-containing proteins. As these genes do not bear sequence similarity to *h3h* which is also a NAD(P)H-dependent enzyme, they may likely involve in other biochemical processes that is required during bioluminescence.

Differences in bioluminescent intensity have been recorded in tissues of fungi both in nature[4,40,41,45,46] and—for *M. kentingensis*—in a laboratory environment, in which the life cycle can be completed (**Fig. 6b**). To investigate putative roles of bioluminescence across developmental stages, additional transcriptome profiling was carried out in the primordia, young fruiting body, and cap (pileus) and stipe of the mature fruiting body of *M. kentingensis*. Bioluminescence was stronger in the cap than in the stipe, so we expected the luciferase cluster genes to have higher expression in the cap tissue. However, *luz* and *h3h* showed opposite expression patterns (**Fig. 6c** and Supplementary Table 14), suggesting that there may be other regulators involved in bioluminescence in *M. kentingensis*.

The regulation of bioluminescence in *M. kentingensis* during development was determined by performing a weighted correlation network analysis (WGCNA[47,48]), which identified 67 modules of co-expressed genes in these stages (Supplementary Fig. 14). All members of the luciferase cluster *luz*, *h3h*, *cyp450*, and *hisps* belonged to the same module (Module50; **Fig. 6c**) of 57 genes, suggesting that the expression of the luciferase cluster members are orchestrated during developmental stages. Only two genes belonging to OG0001818 (acid protease) and OG0000000 (short-chain dehydrogenase) which were part of the 29 aforementioned OGs involved in bioluminescence of the mycelium across *Mycena*. Six genes in this module were annotated as carbohydrate-active enzymes (Supplementary Table 15): one GH75 (chitosanase), one AA1_2 (Laccase; ferroxidases), two GH16, and two genes with two CBM52. GH16 (glucanases) and AA1 (laccases) are known to be regulated during fruiting body development[49], implying a possible link between cell wall remodelling during development and bioluminescence. In addition, we re-analysed the transcriptomes of *Armillaria ostoyae* across different developmental stages from Sipos *et al*. (2017)[8]. Consistent with observation that bioluminescence was only observed in mycelia and rhizomorphs in *A. ostoyae*[50,51], the expression of *luz*, *h3h*, *cyp450*, and *cph* showed highest expression before the primordia stage (Supplementary Fig. 15).

Together, these results imply that luciferase cluster was differentially regulated during developmental stages; extent of regulation were also different in bioluminescent species of different lineages.

Fig. 6| Expression analysis to identify genes involved in bioluminescence. a, Conserved upregulated OGs. Differentially-expressed genes (DEGs) between mycelia with different bioluminescent intensities were identified in four bioluminescent *Mycena* species, and all 29 OGs—except OG0009249 and OG0000706—contain at least one upregulated gene. A detailed annotation of the genes in the OGs is listed in Supplementary Table 13. The raw gene number in each OG is shown in Supplementary Fig. 13. **b**, Tissues used for transcriptomic data analysis in *M. kentingensis*. The left and right side are the tissues under light and dark conditions, respectively (captured by a Nikon D7000). The camera setting for each tissue: mycelium, Sigma 17-50mm ISO100 f2.8 with 16 min exposure time; primordia, AF-S Micro Nikkor 60mm ISO800 f/11with 122.4 sec exposure time; YFB, AF-S Micro Nikkor 60mm ISO800 f/11with 60.6 sec exposure time; FB, AF-S Micro Nikkor 60mm ISO800, f/11 with 9.3 sec exposure time. YFB, young fruiting body (0.5-1 cm). FB, mature fruiting body (> 1 cm). FB-cap, cap from FB. FB-stipe, stipe from FB. **c**, Expression profile of luciferase cluster across developmental stages of *M. kentingensis*. Bold lines indicate four genes in the luciferase cluster. These four genes and the other 53 genes (yellow) were assigned into the same module (Module50) with similar expression patterns. The genes located up- or downstream (grey) of the luciferin biosynthesis cluster had lower expression levels than the four genes in the cluster.



450

451

452 **Gene families associated with the evolution of mycenoid species**

453 We sought to identify proteins specific to mycenoid species by annotating protein
454 family (Pfam) domains and comparing them with those of species outside this lineage
455 (Supplementary Table 16). A total of 537 Pfam domains were enriched in the mycenoid
456 lineage (one-fold by Wilcoxon rank sum test with P<0.01; Supplementary Table 17) of

which 3–17 were species-specific. Acyl_transf_3 (acyltransferase family; PF01757), contained in a range of acyltransferase enzymes, was the only domain found in all six mycenoid species. The closest homologs were found in ascomycetous *Cadophora*, *Pseudogymnoascus*, or *Phialocephala* (31-35% identity with 73-100% coverage). Four of the enriched domains are known pathogenesis-related domains expanded in pathogenic Agaricales *Moniliophthora*[52] and *Armillaria* species[8]: COesterase (PF00135; Carboxylesterase family), Thaumatin (PF00314), NPP1 (PF05630; necrosis-inducing protein), and RTA1 (PF04479; RTA1-like protein) (Supplementary Fig. 16). Moreover, *M. sanguinolenta* and *M. venus* contained over 100 and 17 copies of COesterase and Thaumatin (median 37 and 4 copies in other fungal species of this study), respectively.

Wood decay gene repertoires in *Mycena* species

Since studies on wood degradation in *Mycena* are lacking[53], we annotated the carbohydrate-active enzyme (CAZyme) repertoire of *Mycena* proteomes to understand their ecological role. Class II peroxidases (AA2) were identified in all *Mycena* species, suggesting these species can degrade lignin and leave white coloured cellulose (white rot) behind[54]. In particular, *M. sanguinolenta* and *M. venus* possess 66 and 72 copies of AA2, respectively, which are much higher than in other closely-related white-rot fungi, such as *Armillaria* species (10–14 copies). *Mycena* species also contain 11–20 copies of AA5_1 (copper radical oxidases), 18–121 copies of AA3_2 (GMC oxidoreductase), and 10–34 copies of AA1_1 (Laccases)[55]. These enzymes are capable of degrading or modifying lignin and are generally in higher abundance in white-rot than in brown-rot fungi[55,56]. For cellulose degradation, white-rot generally has more cellulolytic genes—i.e., GH6 and GH7—and lytic polysaccharide monooxygenase of AA9 than does brown-rot fungi[55]. Our *Mycena* species have similar numbers of these genes as other white-rot fungi (Supplementary Fig. 17).

Discussion

Bioluminescence is one of the most unusual and fascinating traits in fungi, but the evolutionary history of luciferase cluster remains elusive. Here, we produced highly contiguous genome assemblies using Nanopore technology and annotations for five of the *Mycena* species to examine their genome dynamics. The results of phylogenomic analyses on these genomes have important implications for the origin of luciferase. We argue that luciferase originated in the last common ancestor of mycenoid and marasmioid, which disagrees the theory of Kotlobay *et al.*[6], that it originated at the basal clade of Agaricales. Our results are in good general agreement with comparative

genomic analyses around this group[8], and is the more evolutionarily parsimonious scenario that does not require an extensive loss of luciferase across the entire Agaricales lineage (**Fig. 1b** and **1c**).

Until 2019, only 68 species of bioluminescent fungi were known[40] across over 600 non-bioluminescent ones in Mycenaceae. Our reconstructed evolution of the luciferase cluster model shows that the luciferase cluster originated in the dispensable region of genomes (**Fig. 3**), making it susceptible to rearrangement, which suggests it is highly prone to loss and explains why most mycenoid species are non-bioluminescent. This is consistent with a previous report that the main evolutionary process in fungal gene clusters is vertical evolution followed by differential loss[57]. Interestingly, synteny was retained in luciferase clusters and adjacent genes of *Armillaria* species (**Fig. 4**), which are better known for their roles as plant pathogens[8]. Indeed, bioluminescence was identified in all nine examined *Armillaria* species[41]. We speculate that the majority of unquantified *Armillaria* species may also exhibit bioluminescence. The repeated duplication and relocation of *cph* that we observed in the luciferase cluster is under selection pressure, suggesting that bioluminescence has adaptive importance in these species. A systematic quantification of bioluminescence and more complete genome assemblies will help reconstruct the evolutionary events that contributed to the polymorphism and functional diversity in the luciferase clusters.

Chromosome-length polymorphisms are uncommon in Agaricales but have been observed in the pathogen *Moniliophthora perniciosa*[58] causing witches' broom disease and model mushroom *Coprinopsis cinerea*[59]. Synteny analyses across *Mycena* genomes revealed the first documented case of 'two speed' genomes in the Agaricales. Despite all the studied *Mycena* species are saprotrophs, the compartmentalization of genomes exhibiting contrasting evolutionary scenarios and the presence of lineage-specific chromosomes are hallmarks of plant pathogens throughout fungal tree of life[60]. We argue that in addition to gene family expansion and adaptation due to diversifying lifestyle, the increased rates of evolution in particular regions have also led to higher propensity to the loss of luciferase cluster in the mycenoid lineage. Two of the five *Mycena* species had large genomes, which was a result of a transposable element expansion and gene family dynamics (**Fig. 1d**). In plants, animals and fungi, DNA methylation is known to suppress TE activity[61,62]. In *Mycena*, we observed numerous relic TEs containing sequences similar to classic TEs but lacking functional domains, and these relic TEs had lower levels of CG methylation than did the classic TEs (**Fig. 2b**). We hypothesize that the relic TEs were still repressed by DNA methylation, yielding overall lower levels of CG methylation in active TEs, which consequently

resulted in genome expansion and increased dispensable region offering more plasticity in their evolution.

Researchers have long been puzzled over the ecological role of bioluminescence in fungi. One explanation that has been put forth for *Neonothopanus gardneri* is that bioluminescence follows a circadian rhythm to increases spore dispersal by attracting arthropods in the evening[42]. If true, this is most likely a derived adaptation, as most Agaricomycetes like *Omphalotus nidiformis* disperse spores via wind, display bioluminescence continuously, and do not attract insects[63]. Besides, attraction is insufficient to explain luminescence in the mycelium. *Mycena kentingensis* and *Armillaria ostoyae* are thought to constitutively express the luciferase cluster throughout development[64]. In addition to the luciferase cluster, only a handful of genes were identified to associate with fungal bioluminescence (**Fig6**). If fungal bioluminescence originated as a by-product of a biological process that is currently unknown, the ecological role was likely to be initially limited and has undergone subsequent losses in many species, especially in the *Mycena* lineage. For those that have retained bioluminescence, we speculate that its ecological role may be species-specific and together converge on the maintenance of luciferase cluster.

In summary, our comparative analyses have led us to propose an evolutionary model pinpointing changes in the luciferase cluster across all published bioluminescent species. Our findings offer insights into the evolution of a gene cluster spanning over 160 million years and the retained luciferases were under strong purifying selection. Our *Mycena* genome sequences may complement ongoing research on the application of bioluminescent pathways[6] and shed light on the ecological role of bioluminescence in fungi.

Methods

Strains and fungal materials

M. kentingensis, *M. venus, M. sanguinolenta, M. indigotica* and *M. chlorophos* were isolated from fruiting bodies collected from forest in Taiwan. *M. indigotica* was isolated from basidiospores. The mycelia were grown and maintained on potato dextrose agar (PDA) plates at 25°C. To identify the pattern of bioluminescence, a piece of mycelium from each species was inoculated in the centre of a sheet of sterilized dialysis cellulose membrane (8030-32, Cellu Sep T-Series) on a 3 cm PDA agar plate at 25°C. The diameter of the mycelium was measured and its

bioluminescence was recorded with a Glomax 20/20 luminometer (Promega BioSystems Sunnyvale, Inc., USA) for seven days (Supplementary Table18). The taxonomic status of species was reconfirmed by sequencing the internal transcribed spacer (ITS) with the primer pair SR6R(5'-AAGWAAAAGTCGTAACAAGG-3')/ITS4(5'-TCCTCCGCTTATTGATATGC-3'). Using the other available *Mycena* ITS sequences, all sequences were aligned by MAFFT[65] (ver. 7.310) and trimmed by trimAl[66] (1.2rev59; with option -automated1). The ITS phylogeny was constructed by IQ-TREE[67,68] (ver. 1.6.10; with option -bb 10000 -alrt 1000).

Genomic DNA extraction and sequencing

Genomic DNA was extracted using the traditional CTAB and chloroform extraction method. Briefly, 0.1 g mycelium was grinded with liquid nitrogen and then mixed with CTAB extraction buffer (0.1 M tris, 0.7 M NaCl, 10 mM EDTA, 1% CTAB, 1% Beta-Mercaptoethanol). After incubating at 65°C for 30 min, an equal volume of chloroform was added, then the mixture was centrifuged at 8000 rcf for 10 min. The supernatant was mixed with an equal volume of isopropanol and the DNA was precipitated. After washing with 70% EtOH, the DNA was dissolved with nuclease free water. Genome sequencing was carried out in two platforms. First, paired-end libraries were constructed using the KAPA LTP library preparation kits (#KK8232, KAPA Biosystems). All libraries were prepared in the High Throughput Genomics Core at Biodiversity Research Center, Academia Sinica and sequenced on an Illumina HiSeq 2500 platform. A total of 51.6 Gb of 150- or 300-bp read pairs were generated. Second, Oxford Nanopore libraries were prepared using SQK-LSK108 and sequenced on a GridION instrument. Basecalling of Nanopore raw signals was performed using Guppy (ver. 3.2.4) into a total 67.7 Gb of raw sequences at least 1 kb or longer. A summary of the sequencing data is shown in Supplementary Table 1.

RNA extraction and sequencing

Bioluminescent mycelia were collected in two ways. i) For *M. chlorophos* and *M. kentingensis*, a piece of mycelium was inoculated at the centre of a sheet of sterilized dialysis cellulose membrane (8030-32, Cellu Sep T-Series) on PDA agar plates at 25°C. The plates were cultured for 10 and 14–18 days for *M. chlorophos* and *M. kentingensis*, respectively. For *M. kentingensis*, bioluminescence was detected by camera (Nikon D7000, Sigma 17-50mm ISO100 f2.8 with 16 min exposure time) (**Fig. 6b**). The mycelia with low or high bioluminescent intensities which occurred spontaneously were collected from two separated plates inoculated on the same day. In *M. chlorophos*, bioluminescence was detected by luminometer. Mycelium with low bioluminescence showed the intensity of 7-14 Relative Light Unit (RLU)/mg, and the mycelium with

high bioluminescence showed the intensity of 5,000-10,000 RLU/mg (Supplementary Table 3). Three replicates were collected. ii) For *M. sanguinolenta* and *M. venus*, a piece of mycelium was inoculated at the centre of a sheet of sterilized dialysis cellulose membrane on PDA agar plates at 25°C. The plates were cultured for 13–17 days, and the bioluminescent features were detected by CCD camera; the tissues were collected and their luminescence intensity was recorded with a Glomax 20/20 luminometer (Promega BioSystems Sunnyvale, Inc., USA). A total of 12 samples with different bioluminescence intensities were collected (Supplementary Table 3). After homogenizing 5–10 mg of tissues by liquid nitrogen, total RNA was extracted using the Direct-zol RNA Miniprep (Zymo Research). Concentrations were measured by Qubit fluorometer (Invitrogen USA), and quality was assessed by the BioAnalyzer 2100 RNA Nano kit (Agilent, USA) with RIN values higher than 8.0. The paired-end libraries were constructed using the TruSeq Stranded mRNA library prep kit (#20020594, Illumina, San Diego, USA) with standard protocol and sequenced by Illumina HiSeq 2500 (Illumina, USA) to produce 150-bp paired-end reads.

RNA extraction and sequencing from the *M. kentingensis* fruiting body

Fruiting body production of *M. kentingensis* was modified from previous studies[69,70]. Mycelia, grown on PDA for 8–15 days, was then inoculated onto sterilized commercially available peat soil mixed with 10% rice bran and 50% water in a jar. Mycelium samples were grown at 25°C for 3–4 weeks and then transferred into fresh compost. The culture was sprayed with sterilized water daily until the fruiting body formed. Four kinds of tissue were collected: (1) primordia, (2) young fruiting body (YFB, 0.5–1 cm), (3) cap and (4) stipe of mature fruiting body (> 1 cm). For each batch of culture, 15–20 primordia, 6–11 YFB, and 8–12 caps and stipes from mature fruiting bodies were pooled to measure their weight and bioluminescent intensity, and the RNA was extracted using Trizol extraction and lithium chloride purification method. Three replicates were produced. The paired-end libraries were constructed using the TruSeq Stranded mRNA library prep kit (#20020594, Illumina, San Diego, USA) with standard protocol and sequenced by Illumina HiSeq 2500 (Illumina, USA) to produce 150-bp paired-end reads.

***De novo* assemblies of *Mycena* species**

Oxford Nanopore reads were assembled using the Canu[71] (ver. 1.8) assembler. Consensus sequences of the assemblies were polished first by five iterations of Racon[72] (ver. 1.3.2) followed by Medaka (ver. 0.7.1; https://github.com/nanoporetech/medaka) using Oxford Nanopore reads. HaploMerger2[73] (ver. 20180603) was then run on to generate haploid assemblies. Finally, the consensus sequences were further corrected

with Illumina reads using Pilon[74] (ver. 1.22). Throughout each stage the genome completeness was assessed using fungi and basidiomycete dataset of BUSCO[15] (ver. 3.0.2). Putative telomeric repeats were searched for copy number repeats less than 10 mers using tandem repeat finder[75] (ver. 4.09; options: 2 7 7 80 10 50 500). The hexamer TTAGGG was identified (Supplementary Table 19).

Gene predictions and functional annotation

Protein sequences from Uniprot fungi (32,991 sequences; downloaded 20th December 2018) and *Coprinopsis cinerea, Pleurotus ostreatus* PC15 (v2.0), *Schizophyllum commune and Armillaria mellea* from MycoCosm[76] portal were downloaded as reference proteome. Transcriptome reads were first mapped to the corresponding genome assemblies using STAR[77,78] (ver. 2.5.3a), and subsequently assembled into transcripts using Trinity[79] (ver. 2.3.2; guided approach), Stringtie[80] (ver. 1.3.1c), CLASS2[81] (ver. 2.1.7) and Cufflinks[82] (ver. 2.2.1). The samples used for input are listed in Supplementary Table 3. Transcripts generated from Trinity were aligned to the references using GMAP[83]. All transcripts were merged, filtered and picked using MIKADO[84] (ver. 1.1). The gene predictor Augustus[85] (ver. 3.2.1) and gmhmm[86] (ver. 3.56) were trained using BRAKER2[87] (option fungi and softmasked), and SNAP[88] was trained using the assembled transcripts with MAKER2[14] (ver. 2.31.9). The assembled transcripts, reference proteomes and BRAKER2 annotations were combined as evidence hints for input in the MAKER2[14] annotation pipeline. MAKER2[14] invoked the three trained gene predictors to generate a final set of gene annotation. Amino acid sequences of the proteome were functionally annotated using Blast2GO[89] and eggnog-mapper[90] (ver. 1.0.3).

Identification of repetitive elements

Consensus (library) sequences of repetitive elements were identified using the pipeline described in Berriman *et al*[91]. Full LTR retrotransposons in *Mycena* species were defined as i) initially identified by LTRharvest[92] and ii) presence of known reverse transcriptase domains identified by Pfam[93] (ver. 31.0). Repeat contents were quantified using RepeatMasker[94] (ver. open-4.0.7). Proportions of repeat content along the scaffolds were calculated using Bedtools[95]. A phylogenetic tree was built by first aligning all the putative RVT domain sequences using MAFFT[65] (ver. 7.310; --genafpair --ep 0) and FastTree[96] with the JTT model on the aligned sequences, and were visualised using the ggtree[97] package in R.

Methylation analyses

To construct a BS-seq library, the fragmented DNA was first ligated with a premethylated TruSeq DNA adapter (Illumina). The ligated DNA fragments were bisulfite converted using the EZ DNA methylation kit (Zymo Research), followed by PCR amplification. The BS-seq libraries were sequenced on an Illumina HiSeq 2500 sequencer. The bisulfite conversion efficiency reached approximately 99% in all of our libraries (Supplementary Table 20). High-quality paired-end reads were aligned to the genome assemblies of *M. kentingensis* using the bisulfite specific aligner BS-Seeker2[98]. Only uniquely mapped reads were retained. The cytosines covered by at least four reads were included in the data analysis, and the DNA methylation level for each cytosine was estimated as #C/(#C+#T), where #C is the number of methylated reads and #T is the number of unmethylated reads.

One or two Nanopore flowcells for each Mycena species were selected to infer methylation information using deepsignal[99] (ver. 0.1.5) (*M. kentingensis*: FAH31207, *M. chlorophos*: FAH31470, *M. indigotica*: FAH31228, *M. sanguinolenta*: FAK22405 and FAH31211, *M. venus*: FAK22389 and FAH31302). The machine learning-based model was trained with one bisulfite dataset (YJMC0389) and one Nanopore dataset (FAH31207) of *M. kentingensis*. The bisulfite result was first filtered for depth >20, then methylation levels >0.9 and <0.01 were selected for positive and negative validation datasets, respectively. All seven flowcells were called for methylation information with a customized model and default arguments. A minimal depth of 4 was applied to the results for further analysis. In the estimates of DNA methylation levels between Nanopore long-reads and the Illumina BS-seqs, the Pearson correlation coefficient was as high as 0.96 in the methylomes of *M. kentingensis* (Supplementary Fig. 18).

Orthogroup inference and analysis of protein family domains

Orthologous groups (OGs) among 42 species were identified using OrthoFinder[16,17] (ver. 2.2.7). CAFÉ[18] (ver. 4.2.1) was used to predict the expansion and contraction of gene numbers of OGs based on the topological gene tree. Unique OGs among the node identified by CAFÉ were analysed by UpSetR[100]. The phylogenetic tree was visualized by the ggtree[97,101] package in R. Protein domains of each gene were identified by pfam_scan.pl ver. 1.6 by comparing them against Pfam ver. 32.0 db[93]. To compare them to plant pathogenic fungi, the Pfam domains from *Moniliophthora perniciosa* FA55313 (Monpe1_1)[52] from JGI and *Moniliophthora roreri* (Monro) from BioProject: PRJNA279170 were also annotated. Enrichment of Pfam domain number between two sets of interest was assessed by the Wilcoxon rank-sum test ($P \leq$

0.05). We compared the Pfam copy number between six mycenoid species and the other 37 species. Gene ontology enrichments were identified for these genes using TopGO[102]. Genes encoding carbohydrate-active enzymes were identified according to the Carbohydrate-Active enZYmes (CAZy) Database[103] by searching for sequence homologs with DIAMOND[104] or HMMER[105].

Phylogenomic analyses

A total of 42 sets of amino acid sequences from 360 single-copy OGs were aligned independently using MAFFT[65] (ver. 7.271; option --maxiterate 1000). A total of three approaches were used to infer the species tree. The first two approaches relied on maximum likelihood phylogenies from individual OG alignments computed using RAxML-ng[106] (ver. 0.9.0; options: --all --model LG+I+F+G4 --seed 1234 --tree pars 10 --bs-trees 100) with 100 bootstrap replicates. The best phylogeny and bootstrap replicates were separately used to infer a consensus tree using ASTRAL-III[20]. Finally, a maximum likelihood phylogeny from the concatenated amino acid alignments of the single-copy orthogroups was constructed with 100 bootstrap replicates using RAxML-ng[106] (ver. 0.9.0; options: --all --seed 1234 --tree pars 10 --bs-trees 100 with --model LG+I+F+G4 partitioned with each OG alignment).

Estimation of divergence time

The divergence time of each tree node was inferred using MCMCtree in PAML[21] package (ver. 4.9g with approximate likelihood[107]; the JC69 model and the rest were default). The species tree and concatenated translated nucleotide alignments of 360 single-copy-orthologs were used as the input for MCMCtree. The phylogeny was calibrated using fossil records by placing soft minimum bounds at the ancestral node of: i) marasmioid (using *Archaeomarasmius legettii* 94–90 Ma[108]; 90 was used), ii) Agaricales (using *Palaeoagaricites antiquus* 110–100 Ma[109]; 100 was used), iii) Taxon A (~99 Ma[110]; 95 was used), and iv) a soft bound of 200 Ma for the phylogeny. The entire analysis was run five times to check for convergence.

Synteny analyses

Linkage groups (LGs) between *M. indigotica* and *Armillaria ectypa,* and between *M. indigotica* and *Pleurotus ostreatus* were assigned based on the majority of the single-copy orthologues (Supplementary Figure 7 and 8). Scaffolds containing fewer than 10 single-copy orthologues, shorter than 500 kb or species N90 were excluded. Linkage groups within *Mycena* were assigned based on majority and at least 10% of single-copy orthologue links with *M. indigotica* scaffolds. Subsequent scaffolds were identified as

the same linkage group if they contained a majority of pairwise one-to-one single-copy orthologues belonging to the *M. indigotica* LG.

As gene collinearity among *Mycena* species became less conserved, synteny blocks of each *Mycena* species were defined based on merging of adjacent pairwise single-copy orthologues to its closest-related species. For instance, synteny blocks of *M. chlorophos* were based on single-copy orthologues against *M. indigotica*. For every ortholog, the distance to the next closest single-copy orthologue was calculated to take into account segment duplications of genes or gene insertion/deletions. Synteny blocks within the 90% quantile of ortholog distance were further merged and classified as core regions. Dispensable regions were defined as synteny breaks under this criterion. Based on these criteria, 68.5–85.5% of the genomes of *Mycena* species were dispensable (Supplementary Table 21). Synteny around luciferase cluster was based on orthogroup sharing and plotted using the R genoPlotR[111] package.

Evolution of gene families related to the luciferase gene cluster

The sequences of five orthologues in the luciferase family—hispidin-3-hydroxylase, cytochrome P450, hispidin synthase, and caffeylpyruvate hydrolase—were constructed and the sequences were aligned by MAFFT[65] (ver. 7.310) and trimmed by trimAl[66] (1.2rev59 ; with option -automated1). The protein trees were constructed by IQ-TREE [67,68] (ver. 1.6.10; with option -bb 10000 -alrt 1000). The evidence for selection across gene families was tested using the HyPhy[112,113] platform in the webserver of datamonkey[114-116]. According to the recombination breakpoints analysed by the Genetic Algorithm for Recombination Detection[117] (GARD), the alignment was trimmed for analysing selection using Single-Likelihood Ancestor Counting[118] (SLAC) and Mixed Effects Model of Evolution[119] (MEME) with $P<0.1$.

RNAseq analysis of differential bioluminescent mycelium

Quality trimming of the RNA sequencing reads was conducted using Trimmomatic[120]. The sequencing reads were mapped to the genome using STAR[77,78] (ver. STAR_2.5.1b_modified; default parameters). Raw read counts of the gene models were quantified by FeatureCounts[121] (ver. v1.5.0; -p -s 2 -t exon). For *M. kentingensis* and *M. chlorophos*, the differential expressed genes (DEGs) were analysed using DESeq2[122]. Genes with fold change (FC) > 0 and FDR ≤ 0.05 were defined as DEG. *For M. sanguinolenta* and *M. venus*, the DEGs were identified by the Pearson correlation coefficient between the bioluminescence intensity (relative light unit; RLU) normalized

by weight (RLU/mg) and log transformation of counts per million. Genes with correlation coefficient > 0.7 and $P<0.01$ were defined as DEGs.

RNA analysis of *M. kentingensis* and *Armillaria ostoyae* developmental stages

The reads from transcriptomes of the primordia, young fruiting body, and cap and stipe of mature fruiting body were conducted by the same method of manipulating the reads from transcriptomes of mycelium. To identify co-expressed genes among transcriptomes, the transformation of transcripts per million (TPM) from six different tissues—mycelia with high bioluminescence and low bioluminescence, primordia, young fruiting body, and fruiting body cap and stipe were calculated. The lowest 25% expressed gene across all samples were excluded and co expression was analysed using weighted gene co-expression network analysis (WGCNA)[47,48] package in R (maxBlockSize = 10000, power = 20, networkType = signed , TOMType = signed, minModuleSize = 30). The Illumina reads among ten stages from *Armillaria ostoyae* were also downloaded from NCBI's GEO Archive (http://www.ncbi.nlm.nih.gov/geo under accession GSE100213) and also analysed by the same pipeline of *M. kentingensis* to identify co-expressed genes among the transcriptomes.

Authors contribution

I.J.T. and H.M.K. conceived the study. I.J.T. led the study. H.M.K., C.C.C., G.S. and H.W.K. collected and identified *Mycena* species around Taiwan. H.M.K, P.H.W. and C.I.L. conducted the experiments. M.J.L. and J.Y.L. designed the illumina sequencing experiment. H.H.L. and I.J.T. performed the assemblies and annotations of the *Mycena* genomes. H.M.K., H.H.L., Y.C.L. and I.J.T. conducted the repeat analysis. L.C.N. and I.J.T. carried out phylogenomics analyses and the divergence time estimation. H.M.K., H.H.L., Y.C.L. and I.J.T. carried out comparative genomic analyses. H.M.K. and M.R.L. analysed the RNA-seq data. H.H.L., R.J.L., J.W.H., P.Y.C. and H.M.K. carried out the methylation analyses. H.M.K. and I.J.T. wrote the manuscript with input from L.C.N and P.Y.C.

Acknowledgement

We thank Chia-Ning Shen for lending us the luminometer for the duration of the project. We thank Bi-Chang Chen for providing the CCD camera to qualify bioluminescence. We thank Chi-Yu Chen and Jie-Hao Ou for their useful advice on culturing *Mycena* fungi. We are grateful to the '1000 Fungal Genomes – Deep Sequencing of Ecologically-relevant Dikarya' consortium for access to unpublished genome data. We thank Gregory Bonito, Hui-Ling Liao, Alejandro Rojas and Rytas Vilgalys for permission to use the *Flagelloscypha* sp. FlaPMI526_1 assembly from JGI. We thank Mary Catherine Aime for permission to use the *G. necrorhiza* MCA 3950. assembly from JGI. The genome sequence data were produced by the US Department of Energy Joint Genome Institute (JGI) in collaboration with the user community. We thank Chia-Lin Chung, Ben-Yang Liao and John Wang for commenting the earlier version of the manuscript. P.Y.C was supported by Ministry of Science and Technology, Taiwan, under Grant No. 106-2311-B-001 -035 -MY3 and 108-2313-B-001 -013 -MY3. H.M.K was supported by postdoctoral fellowship, Academia Sinica. I.J.T was supported by Career Development Award AS-CDA-107-L01, Academia Sinica.

Data availability

Genome assembly and annotation of five *Mycena* species was deposited in the National Centre for Biotechnology Information BioProject database (accession no. PRJNA623720) pending final checks.

References

1 He, M. *et al.* Notes, outline and divergence times of Basidiomycota. *Fungal Diversity*, doi:10.1007/s13225-019-00435-4 (2019).

2 Wei, C. & Kirschner, R. A new Mycena species with blue basidiomata and porioid hymenophore from Taiwan. *Mycoscience* **60**, 10-13, doi:10.1016/j.myc.2018.06.001 (2019).

3 Thoen, E. *et al.* In vitro evidence of root colonization suggests ecological versatility in the genus Mycena. *New Phytologist*, doi:10.1111/nph.16545 (2020).

4 Chew, A., Desjardin, D., Tan, Y., Musa, M. & Sabaratnam, V. Bioluminescent fungi from Peninsular Malaysia-a taxonomic and phylogenetic overview.

Fungal Diversity **70**, 149-187, doi:10.1007/s13225-014-0302-9 (2015).

5 Kaskova, Z. M. *et al.* Mechanism and color modulation of fungal bioluminescence. *Sci Adv* **3**, e1602847, doi:10.1126/sciadv.1602847 (2017).

6 Kotlobay, A. A. *et al.* Genetically encodable bioluminescent system from fungi. *Proc Natl Acad Sci U S A* **115**, 12728-12732, doi:10.1073/pnas.1803615115 (2018).

7 Varga, T. *et al.* Megaphylogeny resolves global patterns of mushroom evolution. *Nat Ecol Evol* **3**, 668-678, doi:10.1038/s41559-019-0834-1 (2019).

8 Sipos, G. *et al.* Genome expansion and lineage-specific genetic innovations in the forest pathogenic fungi Armillaria. *Nat Ecol Evol* **1**, 1931-1941, doi:10.1038/s41559-017-0347-8 (2017).

9 Floudas, D. *et al.* Evolution of novel wood decay mechanisms in Agaricales revealed by the genome sequences of Fistulina hepatica and Cylindrobasidium torrendii. *Fungal Genet Biol* **76**, 78-92, doi:10.1016/j.fgb.2015.02.002 (2015).

10 Park, Y. J. *et al.* Whole genome and global gene expression analyses of the model mushroom Flammulina velutipes reveal a high capacity for lignocellulose degradation. *PloS one* **9**, e93560, doi:10.1371/journal.pone.0093560 (2014).

11 Liu, D., Hunt, M. & Tsai, I. J. Inferring synteny between genome assemblies: a systematic evaluation. *Bmc Bioinformatics* **19**, 26, doi:10.1186/s12859-018-2026-4 (2018).

12 Heinzelmann, R., Rigling, D., Sipos, G., Münsterkötter, M. & Croll, D. Chromosomal assembly and analyses of genome-wide recombination rates in the forest pathogenic fungus Armillaria ostoyae. *Heredity*, doi:10.1038/s41437-020-0306-z (2020).

13 Perez, G., Pangilinan, J., Pisabarro, A. G. & Ramirez, L. Telomere organization in the ligninolytic basidiomycete Pleurotus ostreatus. *Appl Environ Microbiol* **75**, 1427-1436, doi:10.1128/AEM.01889-08 (2009).

14 Holt, C. & Yandell, M. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC bioinformatics* **12**, 491, doi:10.1186/1471-2105-12-491 (2011).

15 Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210-3212, doi:10.1093/bioinformatics/btv351 (2015).

16 Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol* **20**, 238, doi:10.1186/s13059-019-1832-y (2019).

17 Emms, D. M. & Kelly, S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol* **16**, 157, doi:10.1186/s13059-015-0721-2 (2015).

18 De Bie, T., Cristianini, N., Demuth, J. P. & Hahn, M. W. CAFE: a computational tool for the study of gene family evolution. *Bioinformatics* **22**, 1269-1271, doi:10.1093/bioinformatics/btl097 (2006).

19 Stamatakis, A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics (Oxford, England)* **22**, 2688-2690, doi:10.1093/bioinformatics/btl446 (2006).

20 Mirarab, S. & Warnow, T. ASTRAL-II: coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. *Bioinformatics* **31**, i44-52, doi:10.1093/bioinformatics/btv234 (2015).

21 Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* **24**, 1586-1591, doi:10.1093/molbev/msm088 (2007).

22 Lupia, R., Lidgard, S. & Crane, P. Comparing palynological abundance and diversity: implications for biotic replacement during the Cretaceous angiosperm radiation. *Paleobiology* **25**, 305-340, doi:10.1017/S009483730002131X (1999).

23 Duplessis, S. *et al.* Obligate biotrophy features unraveled by the genomic analysis of rust fungi. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 9166-9171, doi:10.1073/pnas.1019315108 (2011).

24 Zheng, A. *et al.* The evolution and pathogenic mechanisms of the rice sheath blight pathogen. *Nature Communications* **4**, doi:10.1038/ncomms2427 (2013).

25 Stajich, J. E. *et al.* Insights into evolution of multicellular fungi from the assembled chromosomes of the mushroom Coprinopsis cinerea (Coprinus cinereus). *Proceedings of the National Academy of Sciences of the United States of America* **107**, 11889-11894, doi:10.1073/pnas.1003391107 (2010).

26 Chung, C. L. *et al.* Comparative and population genomic landscape of Phellinus noxius: A hypervariable fungus causing root rot in trees. *Mol Ecol*, doi:10.1111/mec.14359 (2017).

27 Zhang, X. *et al.* Genome-wide high-resolution mapping and functional analysis of DNA methylation in arabidopsis. *Cell* **126**, 1189-1201, doi:10.1016/j.cell.2006.08.003 (2006).

28 Wang, Y. P., Wang, X. Y., Lee, T. H., Mansoor, S. & Paterson, A. H. Gene body methylation shows distinct patterns associated with different gene origins and duplication modes and has a heterogeneous relationship with gene expression in Oryza sativa (rice). *New Phytologist* **198**, 274-283, doi:10.1111/nph.12137 (2013).

29 Montanini, B. *et al.* Non-exhaustive DNA methylation-mediated transposon silencing in the black truffle genome, a complex fungal genome with massive repeat element content. *Genome Biol* **15**, 411, doi:10.1186/s13059-014-0411-5 (2014).

30 Jones, P. A. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat Rev Genet* **13**, 484-492, doi:10.1038/nrg3230 (2012).

31 Elango, N., Hunt, B. G., Goodisman, M. A. D. & Yi, S. V. DNA methylation is widespread and associated with differential gene expression in castes of the honeybee, Apis mellifera. *P Natl Acad Sci USA* **106**, 11206-11211, doi:10.1073/pnas.0900301106 (2009).

32 Bewick, A. J. *et al.* Diversity of cytosine methylation across the fungal tree of life. *Nat Ecol Evol* **3**, 479-490, doi:10.1038/s41559-019-0810-9 (2019).

33 Willing, E. M. *et al.* Genome expansion of Arabis alpina linked with retrotransposition and reduced symmetric DNA methylation. *Nat Plants* **1**, 14023, doi:10.1038/nplants.2014.23 (2015).

34 Wang, H. *et al.* CG gene body DNA methylation changes and evolution of duplicated genes in cassava. *Proc Natl Acad Sci U S A* **112**, 13729-13734, doi:10.1073/pnas.1519067112 (2015).

35 Ma, L. J. *et al.* Comparative genomics reveals mobile pathogenicity chromosomes in Fusarium. *Nature* **464**, 367-373, doi:10.1038/nature08850 (2010).

36 Deshpande, A. R., Pochapsky, T. C. & Ringe, D. The Metal Drives the Chemistry: Dual Functions of Acireductone Dioxygenase. *Chem Rev* **117**, 10474-10501, doi:10.1021/acs.chemrev.7b00117 (2017).

37 Dong, S., Raffaele, S. & Kamoun, S. The two-speed genomes of filamentous pathogens: waltz with plants. *Curr Opin Genet Dev* **35**, 57-65, doi:10.1016/j.gde.2015.09.001 (2015).

38 Pedersen, B. S. & Quinlan, A. R. Mosdepth: quick coverage calculation for genomes and exomes. *Bioinformatics* **34**, 867-868, doi:10.1093/bioinformatics/btx699 (2018).

39 Rokas, A., Wisecaver, J. H. & Lind, A. L. The birth, evolution and death of metabolic gene clusters in fungi. *Nat Rev Microbiol*, doi:10.1038/s41579-018-0075-3 (2018).

40 Cortes-Perez, A. *et al.* New species and records of bioluminescent Mycena from Mexico. *Mycologia* **111**, 319-338, doi:10.1080/00275514.2018.1554172 (2019).

41 Mihail, J. D. Bioluminescence patterns among North American Armillaria species. *Fungal Biol* **119**, 528-537, doi:10.1016/j.funbio.2015.02.004 (2015).

42 Oliveira, A. G. *et al.* Circadian control sheds light on fungal bioluminescence.

Curr Biol **25**, 964-968, doi:10.1016/j.cub.2015.02.021 (2015).

43 Sharpe, M. L., Dearden, P. K., Gimenez, G. & Krause, K. L. Comparative RNA seq analysis of the New Zealand glowworm Arachnocampa luminosa reveals bioluminescence-related genes. *BMC Genomics* **16**, 825, doi:10.1186/s12864-015-2006-2 (2015).

44 Fallon, T. R. *et al.* Firefly genomes illuminate parallel origins of bioluminescence in beetles. *Elife* **7**, doi:10.7554/eLife.36495 (2018).

45 Desjardin, D. E., Oliveira, A. G. & Stevani, C. V. Fungi bioluminescence revisited. *Photochem Photobiol Sci* **7**, 170-182, doi:10.1039/b713328f (2008).

46 Bermudes, D., Petersen, R. H. & Nealson, K. H. Low-Level Bioluminescence Detected in Mycena-Haematopus Basidiocarps. *Mycologia* **84**, 799-802, doi:Doi 10.2307/3760392 (1992).

47 Zhang, B. & Horvath, S. A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol* **4**, Article17, doi:10.2202/1544-6115.1128 (2005).

48 Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *Bmc Bioinformatics* **9**, 559, doi:10.1186/1471-2105-9-559 (2008).

49 Krizsan, K. *et al.* Transcriptomic atlas of mushroom development reveals conserved genes behind complex multicellularity in fungi. *Proc Natl Acad Sci U S A* **116**, 7409-7418, doi:10.1073/pnas.1817822116 (2019).

50 Harvey, E. N. *Bioluminescence*. (Academic Press, 1952).

51 Wassink, E. C. in *Bioluminescence in action* (ed P.J. Herring) 171–197 (Academic Press, 1978).

52 Mondego, J. M. *et al.* A genome survey of Moniliophthora perniciosa gives new insights into Witches' Broom Disease of cacao. *BMC Genomics* **9**, 548, doi:10.1186/1471-2164-9-548 (2008).

53 Worrall, J. J., Anagnost, S. E. & Zabel, R. A. Comparison of wood decay among diverse lignicolous fungi. *Mycologia* **89**, 199-219 (1996).

54 Floudas, D. *et al.* The Paleozoic origin of enzymatic lignin decomposition reconstructed from 31 fungal genomes. *Science* **336**, 1715-1719, doi:10.1126/science.1221748 (2012).

55 Riley, R. *et al.* Extensive sampling of basidiomycete genomes demonstrates inadequacy of the white-rot/brown-rot paradigm for wood decay fungi. *Proceedings of the National Academy of Sciences*, doi:10.1073/pnas.1400592111 (2014).

56 Miyauchi, S. *et al.* Integrative visual omics of the white-rot fungus Polyporus brumalis exposes the biotechnological potential of its oxidative enzymes for

delignifying raw plant biomass. *Biotechnol Biofuels* **11**, 201, doi:10.1186/s13068-018-1198-5 (2018).

57 Marcet-Houben, M. & Gabaldon, T. Evolutionary and functional patterns of shared gene neighbourhood in fungi. *Nat Microbiol* **4**, 2383-2392, doi:10.1038/s41564-019-0552-0 (2019).

58 Rincones, J. *et al.* Genetic variability and chromosome-length polymorphisms of the witches' broom pathogen Crinipellis perniciosa from various plant hosts in South America. *Mycological Research* **110**, 821-832, doi:10.1016/j.mycres.2006.05.002 (2006).

59 Zolan, M. E., Heyler, N. K. & Stassen, N. Y. Inheritance of chromosome-length polymorphisms in Coprinus cinereus. *Genetics* **137**, 87-94 (1994).

60 Moller, M. & Stukenbrock, E. H. Evolution and genome architecture in fungal plant pathogens. *Nat Rev Microbiol*, doi:10.1038/nrmicro.2017.76 (2017).

61 Zamudio, N. & Bourc'his, D. Transposable elements in the mammalian germline: a comfortable niche or a deadly trap? *Heredity (Edinb)* **105**, 92-104, doi:10.1038/hdy.2010.53 (2010).

62 Zemach, A., McDaniel, I. E., Silva, P. & Zilberman, D. Genome-Wide Evolutionary Analysis of Eukaryotic DNA Methylation. *Science* **328**, 916-919, doi:10.1126/science.1186366 (2010).

63 Weinstein, P., Delean, S., Wood, T. & Austin, A. D. Bioluminescence in the ghost fungus Omphalotus nidiformis does not attract potential spore dispersing insects. *IMA Fungus* **7**, 229-234, doi:10.5598/imafungus.2016.07.02.01 (2016).

64 Purtov, K. V., Petushkov, V. N., Rodionova, N. S. & Gitelson, J. I. Why does the bioluminescent fungus Armillaria mellea have luminous mycelium but nonluminous fruiting body? *Dokl Biochem Biophys* **474**, 217-219, doi:10.1134/S1607672917030176 (2017).

65 Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution* **30**, 772-780, doi:10.1093/molbev/mst010 (2013).

66 Capella-Gutierrez, S., Silla-Martinez, J. M. & Gabaldon, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972-1973, doi:10.1093/bioinformatics/btp348 (2009).

67 Nguyen, L. T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular biology and evolution* **32**, 268-274, doi:10.1093/molbev/msu300 (2015).

68 Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q. & Vinh, L. S. UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Molecular*

biology and evolution **35**, 518-522, doi:10.1093/molbev/msx281 (2018).

69 Mori, K., Kojima, S., Maki, S., Hirano, T. & Niwa, H. Bioluminescence characteristics of the fruiting body of Mycena chlorophos. *Luminescence* **26**, 604-610, doi:10.1002/bio.1280 (2011).

70 Shih, Y. S., Chen, C. Y., Lin, W. W. & Kao, H. W. Mycena kentingensis, a new species of luminous mushroom in Taiwan, with reference to its culture method. *Mycol Prog* **13**, 429-435, doi:DOI 10.1007/s11557-013-0939-x (2014).

71 Koren, S. *et al.* Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res* **27**, 722-736, doi:10.1101/gr.215087.116 (2017).

72 Vaser, R., Sović, I., Nagarajan, N. & Šikić, M. Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res* **27**, 737-746, doi:10.1101/gr.214270.116 (2017).

73 Huang, S., Kang, M. & Xu, A. HaploMerger2: rebuilding both haploid sub-assemblies from high-heterozygosity diploid genome assembly. *Bioinformatics*, doi:10.1093/bioinformatics/btx220 (2017).

74 Walker, B. J. *et al.* Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *Plos One* **9**, e112963, doi:10.1371/journal.pone.0112963 (2014).

75 Benson, G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res* **27**, 573-580 (1999).

76 Grigoriev, I. V. *et al.* MycoCosm portal: gearing up for 1000 fungal genomes. *Nucleic Acids Res* **42**, D699-704, doi:10.1093/nar/gkt1183 (2014).

77 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21, doi:10.1093/bioinformatics/bts635 (2013).

78 Dobin, A. & Gingeras, T. R. Mapping RNA-seq Reads with STAR. *Curr Protoc Bioinformatics* **51**, 11 14 11-11 14 19, doi:10.1002/0471250953.bi1114s51 (2015).

79 Haas, B. J. *et al. De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc* **8**, 1494-1512, doi:10.1038/nprot.2013.084 (2013).

80 Pertea, M. *et al.* StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol* **33**, 290-295, doi:10.1038/nbt.3122 (2015).

81 Song, L., Sabunciyan, S. & Florea, L. CLASS2: accurate and efficient splice variant annotation from RNA-seq reads. *Nucleic Acids Res* **44**, e98, doi:10.1093/nar/gkw158 (2016).

82 Trapnell, C. *et al.* Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature protocols* **7**, 562-578,

doi:10.1038/nprot.2012.016 (2012).

83 Wu, T. D. & Watanabe, C. K. GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics* **21**, 1859-1875, doi:10.1093/bioinformatics/bti310 (2005).

84 Venturini, L., Caim, S., Kaithakottil, G. G., Mapleson, D. L. & Swarbreck, D. Leveraging multiple transcriptome assembly methods for improved gene structure annotation. *GigaScience* **7**, giy093, doi:10.1093/gigascience/giy093 (2018).

85 Stanke, M., Tzvetkova, A. & Morgenstern, B. AUGUSTUS at EGASP: using EST, protein and genomic alignments for improved gene prediction in the human genome. *Genome Biol* **7**, S11, doi:10.1186/gb-2006-7-s1-s11 (2006).

86 Ter-Hovhannisyan, V., Lomsadze, A., Chernoff, Y. O. & Borodovsky, M. Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training. *Genome Research* **18**, 1979-1990, doi:10.1101/gr.081612.108 (2008).

87 Hoff, K. J., Lange, S., Lomsadze, A., Borodovsky, M. & Stanke, M. BRAKER1: unsupervised RNA-Seq-Based genome annotation with GeneMark-ET and AUGUSTUS. *Bioinformatics* **32**, 767-769, doi:10.1093/bioinformatics/btv661 (2016).

88 Korf, I. Gene finding in novel genomes. *BMC Bioinformatics* **5**, 59, doi:10.1186/1471-2105-5-59 (2004).

89 Conesa, A. *et al.* Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**, 3674-3676, doi:10.1093/bioinformatics/bti610 (2005).

90 Huerta-Cepas, J. *et al.* Fast genome-wide functional annotation through orthology assignment by eggNOG-Mapper. *Mol Biol Evol* **34**, 2115-2122, doi:10.1093/molbev/msx148 (2017).

91 Berriman, M., Coghlan, A. & Jason Tsai, I. J. Creation of a comprehensive repeat library for a newly sequenced parasitic worm genome. *Protocol Exchange*, doi:10.1038/protex.2018.054 (2018).

92 Ellinghaus, D., Kurtz, S. & Willhoeft, U. LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. *BMC bioinformatics* **9**, 18, doi:10.1186/1471-2105-9-18 (2008).

93 El-Gebali, S. *et al.* The Pfam protein families database in 2019. *Nucleic Acids Res* **47**, D427-D432, doi:10.1093/nar/gky995 (2019).

94 Smit, A., Hubley, R. & Green, P. *RepeatMasker Open-4.0*. (2013-2015).

95 Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842,

doi:10.1093/bioinformatics/btq033 (2010).

96 Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2-approximately maximum-likelihood trees for large alignments. *Plos One* **5**, e9490, doi:ARTN e9490 10.1371/journal.pone.0009490 (2010).

97 Yu, G., Lam, T., Zhu, H. & Guan, Y. Two Methods for Mapping and Visualizing Associated Data on Phylogeny Using Ggtree. *Molecular Biology and Evolution* **35**, 3041-3043, doi:10.1093/molbev/msy194 (2018).

98 Guo, W. *et al.* BS-Seeker2: a versatile aligning pipeline for bisulfite sequencing data. *BMC Genomics* **14**, 774, doi:10.1186/1471-2164-14-774 (2013).

99 Ni, P. *et al.* DeepSignal: detecting DNA methylation state from Nanopore sequencing reads using deep-learning. *Bioinformatics* **35**, 4586-4595, doi:10.1093/bioinformatics/btz276 (2019).

100 Conway, J. R., Lex, A. & Gehlenborg, N. UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics* **33**, 2938-2940, doi:10.1093/bioinformatics/btx364 (2017).

101 Yu, G., Smith, D., Zhu, H., Guan, Y. & Lam, T. T. ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol Evol* **8**, doi:10.1111/2041-210X.12628 (2017).

102 Alexa, A., Rahnenfuhrer, J. & Lengauer, T. Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics* **22**, 1600-1607, doi:10.1093/bioinformatics/btl140 (2006).

103 Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M. & Henrissat, B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res* **42**, D490-495, doi:10.1093/nar/gkt1178 (2014).

104 Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using DIAMOND. *Nat Methods* **12**, 59-60, doi:10.1038/nmeth.3176 (2015).

105 Finn, R. D., Clements, J. & Eddy, S. R. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res* **39**, W29-37, doi:10.1093/nar/gkr367 (2011).

106 Kozlov, A., Darriba, D., Flouri, T., Morel, B. & Stamatakis, A. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* **35**, 4453-4455, doi:10.1093/bioinformatics/btz305 (2019).

107 dos Reis, M. & Yang, Z. Approximate Likelihood Calculation on a Phylogeny for Bayesian Estimation of Divergence Times. *Molecular Biology and Evolution* **28**, 2161-2172, doi:10.1093/molbev/msr045 (2011).

108 Kumar, S., Stecher, G., Suleski, M. & Hedges, S. B. TimeTree: a resource for

timelines, timetrees, and divergence times. *Mol Biol Evol* **34**, 1812-1819, doi:10.1093/molbev/msx116 (2017).

109 Poinar, G. & Buckley, R. Evidence of mycoparasitism and hypermycoparasitism in Early Cretaceous amber. *Mycological Research* **111**, 503-506, doi:10.1016/j.mycres.2007.02.004 (2007).

110 Cai, C., Leschen, R., Hibbett, D., Xia, F. & Huang, D. Mycophagous rove beetles highlight diverse mushrooms in the Cretaceous. *Nature Communications* **8**, doi:10.1038/ncomms14894 (2017).

111 Guy, L., Kultima, J. R. & Andersson, S. G. genoPlotR: comparative gene and genome visualization in R. *Bioinformatics* **26**, 2334-2335, doi:10.1093/bioinformatics/btq413 (2010).

112 Kosakovsky Pond, S. L. *et al.* HyPhy 2.5-A Customizable Platform for Evolutionary Hypothesis Testing Using Phylogenies. *Molecular biology and evolution* **37**, 295-299, doi:10.1093/molbev/msz197 (2020).

113 Pond, S. L., Frost, S. D. & Muse, S. V. HyPhy: hypothesis testing using phylogenies. *Bioinformatics* **21**, 676-679, doi:10.1093/bioinformatics/bti079 (2005).

114 Weaver, S. *et al.* Datamonkey 2.0: A Modern Web Application for Characterizing Selective and Other Evolutionary Processes. *Molecular biology and evolution* **35**, 773-777, doi:10.1093/molbev/msx335 (2018).

115 Delport, W., Poon, A. F., Frost, S. D. & Kosakovsky Pond, S. L. Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. *Bioinformatics* **26**, 2455-2457, doi:10.1093/bioinformatics/btq429 (2010).

116 Pond, S. L. & Frost, S. D. Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics* **21**, 2531-2533, doi:10.1093/bioinformatics/bti320 (2005).

117 Kosakovsky Pond, S. L., Posada, D., Gravenor, M. B., Woelk, C. H. & Frost, S. D. GARD: a genetic algorithm for recombination detection. *Bioinformatics* **22**, 3096-3098, doi:10.1093/bioinformatics/btl474 (2006).

118 Kosakovsky Pond, S. L. & Frost, S. D. Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Molecular biology and evolution* **22**, 1208-1222, doi:10.1093/molbev/msi105 (2005).

119 Murrell, B. *et al.* Detecting individual sites subject to episodic diversifying selection. *Plos Genet* **8**, e1002764, doi:10.1371/journal.pgen.1002764 (2012).

120 Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-2120, doi:10.1093/bioinformatics/btu170 (2014).

121 Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose

program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923-930, doi:10.1093/bioinformatics/btt656 (2014).

122 Meinhardt, L. W. *et al.* Genome and secretome analysis of the hemibiotrophic fungal pathogen, Moniliophthora roreri, which causes frosty pod rot disease of cacao: mechanisms of the biotrophic and necrotrophic phases. *BMC Genomics* **15**, 164, doi:10.1186/1471-2164-15-164 (2014).