

Ancient sex compatibility genes underlie tissue differentiation in mushrooms

László Nagy

lnagy@funkenomelab.com

Synthetic and Systems Biology Unit, Biological Research Center <https://orcid.org/0000-0002-4102-8566>

Torda Varga

Hungarian Academy of Sciences <https://orcid.org/0000-0002-2597-9126>

Hongli Wu

HUN-REN Biological Research Centre

Anna Szekeres

HUN-REN Biological Research Centre

Árpád Csernetics

HUN-REN Biological Research Centre

Csenge Földi

Synthetic and Systems Biology Unit, Biological Research Center <https://orcid.org/0000-0002-9982-0679>

Máté Virágh

HUN-REN Biological Research Centre

Nikolett Miklovics

HUN-REN Biological Research Centre

Xiao-Bin Liu

HUN-REN Biological Research Centre

Zhihao Hou

State Key Laboratory of Efficient Utilization of Arable Land in China, Institute of Agricultural Resources and Regional Planning, Chinese Academy of Agricultural Sciences, Beijing 100081, China

Zsolt Merényi

Synthetic and Systems Biology Unit, Biological Research Center <https://orcid.org/0000-0003-1114-3739>

Gábor Berend

Institute of Informatics, University of Szeged, Szeged 6726, Hungary

Blythe Angus

Digital Biology Group, Faculty of Life Sciences, University of Dundee, Dundee, United Kingdom

Stefan Manolache

University of Dundee

Victor Rocés

Digital Biology Group, Faculty of Life Sciences, University of Dundee, Dundee, United Kingdom

András Vágner

HUN-REN Biological Research Centre

Ji-Peng Li

HUN-REN Biological Research Centre <https://orcid.org/0000-0002-3250-7346>

Zsuzsanna Darula

HUN-REN Biological Research Centre

Dorottya Szafián

HUN-REN Biological Research Centre

Edit Ábrahám

HUN-REN Biological Research Centre <https://orcid.org/0000-0002-1268-760X>

Péter Horváth

Synthetic and Systems Biology Unit, Institute of Biochemistry Biological Research Centre, 6726 Szeged

<https://orcid.org/0000-0002-4492-1798>

Zsolt Kristóffy

HUN-REN Biological Research Centre

Jinwei Zhang

HUN-REN Biological Research Centre

Balázs Bálint

MTA Biological Research Centre <https://orcid.org/0000-0002-5102-0787>

Botond Hegedüs

Synthetic and Systems Biology Unit, Biological Research Center <https://orcid.org/0000-0002-8553-8234>

Nikolett Zsibrita

HUN-REN Biological Research Centre

Hajk-Georg Drost

Division of Computational Biology, School of Life Sciences, University of Dundee

<https://orcid.org/0000-0002-1567-306X>

Zoltan Lipinszki

HUN-REN Biological Research Centre <https://orcid.org/0000-0002-2067-0832>

Biological Sciences - Article

Keywords:

Posted Date: September 9th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10877730/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: There is **NO** Competing Interest.

Ancient sex compatibility genes underlie tissue differentiation in mushrooms

Torda Varga¹, Hongli Wu^{2*}, Anna Szekeres^{2*}, Árpád Csernetics^{2*}, Csenge Földi^{2*}, Máté Virágh^{2*}, Nikolett Miklovics^{2*}, Xiao-Bin Liu^{2*}, Zhihao Hou^{2*\$}, Zsolt Merényi², Gábor Berend³, Blythe Angus⁴, Stefan Manolache⁴, Victor Roces⁴, András Vágner², Ji-Peng Li², Zsuzsanna Darula^{2,5}, Dorottya Szafián², Edit Ábrahám², Péter Horváth², Zsolt Kristóffy², Jinwei Zhang², Balázs Bálint², Botond Hegedüs², Nikolett Zsibrita², Hajk Drost^{4,6}, Zoltán Lipinszki², László G. Nagy^{2,7,8&}

¹ Royal Botanic Gardens Kew, Richmond, TW9 3DS UK

² HUN-REN Biological Research Centre, Szeged 6726, Hungary

³ Institute of Informatics, University of Szeged, Szeged 6726, Hungary

⁴ Digital Biology Group, Faculty of Life Sciences, University of Dundee, Dundee, United Kingdom

⁵ Single Cell Omics Advanced Core Facility, Hungarian Centre of Excellence for Molecular Medicine, Szeged 6728, Hungary

⁶ Computational Biology Group, Max Planck Institute for Biology Tübingen, Tübingen, Germany

⁷ Korea University, Seoul 02841, Republic of Korea

⁸ Shandong Agricultural University, Tai'an, China

* These authors contributed equally

& Author for correspondence: lnagy@funkenomelab.com

\$ Current address: State Key Laboratory of Efficient Utilization of Arable Land in China, Institute of Agricultural Resources and Regional Planning, Chinese Academy of Agricultural Sciences, Beijing 100081, China

Summary – Complex multicellular development through tissue differentiation and organogenesis is among the most complex, but least understood genetic mechanisms of living systems. Development across different organisms converges on a miniature blueprint of the mature organism, which, in animals, plants and brown algae, is mostly achieved via embryogenesis, resulting in distinct tissues and organs^{1,2}. Fungal fruiting bodies are complex multicellular structures, but in contrast to other organisms, they are assembled via apical growth of tubular hyphae³, a unique mechanism for the fungal kingdom. How hyphal growth was adapted to enable complex development, whether regions of the fruiting body function as real tissues and what are the key regulators of differentiation - equivalents of Hox genes of animals or ABC genes of plants - have remained unknown. Here we show that fruiting bodies of mushroom-forming fungi evolved hallmarks of *bona fide* tissues, such as distinct transcriptomic footprints, functions, and specialization and we uncover tissue-specific transcriptome patterns and genetic networks underlying their development. We discover that mating pheromones and their receptors, as well as a gene family required for pheromone maturation, are required for tissue differentiation and form distinct, tissue-specific pathways driving differentiation. Ancestrally, these genes regulate mating behaviour, indicating co-option during evolution and that Agaricomycetes use the same genetic system to control sexual compatibility and tissue development. This study uncovers how evolution produced complex tissues from apically growing hyphae, and illustrates that selection can drive the convergent evolution of tissues across organisms with fundamentally different cellular architectures.

Introduction

The ability to grow and develop complex multicellular structures is found across diverse clades (animals, plants, fungi, brown and red algae) and in distinct biological contexts, such as embryogenesis, flower development and the production of fungal fruiting bodies^{2,4,5}. Notably, certain features of complex multicellular development, such as symmetry breaking, pattern formation, and tissue differentiation, occur in all complex multicellular clades^{1,6}. Tissue differentiation is arguably one of the most intricate of these processes, and its independent evolution in unrelated clades points to general principles of the evolution of complex development¹. The general principles of development, such as the necessity of division of labor, that of functional specialization, the co-option of ancient genes or the impact of regulatory mutations, remain strongly tethered to animal and plant models. How can we define general principles without considering other organismal groups that produce complex structures?

Across the tree of life, the molecular mechanisms driving tissue development are the least known among fungi, in which sexual fruiting bodies – technically sporocarps or asco-/basidiomes – reach the highest morphological complexity^{7,8}, with at least 30 morphologically distinguishable cell types⁹. Fruiting bodies develop from apically growing hyphae, a unique solution for developing complex structures. However, because growth in fruiting bodies remains hyphal, mushrooms are traditionally considered not to have *bona fide* tissues, only pseudotissues or plectenchyma^{9–12}. Yet hyphae form morphologically and functionally distinct regions in fruiting bodies, which raises the questions: can these be considered *bona fide* tissues and how fungi evolved to coordinate the growth of thousands of apically growing hyphae into highly differentiated fruiting bodies?

To address this question, we study Agaricomycetes (mushroom-forming fungi), a class of ca. 40,000 species of mushrooms toadstools, puffballs, and most cultivated crop mushrooms³. Agaricomycetes produce the morphologically most complex fruiting bodies and have over 350 million years of evolutionary history¹³, but despite their diversity and importance, the fundamental principles underlying fruiting body differentiation remain unknown. Fruiting bodies of mushroom-forming fungi are sexual reproductive structures that facilitate spore dispersal, e.g. by lifting spore-producing cells above ground. They develop in response to internal and external factors and start with the aggregation and hyperbranching of competent hyphal tips^{14,15}. Branching generates a loose hyphal aggregate, which then differentiates into a compact inner and a loose outer layer. In simple terms, the

inner region (nodulus) gives rise to the cap, stipe, gills and other structures, whereas the outer region participates in forming layers of protective sheath, the universal veil (UV).

In *Coprinopsis cinerea*, an emerging agaricomycete model (Fig 1/a), a one mm tall so called 'primordium' already contains all differentiated regions (for simplicity, we will refer to these as tissues) that are organised in a manner that resembles a mature mushroom, even though it is only 1/100th of the size. This early establishment of a blueprint of the mature fruiting body was a key evolutionary innovation that spurred lineage diversification in some clades of Agaricomycetes¹⁶. Additionally, the occurrence of key differentiation events within a protected environment resembles animal embryogenesis and flower morphogenesis. However, master regulators of development and the mechanisms controlling tissue differentiation, such as the analogues of Hox genes in animals or ABC genes in flowers, have remained unknown.

Fruiting body development is preceded by mating of two compatible hyphae¹⁷. Mating compatibility in Agaricomycetes is controlled by two unlinked loci, the HD (homeodomain) and P/R (Pheromone/Receptor) loci. First, hyphal fusion between individuals can occur if the P/R locus carries different alleles of pheromone and pheromone receptor genes. After fusion, heterodimerization of homeodomain proteins (HD1 and HD2) encoded by compatible alleles establishes the transcriptional program required for maintaining a sexually competent dikaryotic strain^{18,19}. Genes in the P/R locus are highly paralogous with each paralog displaying multiallelism in the population. Both multiallelism and paralogy have been proposed to promote outbreeding, even though multiallelism alone would be sufficient for that²⁰, leaving the role of paralogy incompletely explained.

Here, we employ ultralow-input laser-capture microdissection (LCM) coupled with RNA-seq in *C. cinerea* to chart gene expression at high resolution across development, from a ~100µm wide, internally homogeneous hyphal knot to a ~2mm tall primordium that contains all major tissues (Fig. 1/a, Supplementary Fig. 1). We detect widespread tissue-enhanced gene expression with late conservation of development and provide evidence for division of labour and specialisation during development. We demonstrate that an ancient pheromone-receptor system, which regulates mate recognition, has acquired a key role in modulating tissue development. Therefore, this study illuminates the first tissue-differentiation pathways in mushroom-forming fungi, uncovering how fungi solved the challenge of complex development via hyphal growth.

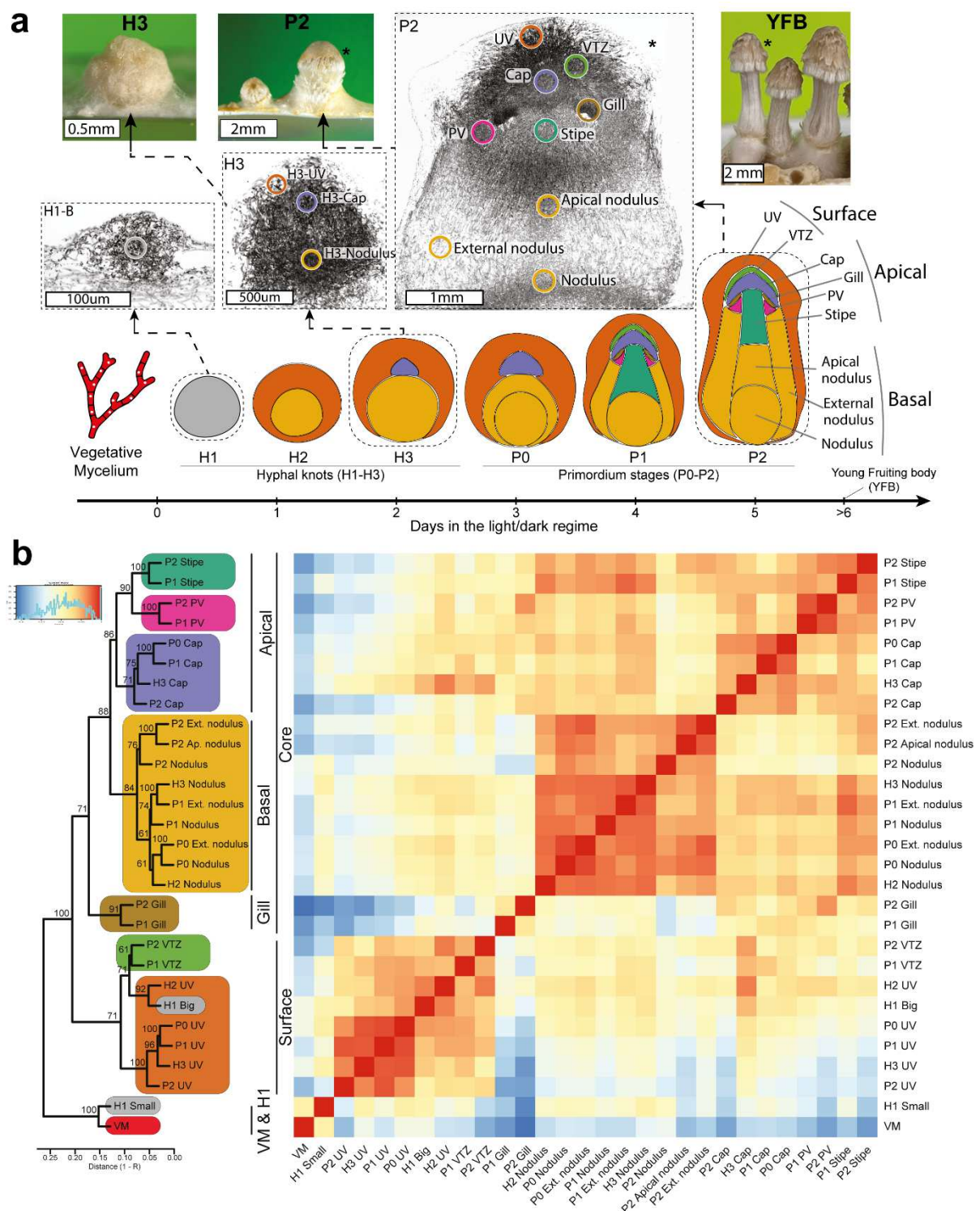


Fig. 1. Tissue-restricted gene expression patterns in a model mushroom. **a.** Nine tissues of seven developmental stages of *Coprinopsis cinerea*, were dissected from ca. 10 μ m thick sections using laser-capture microdissection. The Top row shows the stereomicroscope photographs (green background) of major developmental stages. H3 (tertiary hyphal knot), where the first histological sign of a compact tissue is designated as a cap. Primordium stage 2 (P2), where all tissue types are present and develop into the young fruiting body (YFB). The asterisk denotes the mushroom shape, i.e., cap and stipe. The middle row depicts light-microscopic images of the larger primary hyphal knot (H1-B) and those of the H3 and P2 stages. Coloured circles highlight tissues sampled for LCM-RNA-seq. The bottom row outlines spatio-temporal patterns of the seven developmental stages and nine tissue types included in this study. VM: vegetative mycelium; H1: primary hyphal knot (note there are two distinct H1 populations); H2: secondary hyphal knot; H3: tertiary hyphal knot; P0, P1 and P2: primordium stages 0, 1 and 2; UV: universal veil; VTZ: veil transition zone; Ext. nodule: external nodule;

PV: partial veil. See Supplementary Fig. 1 for all developmental stages and tissue samples. **b.** The hierarchical clustering of 29 sample types (each with 3-5 biological replicates averaged) into nine main tissue types based on the ultra-low input RNA-seq. The heatmap depicts similarity based on Pearson correlations between mean gene-expression profiles. The dendrogram is coloured to match the tissue types shown in panel a.

Results

Transcriptionally distinct tissues emerge during the development of *C. cinerea* fruiting bodies

In mushrooms, including *C. cinerea*, no prior information on tissue-marker genes is available, thus we made LCM-RNA-seq our method of choice for obtaining topologically labeled samples, in contrast to marker gene dependent droplet-based scRNA-seq approaches. We sampled the early development of *C. cinerea* at seven developmental stages and up to nine tissue types (Fig. 1/a, Supplementary Fig. 1). Our sampling includes the most important developmental transitions: differentiation of a homogeneous mass into two cell layers, the emergence of key tissues and a blueprint for the mature fruiting body. Cell differentiation establishes all major tissue types by the primordium 1 and 2 stages (P1 and P2), consequently our sampling comprises the most tissue types at these time points. After replicate homogeneity assessments via spike-in RNAs and principal component analyses (PCA) (Supplementary Fig. 2, Supplementary Table 1, Supplementary Table 2) we retained 85 high-quality samples, for which we calculated gene expression levels using DESeq2 rlog transformation and qsmooth normalisation (Supplementary Note 1). A generalized linear model (GLM) revealed that tissue identity explains transcriptome patterns better than developmental stage (Mantel $r = 0.038$, $P = 0.247$, mean Silhouette value = -0.211 , Supplementary Fig. 3).

Tissue similarities are expected to reflect shared developmental history²¹, which is currently unknown in *C. cinerea*. Hierarchical clustering (Approximately Unbiased (AU) $P \geq 0.95$) and testing alternative groupings (Mantel $r = 0.574$, $P = 0.001$; mean Silhouette = 0.266) of samples outlined nine major clusters (Fig. 1/b), indicating that specific gene expression programs are associated with tissue identities. This clustering reveals a differentiation trajectory of tissues in the fruiting body (Fig. 1/b). Vegetative mycelium differs most from all fruiting body tissues (Fig. 1/b), consistent with this not being complex multicellular²². Within the fruiting body, the most divergent cluster includes surface tissues (Fig. 1/a, Supplementary Fig. 1), which provide physical and chemical protection to the developing inner cell types¹⁴, which we refer to as core tissues. The core tissues make up the stereotypical 'mushroom' morphology, whereas surface tissues are only found in certain taxa, including *C. cinerea* (Fig 1/a). Core tissues comprise basal (the nodulus and adjacent regions) and apical parts (cap, gill, stipe, partial veil). The clustering of core tissues indicates that there may be a gene expression program that characterizes this conserved developmental unit.

Overall, these data indicate that tissue development in fruiting bodies follows a clear trajectory and leads to transcriptionally distinguishable regions.

Fruiting body development establishes functionally distinct tissues

To understand what drives the identity of the tissues in *C. cinerea* fruiting bodies, we next focused on dissecting gene expression in different tissues. We identified genes with significantly different expression ($P < 0.01$) relative to the grand mean across all samples, while accounting for developmental stage effects (Fig. 2/a). We refer to these as tissue-enhanced and -repressed genes, similar to existing approaches^{23–25}, but different from the notion of tissue-specificity, which has been criticized²⁶. We chose a fold-change cutoff of 2.5 based on fold-change / Shannon entropy relationships (Supplementary Fig. 4).

We identified a total of 3,726 tissue-enhanced and 3,667 tissue-repressed genes across the nine main tissue types (Fig. 2/b, Supplementary Table 3). Comparison of tissue-enhanced genes to tissue-specific genes (measured by the τ metric) indicated significant overlap between the two

approaches (Hypergeometric test, $P < 10^{-16}$) (Fig. 2/c). Proteomic analysis of separate apical/basal samples corroborates these tissue differences and a significant overlap with tissue-enhanced gene sets (Supplementary Fig. 5, Supplementary Table 4). Collectively, these findings demonstrate that tissue-restricted gene expression is prevalent in *C. cinerea* fruiting bodies.

To elucidate the specific functions of each tissue, we integrated information from experimentally characterized *C. cinerea* genes and a total of 244 and 219 significant GO terms associated with tissue-enhanced and -repressed genes, respectively (Supplementary Table 5) (Fig. 2/b). The universal veil has one of the most unique functional footprints, reflecting low metabolic activity and a specialization in interacting with the environment. Several antifungal and insecticidal defence genes (e.g., *ccl1*, *ccl2*, *cgl1*, *aerolysins*, *pic1*, *ccp2*) were highly expressed (Supplementary Fig. 6), which is in line with the role of surface tissues in physical and chemical protection of core tissues^{16,27}. The universal veil also displayed the highest expression of hydrophobin genes, whose products confer hydrophobicity to the cell surface²⁸ (Supplementary Fig. 7). By contrast, core tissues highly express genes related to metabolism and cellular growth, cell division, ribosome assembly, mitochondria, and chromatin-remodeling (Supplementary Figs. 8-11, Supplementary Table 5). Based on GO, gills were most divergent from other tissues, which is largely driven by genes related to DNA replication and repair (e.g. *rad3*, *rec8*, *smc3*, *dmc1*). The nodulus had high expression of genes for glycogen biosynthesis, reflecting the accumulation of glycogen during early development¹⁴. Genes related to cell wall synthesis and modification showed high expression variation but no clear tendency across tissues (Supplementary Fig. 12, Supplementary Table 5), indicating that cell wall remodelling is active throughout the whole fruiting body and involves tissue-specific gene sets. For example, the chitinases, *ChiIII* and *ChiEN3* were enhanced in the cap, *ChiIII* and *ChiB1* in the universal veil, *ChiEN3* and *ChiB1* in gills, while *ChiEN2* was enhanced in the nodulus.

GO terms corresponding to transcription factors (TFs) and isoprenylcysteine carboxyl methyltransferases (ICMTs; detailed below) showed substantial variation among tissues (Fig. 2/b, Supplementary Table 5). The expression of 93 TF genes was tissue-enhanced or -repressed (Supplementary Fig. 13), and each tissue had a distinct set of tissue-enhanced TFs, which formed clusters corresponding to surface and core tissues (Supplementary Fig. 13), indicating that *C. cinerea* likely employs regulatory programs specific to each tissue. For example, *gcd1*, which regulates UV development²⁹, is both UV- and cap-enhanced. The Gti1/Pac2, C₂H₂ Zn-finger and binuclear (Zn(2)-C6) zinc-fingers TF families are significantly overrepresented (Fisher's exact test, $P < 0.005$, Fig. 2/f, Supplementary Table 6), indicating that these families may be particularly relevant for tissue development.

To understand how regions of the fruiting body compare to *bona fide* tissues, we quantified two hallmarks of the latter: division of labour and specialization. Distances based on gene expression indicate a strong separation of core and surface tissues (Fig. 2/d), whereas GO-based distances indicate strong functional divergence among all tissue pairs (Fig. 2/e). This pattern is robust across different GO categories (Supplementary Fig. 14) and overall indicates that fruiting body regions display functional divergence. Specialization of tissues is often accompanied by selective gene silencing (e.g. in metazoans), which is marked by an increasing number of silenced genes as development progresses. For three tissue types where this was testable, we found that the number of silenced genes (Counts per million (CPM) of reads < 10) increases monotonously (Fig. 2/g), indicating growing repression of genes through time.

Overall, our analyses provide the first high-resolution insights into tissue functions in a mushroom-forming fungus and reveal that fruiting bodies display tissue-specific expression profiles with distinct functional footprints. Expression and GO distances as well as repressed genes indicate that fruiting body regions display functional divergence and specialization through time, similarly to *bona fide* animal and plant tissues.

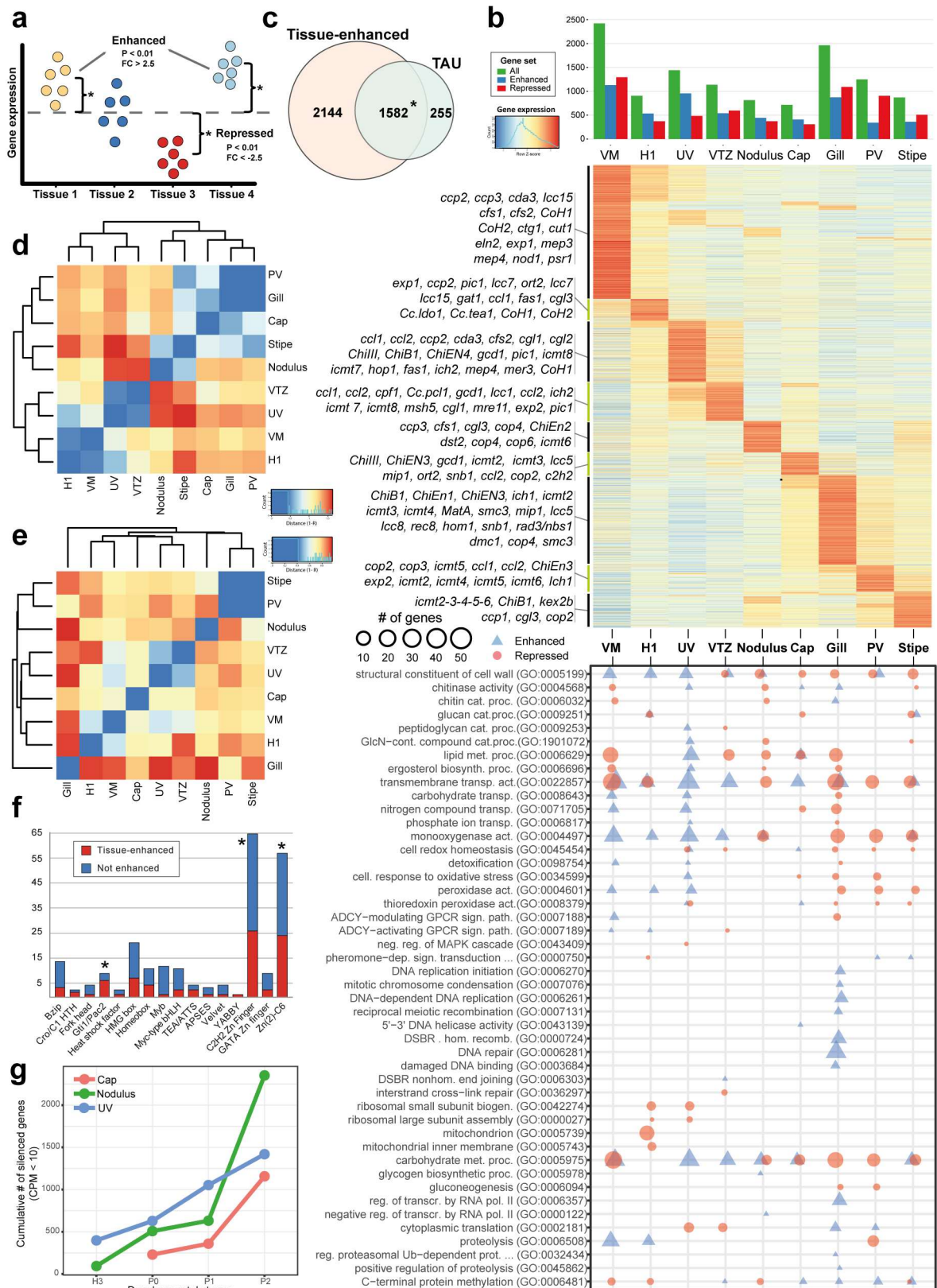


Fig. 2. Functional footprints and divergence between tissues of *C. cinerea*. **a.** The principle of identifying tissue-enhanced and -repressed genes from our RNA-seq data. The dashed line indicates the grand mean across all tissues. FC: Fold change. **b.** Global patterns of tissue-enhanced and -repressed genes. The bar chart depicts the number and distribution of tissue-enhanced and -repressed genes across nine major tissue types. The heatmap shows expression patterns of 3,727 tissue-enhanced genes, highlighting experimentally characterized *C. cinerea* genes on the left. The bottom plot shows a tabular representation of characteristic enriched Gene Ontology terms. See Supplementary Table 5. for the complete list. Tissue types are abbreviated

as follows: VM, vegetative mycelium; UV, universal veil; VTZ, veil transition zone; PV, partial veil. **c.** The overlap between tissue-enhanced genes and tissue-specific genes measured by Tau. * indicates a significant enrichment (Hypergeometric test, $P < 0.05$) of TAU-based tissue-specific genes among tissue-enhanced genes identified by generalised linear model. **d.** Divergence between nine major tissue types measured by the expression fold change of all genes relative to the grand mean. **e.** Divergence between nine major tissue types based on Biological Process GO term enrichment score dissimilarity. In d. and e. blue and red correspond to low and high distance, respectively. **f.** The number and enrichment of transcription factor families among tissue-enhanced genes. * indicates enrichment $P < 0.005$. **g.** The cumulative number of 'off' genes (CPM < 10) during the development of the cap, nodulus and UV.

Pheromones and their receptors are involved in tissue differentiation in mushrooms

ICMTs stand out as one of the strongest signals among tissue-enhanced genes and in GO analyses. ICMTs are part of the CAAX protein maturation pathway (Fig. 3/a), which post-translationally modifies CAAX-motif proteins, such as mating pheromones³⁰. In contrast to yeast and humans, which have a single ICMT gene, *C. cinerea* has eight, all of which exhibit tissue-enhanced or repressed expression in different tissues (Fig. 3/a, Supplementary Fig. 15, Supplementary Table 7).

We hypothesised that ICMT expression patterns might reflect the importance of certain CAAX proteins and their downstream pathways. Among 201 annotated CAAX proteins and 535 CAAX peptides identified *de novo* by 6-frame translation of the genomic sequence (Supplementary Table 8), pheromones had the strongest expression differences and GO signal in our RNA-seq data (Supplementary Tables 5,9). Indeed, most of the pheromone precursor genes show complex expression patterns that vary across tissues and developmental stages (Fig. 3/a, Supplementary Fig. 16, Supplementary Table 9). All seven pheromone-sensing G-protein coupled receptor (GPCR) genes also showed complex expression patterns (Supplementary Table 9). The tissue-enhanced expression of pheromones and GPCRs was unexpected, given that they are conventionally involved in mating, which precedes fruiting body development. We therefore hypothesised that the expression patterns of pheromones, GPCRs, and ICMTs reflect a regulatory role in tissue differentiation, rather than in mating.

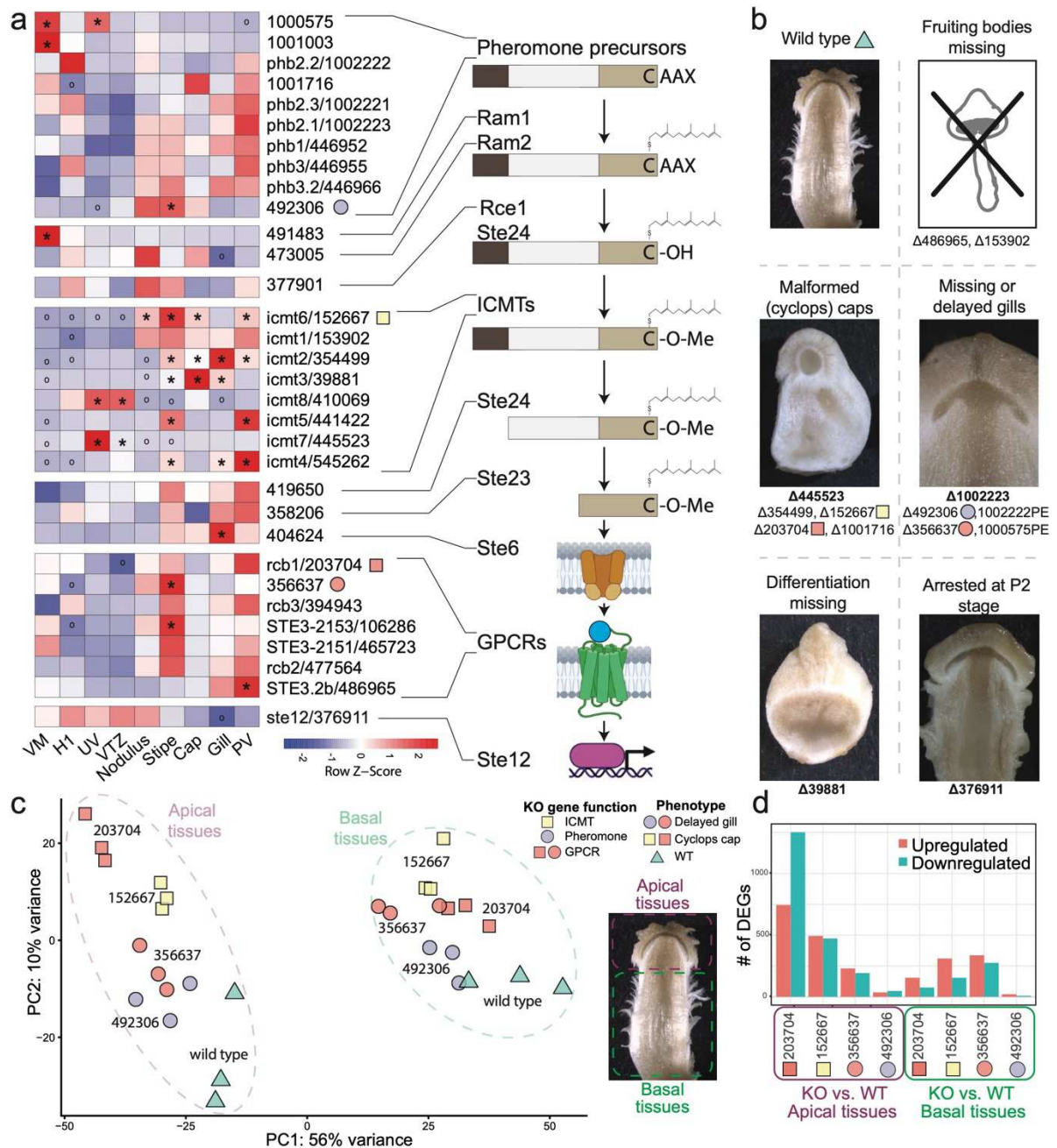


Fig. 3. Pheromones and their associated pathways are involved in tissue differentiation in *C. cinerea*. **a.** The CAAX protein maturation, pheromone precursors and downstream pheromone exporting and sensing systems in *C. cinerea* (right). Me - methyl group. Gene expression heatmap shows their expression across nine main tissue clusters. In the heatmap gene names are given before 'I' (if available), followed by protein IDs. * and ° indicate tissue-enhanced and -repressed genes, respectively. Colored rectangles and circles correspond to the labels on panel c. **b.** P2 stage (or equivalent) phenotypes associated with *icmt*, pheromone and GPCR genes, arranged into main mutant groups. 'Δ' - gene disruption; PE - promoter exchange. Bold indicates the actual mutant shown. Colored rectangles and circles correspond to the labels on panel c. See Supplementary Data 1 for further images of each mutant. **c.** Principal component analysis of wild type and mutant RNA-seq data for apical and basal parts of the fruiting bodies. **d.** The number of differentially expressed genes detected across pairwise comparisons of wild type and mutant RNA-seq data.

To test this hypothesis, we chose eight ICMTs, six pheromone precursors, five GPCRs and one TF gene based on expression patterns and generated 41 stable homozygous gene deletion mutants using CRISPR/Cas9, promoter replacement or overexpression for 20 genes (Supplementary Data 1). We confirmed the gene disruptions by PCR and the absence of transcription by RT-PCR. To

ensure that the observed phenotype is caused by the intended mutation we examined multiple mutants for several genes, as well as complementation mutants for randomly chosen genes. Most of the mutants displayed severe defects in tissue differentiation or the lack of certain tissues (Fig. 3/b), while vegetative growth rate was not affected (except in *Δicmt1*), indicating a specific role of these genes in fruiting bodies. The most severe phenotype was the complete lack of fruiting bodies observed for two genes (*Δicmt1*, *ΔSTE3.2b*), followed by the lack of any internal differentiation for one gene (*Δicmt3*). Malformed and arrested caps, resulting in cyclops-like fruiting bodies, were observed in mutants of four genes (*Δicmt7*, *Δicmt6*, *Δ1001716* and *Δrcb1*). Finally, mutants for two genes did not develop gills (*Δicmt2*, *Δicmt7*) and mutants of three genes (*Δ492306*, *Δphb2.1*, *Δ356637*) lacked gills in stage 2-3 primordia but later developed them. Some delayed-gill mutants could catch up with WT, whereas in others gills remained narrower. While *Δ376911* mutants (orthologous to the Ste12 transcription factor, the terminal effector in the pheromone-sensing MAP kinase pathway) displayed normal tissue development, their fruiting bodies arrested at the primordium 2 stage (Fig. 3/b, Supplementary Data 1). For the above experiments we used the *C. cinerea* Amut43Bmut43 strain, which has unique mating behaviour¹⁴, however, our findings are unlikely to be dependent on the particular genetics of this strain because ICMT-, pheromone- and GPCR-encoding genes showed tissue-enhanced expression patterns also in other examined species (see below, Supplementary Table 10.). Together, these experiments reveal the first molecular mechanisms of tissue differentiation in mushroom fruiting bodies and demonstrate that ICMTs and the mating pheromone pathway are required for tissue differentiation in *C. cinerea*.

To investigate how these genes could mechanistically drive tissue differentiation, we considered two plausible explanations: (1) pheromones/GPCRs act as signal-receptor systems that transmit (spatial) information across tissues, or (2) being required for hyphal fusions that occur in fruiting bodies³¹. In the absence of elaborate tools in *C. cinerea*, we analyzed our RNA-seq data for evidence of either explanation. We found a uniform expression of putative hyphal fusion genes of *C. cinerea* across tissues (Supplementary Table 11), including the universal veil, in which hyphal fusions do not occur³¹, rendering hyphal fusion alone as a mechanistic explanation unlikely.

Next, we investigated whether mutants with the same phenotype (Fig. 3/b) but different genes knocked out have similar transcriptomic signatures, indicative of the existence of distinct ligand-receptor pathways that control the development of different tissues. To do this, we performed RNA-seq on apical and basal regions of mutants with matching phenotypes at the P2 stage (Fig. 3/c): two mutants displaying cyclops-caps (*icmt Δ152667*, GPCR *Δ203704*) and two having delayed gill development (pheromone *Δ492306*, GPCR *Δ356637*), as well as wild type. Principal component analyses revealed that along PC2, samples cluster by phenotype (Fig. 3/c), indicating similar transcriptome states irrespective of gene family membership. These patterns were more pronounced in the apical compared to basal regions, demonstrating that, as expected for both phenotypes the mutated genes are spatially restricted to the apical region. Both the PCA and differentially expressed genes (DEGs, Fig. 3/c,d, Supplementary Table 12.) indicate larger differences in cyclops-cap mutants than delayed-gill mutants relative to wt. GO analysis of DEGs highlighted putative functional consequences of gene deletions, such as differences in cell wall remodeling and transcription regulation (Supplementary Table 13). Apical tissue-enhanced genes from our earlier analyses (Fig. 2) were enriched in apical-downregulated DEGs of cyclops-cap mutants ($P = 1.48 \times 10^{-4} - 1.15 \times 10^{-18}$, hypergeometric test, Supplementary Table 14). Conversely, apical tissue-repressed genes were enriched in apical-upregulated DEGs ($P = 2.5 \times 10^{-3} - 1.52 \times 10^{-81}$, hypergeometric test, Supplementary Table 14), indicating dysregulation of transcription in mutants. Overall, the similarity of phenotypes and underlying transcriptional states indicate that certain ICMTs, pheromone precursors and GPCRs form tissue-specific differentiation pathways, via their ability to act as a ligand-receptor system.

Comparative genomics suggests co-option of mating genes for tissue differentiation

In Agaricomycetes, pheromones and pheromone-sensing GPCRs regulate mating, that is, recognition and fusion of compatible hyphae to form a fertile dikaryon¹⁷. Because mating precedes fruiting body

formation by up to years, our results imply that pheromones, their post-translational modifiers ICMTs, and GPCRs have acquired a new role in tissue differentiation during evolution. Pheromones and GPCRs form multiallelic and paralogous gene families which, under the prevailing model, favors outbreeding and mate discrimination based on compatibility relatedness^{17,19}. However, this model offers a less clear explanation for the role of paralogy²⁰ and provides no explanation for that in ICMTs.

To uncover the role of paralogy in ICMTs, mating pheromones and GPCRs and homeodomain TFs, we analyzed copy number variation in the genomes of 29 non-agaricomycete and 158 agaricomycete fungi. Because the identification of pheromones is notoriously difficult using conventional genome annotation, we used a consensus from 49 neural networks and a custom HMM to robustly identify pheromone precursors in genome-wide collections of translated ORFs (Supplementary Figs. 17-18, Supplementary Table 15). Mushroom-forming fungi have strongly expanded repertoires of ICMTs, pheromones and GPCRs compared to other fungi (Fig. 4/a). Comparative phylogenomic analyses of ICMTs and GPCRs additionally identified a wave of gene duplication coincident with the evolution of fruiting bodies (Fig 4/b).

Was this expansion driven by mating, as widely held¹⁸, or by the evolution of complex fruiting bodies? To address this, we first assessed paralogy in homeodomain TFs, which form the 'A' mating factor in Agaricomycetes, and should display higher copy number variation than pheromones and GPCRs, if mating drove the evolution of paralogy¹⁷. However, we find the opposite: the mean copy number of homeodomain TFs is 1.15 per species in the Agaricomycetes (Fig. 4/a), indicating limited paralogy. Second, we examined gene repertoires of four species which form complex tissues but in which the pheromone-GPCR system no longer regulates mating^{17,20,32,33}. If selection for outcrossing maintains paralogs, these species should have reduced gene repertoires. However, while multiallelism was lost in these species^{17,32}, they do not possess fewer genes of any family (Fig. 4/a), indicating that selective pressures not related to mating must be driving the evolution of ICMT, pheromone and GPCR gene paralogy.

Previous models posited that both multiallelism and paralogy evolved to promote outbreeding¹⁸, even though multiallelism alone would be sufficient^{17,19}. This study now provides an explanation for this unsolved enigma: our data indicate that paralogy is driven by selection for tissue-specific expression, rather than mating alone drives paralogy. Indeed, tissue-specific expression can arise only if promoters of paralogous genes are free to diverge. Thus, we posit that paralogy evolved to support tissue development, which is then an additional driver of mating locus evolution in Agaricomycetes.

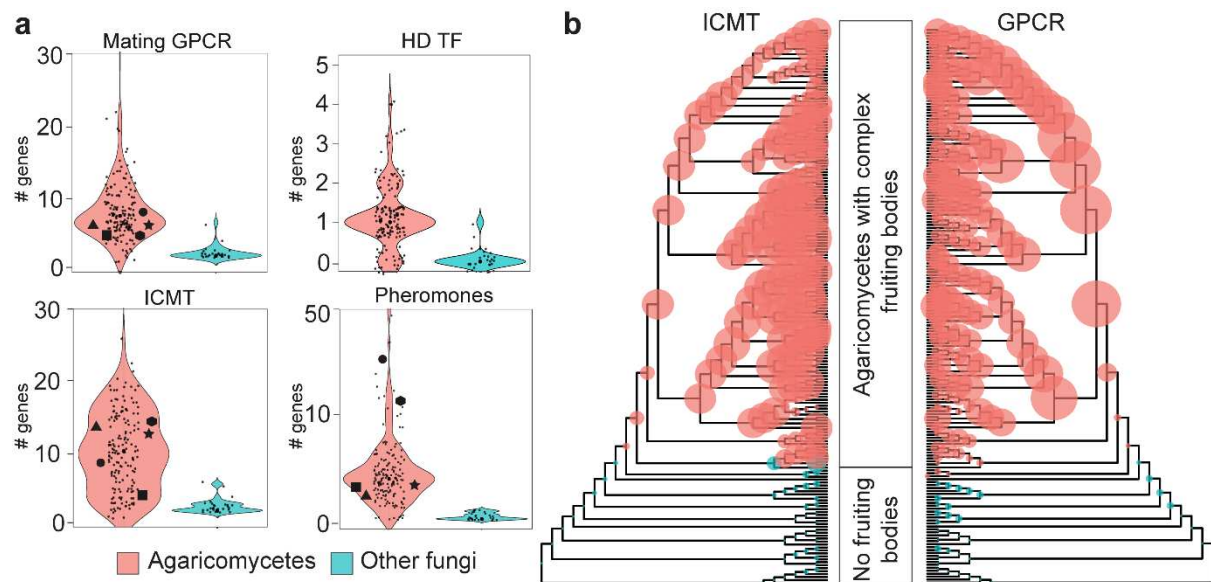


Fig. 4. The evolution of genes underlying tissue development in Agaricomycetes. a. GPCR, ICMT and pheromone diversity across 158 agaricomycete and 29 non-agaricomycete species. GPCR and ICMT copy numbers are based on pfam HMM search, whereas pheromone copy numbers are based on a consensus

prediction of 49 machine learning methods. Enlarged black dots denote *C. cinerea*, while the triangle, rectangle, hexagon and star denote *Heterobasidion annosum*, *Agaricus bisporus*, *Volvariella volvacea* and *Phanerochaete chrysosporium*, respectively, in which pheromones and GPCRs lost their mating function. **b.** Gene duplication and loss reconstruction of the ICMT and pheromone-sensing GPCRs families across Agaricomycetes and related fungi. Circles at nodes are proportional to the extant (on tip branches) and ancestral copy numbers.

Evolutionary conservation of tissue-enhanced gene expression

To assess if tissue-restricted expression is widespread, and to reconcile its existence with the evolution of distinct tissues during the Jurassic explosion of morphologies in Agaricomycetes⁷, we used a comparative transcriptomic approach. Analyses of sequence conservation amongst proteins encoded by tissue-enhanced genes of *C. cinerea* based on 189 fungal genomes revealed that 27.9% of tissue-enhanced genes are older than the emergence of fruiting bodies in the Agaricomycetes, ca 353.15 million years ago (Mya), whereas 42.2% predate the evolution of fruiting bodies with caps and stipes in the Agaricales, ca. 166.37 Mya. This indicates that ancient genes are common in *C. cinerea* tissue transcriptomes. Tissue-enhanced genes and TFs were enriched in one ancient phylostratum (Dikarya), and predominantly in the Agaricales (Supplementary Fig. 19), the latter coinciding with the emergence of core and surface tissues¹⁶ and the radiation of complex morphologies.

To examine whether tissue-enhanced expression is widespread across mushroom-forming fungi, we analyzed published RNA-seq datasets of seven species in which different tissues were sampled separately (Fig. 5/b) using the same DESeq2-GLM as above. Across the seven species 7-36% of the genes were tissue-enhanced (Supplementary Table 16), indicating that tissue-enhanced patterns are widespread among mushroom-forming fungi. Overall, we detected low evolutionary conservation of tissue-enhanced expression: in matching tissues 1.3–11.7% (Supplementary Table 16) and in tissue-agnostic analyses 7.2-18.0% (Supplementary Table 16) of orthologs of the *C. cinerea* tissue-enhanced genes were also tissue-enhanced in other species. This low conservation is comparable to findings in other organisms³⁴, and is expected given the large evolutionary dynamics and homoplasy of fruiting body morphologies⁷. Nevertheless, some highly conserved genes were detected both at the level of gene families and in strict single-copy gene orthogroups. We identified 186 gene families containing on average >40% tissue-enhanced genes in ≥5 species (Fig. 5/d, Supplementary Table 17) and 130 strict single-copy orthogroups that were tissue-enhanced in ≥5 species (Fig. 5/c, Supplementary Table 16). Some gene families, such as ICMTs, hydrophobins and Gti1/Pac2 and TEA/ATTS TFs were top-ranked in both lists and represent candidates for driving tissue differentiation.

To examine evolutionary conservation of *C. cinerea* tissue transcriptomes through the development, we calculated transcriptome age indices (TAI). TAI captured increasing transcriptome conservation as development progresses ($P = 2.88 \times 10^{-16}$, permutation test) (Fig. 5/a), indicating that later stages of fruiting body development (P1 and P2) express evolutionarily more conserved transcriptomes. Conservation is most pronounced in gills, which express the most conserved gene set ($P = 5.23 \times 10^{-6}$, permutation test). Gills harbor basidia, the oldest cell type in fruiting bodies, which produce sexual spores. Other core tissues express less conserved, whereas surface tissues express the least conserved transcriptome (Fig. 5/a), forming a transcriptome age gradient from the gills towards the surface of the fruiting body. The young transcriptome of the surface tissues is consistent with this structure being a recent key innovation in agaricomycete evolution¹⁶ and that evolutionarily young genes involved in defense (*pic1*, *ccp1*, *ccl2*, an aerolysin-like toxin), cell adhesion (*cgl1* and *cgl2*) and hydrophobicity (ten hydrophobins) were among the 96 genes that drive the surface/core TAI-based division most significantly (Supplementary Fig. 20). Overall, our analyses reveal that gills, the last tissue to emerge in fruiting bodies, express the evolutionarily oldest transcriptome, which is different from models postulated for other organisms, but reflects the role of fruiting bodies as structures that provide protection and scaffold for the developing sporogenous tissue.

Taken together, our analyses revealed a broad evolutionary conservation of tissue-enhanced expression and the evolutionary patterns underlying tissue-transcriptomes, highlighting key genes and gene families associated with tissue expression profiles.

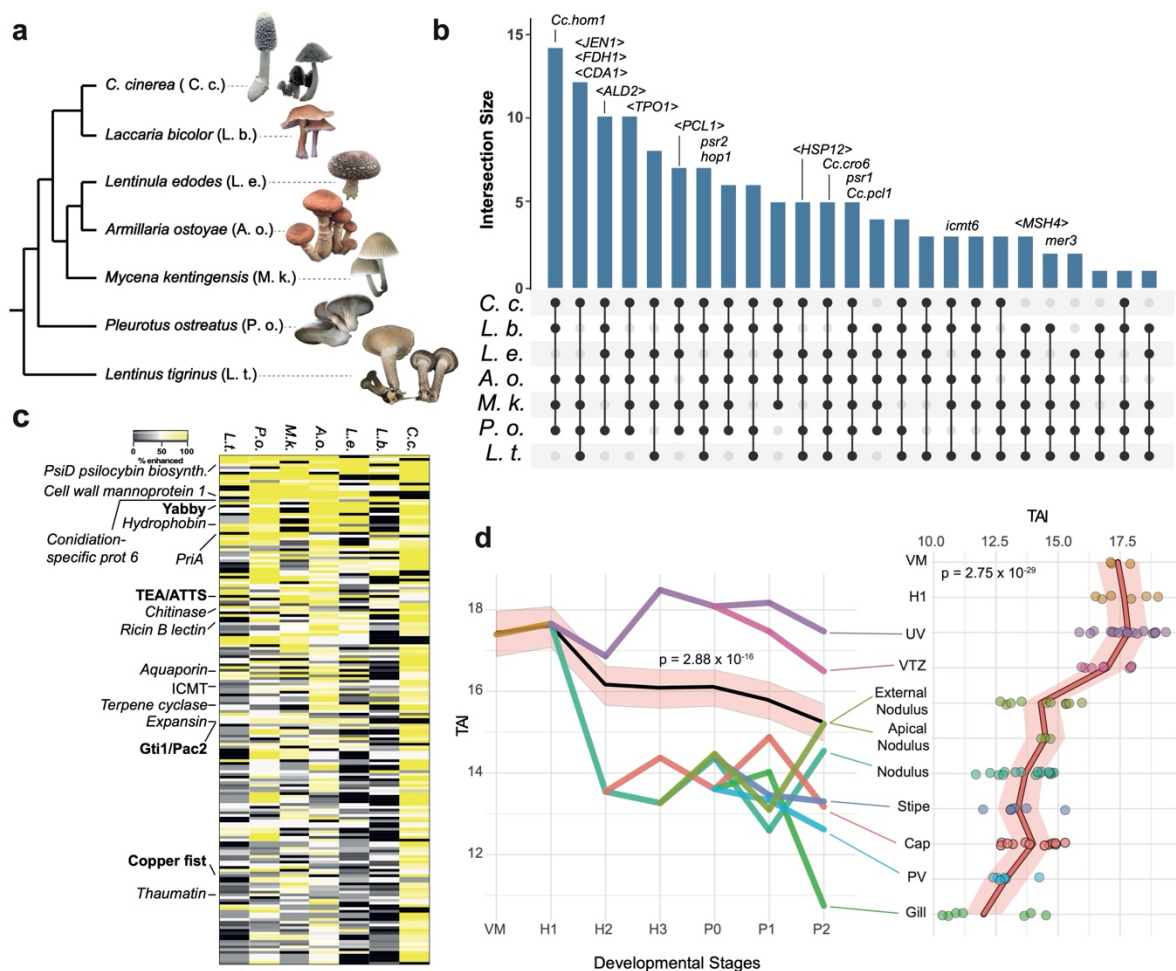


Fig. 5. The conservation of tissue-restricted expression in Agaricomycetes. **a.** Tissue-specific transcriptome state conservation measured by the transcriptome age index (TAI). The left panel shows the TAI as a function of developmental stages. The black line indicates the overall, while colored lines indicate tissue-specific TAI trajectories. The P value indicates the significant fit to the late conservation model implemented in the myTAI package. The right panel shows TAI values by tissue type. Each dot corresponds to a replicate across different developmental stages. **b.** Seven species used in assessments of tissue-enhanced expression conservation. **c.** UpsetR representation of 130 strict single-copy orthogroups that were tissue-enhanced in ≥ 5 species, with key conserved genes highlighted by name. Genes marked by <> are named based on orthology to yeast sequences. **d.** Heatmap showing the percentage of tissue-enhanced genes in 186 gene families that contain on average >40% tissue-enhanced genes in ≥ 5 species. Some key fruiting-related gene families are highlighted. Bold indicates transcription factor families.

Discussion

Genetic mechanisms of tissue differentiation have been studied in great detail in animals and plants, whereas no information has been available in mushroom-forming fungi, a diverse lineage of ca 40,000 species. Here, we present evidence that, despite their hyphal organization, agaricomycete fruiting bodies display hallmarks of *bona fide* tissues: spatiotemporally distinct transcriptional states; unique functional footprints and increasing functions specialization through time. Thus, we postulate that these data blur the boundaries between pseudotissues and tissues.

Our experiments in the model mushroom *C. cinerea* showed that an ancient fungal signaling system comprising pheromones, GPCRs and ICMTs (involved in pheromone maturation) are key genetic components of tissue differentiation. Ancestrally, pheromones and GPCRs regulate sexual compatibility and mate recognition¹⁹ as well as the formation of a sexually competent

dikaryon^{17,19,35,36}. Our results reveal a hitherto undiscovered role of these gene families in tissue differentiation, which we postulate was gained by co-option in early fruiting body-forming fungi. Consequently, Agaricomycetes use the same genetic system for determining sexual compatibility and tissue differentiation, two processes separated in time and space. This is unusual among complex multicellular organisms, which usually have distinct genetic pathways for sex determination and development^{37,38}. What made the pheromone-GPCR system suitable for tissue differentiation in Agaricomycetes? In contrast to animals and plants, in which cell migration and division plane orientation generate complex structures^{2,4}, fungal fruiting bodies are established by regulated hyphal growth and branching. Coordinating these processes requires communication between hyphae. Given the ancestral role of the pheromone/GPCR system in signaling across hyphae³⁵, this pathway may have been poised to be recruited for hypha-based tissues.

How pheromones and GPCRs mechanistically regulate tissue differentiation remains to be explored. Our data raise the possibility that pheromone cocktails convey positional information through the mucilaginous extracellular matrix that exists between fruiting body hyphae³¹, which requires further research in the future. Similarly, intracellular pathways downstream of pheromone-GPCR signaling in tissue differentiation remain unknown. In metazoans and plants complex networks of homeobox and MADS-box TFs are involved in developmental patterning⁴. Fungal homologs in the homeobox superfamily, the mating-type TFs, regulate dikaryon formation but their activity is not sufficient to initiate fruiting body development or cell-type differentiation³⁹. Thus, other gene families, perhaps some of the tissue-enhanced TFs we uncovered, likely underlie tissue patterning.

Unlike in animals and plants, key molecular drivers of tissue differentiation – equivalents of Hox and ABC genes – have remained unknown in mushroom-forming fungi. This study provides transcriptomic insights and the first genetic mechanisms of tissue differentiation and fruiting body development in this group of organisms. Tissues evolved convergently in mushrooms, plants and animals, and this study sheds light on how mushroom-forming fungi evolved the ability to develop complex fruiting bodies, overcoming the physiological constraints imposed by hyphae. The solutions in fungi might therefore illuminate general principles of the evolution of complexity.

Acknowledgements

We appreciate useful discussions with Erika Kothe, Ursula Kues and Marjatta Raudaskoski on the enigma around mating pheromones and GPCRs in Agaricomycetes. The authors thank Árpád Bálint and Krisztina Krizsán for their initial help in setting up the LCM RNA-seq experiment. The Authors acknowledge support by the National Research Development and Innovation Office (Grant No. OTKA 142188, 2022-2.1.1-NL-2022-00005, TKP-2021-EGA-05, TKP2021-EGA-09 and SNN 152010), the European Research Council (grant no. 101086900), Horizon 2020 (SGA No. 739593) and the Hungarian Academy of Sciences (LENDULET Grant 2018-342). T.V. was supported by the Kew Research Fellowship. D.A.Sz. appreciates support of the National Academy of Scientist Education Program of the National Biomedical Foundation. Z.M. (grant no. BO/00269/24/8) and G.B. were supported by the Janos Bolyai Research Scholarship. J.P.L. and Z.H. were supported by China Scholarship Council (Grant No. 202308220156, 202008110168). Z.H. acknowledges the support of the Stipendium Hungaricum Scholarship (grant no. SHE-21070-006/2020). H.G.D thanks the Max Planck Society, the Royal Society (grant no. RSWF\R1\241004) and the BMBF-funded de.NBI Cloud (grant nos. 031A532B, 031A533A, 031A533B, 031A534A, 031A535A, 031A537A, 031A537B, 031A537C, 031A537D, 031A538A) for computational support.

Data availability

Data associated with this paper will be made available before acceptance of the manuscript.

References

1. Rensing, S. A. (Why) Does Evolution Favour Embryogenesis? *Trends in Plant Science* **21**, 562–573 (2016).
2. Knoll, A. H. The Multiple Origins of Complex Multicellularity. *Annu. Rev. Earth Planet. Sci.* **39**, 217–239 (2011).
3. Nagy, L. G. *et al.* The biodiversity, genomics, ecology and evolution of mushroom-forming fungi. *Nat. Rev. Biodivers.* 1–16 (2025) doi:10.1038/s44358-025-00107-z.
4. Meyerowitz, E. M. Plants Compared to Animals: The Broadest Comparative Study of Development. *Science* **295**, 1482–1485 (2002).
5. Nagy, L. G., Varga, T., Csernetics, Á. & Virágh, M. Fungi took a unique evolutionary route to multicellularity: Seven key challenges for fungal multicellular life. *Fungal Biology Reviews* **34**, 151–169 (2020).
6. Irie, N. & Kuratani, S. The developmental hourglass model: a predictor of the basic body plan? *Development* **141**, 4649–4655 (2014).
7. Varga, T. *et al.* Megaphylogeny resolves global patterns of mushroom evolution. *Nat Ecol Evol* **3**, 668–678 (2019).
8. Hibbett, D., Nagy, L. G. & Nilsson, R. H. Fungal diversity, evolution, and classification. *Curr Biol* **35**, R463–R469 (2025).
9. Kües, U. & Navarro-Gonzalez, M. How do Agaricomycetes shape their fruiting bodies? 1. Morphological aspects of development. *Fungal Biology Reviews* **29**, 63–97 (2015).
10. Cléménçon, H. *Cytology and Plectology of the Hymenomycetes*. (J. Cramer, Stuttgart, 2012).
11. Nagy, L. G., Kovács, G. M. & Krizsán, K. Complex multicellularity in fungi: evolutionary convergence, single origin, or both? *Biol Rev Camb Philos Soc* **93**, 1778–1794 (2018).
12. Stajich, J. E. *et al.* The Fungi. *Current Biology* **19**, R840–R845 (2009).
13. Nagy, L. G. *et al.* The biodiversity, genomics, ecology and evolution of mushroom-forming fungi. *Nat. Rev. Biodivers.* **2**, 24–39 (2026).

14. Kües, U. Life History and Developmental Processes in the Basidiomycete *Coprinus cinereus*. *Microbiol Mol Biol Rev* **64**, 316–353 (2000).
15. Sakamoto, Y. Influences of environmental factors on fruiting body induction, development and maturation in mushroom-forming fungi. *Fungal Biology Reviews* **32**, 236–248 (2018).
16. Varga, T., Földi, C., Bense, V. & Nagy, L. G. Radiation of mushroom-forming fungi correlates with novel modes of protecting sexual fruiting bodies. *Fungal Biology* **126**, 556–565 (2022).
17. Nieuwenhuis, B. P. S. *et al.* Evolution of uni- and bifactorial sexual compatibility systems in fungi. *Heredity* **111**, 445–455 (2013).
18. Kües, U. From two to many: Multiple mating types in Basidiomycetes. *Fungal Biology Reviews* **29**, 126–166 (2015).
19. James, T. Y. Why mushrooms have evolved to be so promiscuous: Insights from evolutionary and ecological patterns. *Fungal Biology Reviews* **29**, 167–178 (2015).
20. James, T. Y., Srivilai, P., Kües, U. & Vilgalys, R. Evolution of the bipolar mating system of the mushroom *Coprinellus disseminatus* from its tetrapolar ancestors involves loss of mating-type-specific pheromone receptor function. *Genetics* **172**, 1877–1891 (2006).
21. Musser, J. M. & Wagner, G. P. Character trees from transcriptome data: Origin and individuation of morphological characters and the so-called ‘species signal’. *J Exp Zool B Mol Dev Evol* **324**, 588–604 (2015).
22. Krizsán, 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 (2019).
23. Sonawane, A. R. *et al.* Understanding Tissue-Specific Gene Regulation. *Cell Rep* **21**, 1077–1088 (2017).
24. Drost, H.-G., Janitzka, P., Grosse, I. & Quint, M. Cross-kingdom comparison of the developmental hourglass. *Current Opinion in Genetics & Development* **45**, 69–75 (2017).

- 532 25. Digre, A. & Lindskog, C. The Human Protein Atlas—Spatial localization of the human
533 proteome in health and disease. *Protein Sci* **30**, 218–233 (2021).
- 534 26. Uhlén, M. *et al.* Transcriptomics resources of human tissues and organs. *Mol Syst*
535 *Biol* **12**, 862 (2016).
- 536 27. Nakamori, T. & Suzuki, A. Defensive role of cystidia against *Collembola* in the
537 basidiomycetes *Russula bella* and *Strobilurus ohshima*. *Mycological Research* **111**,
538 1345–1351 (2007).
- 539 28. Linder, M. B., Szilvay, G. R., Nakari-Setälä, T. & Penttilä, M. E. Hydrophobins: the
540 protein-amphiphiles of filamentous fungi. *FEMS Microbiol Rev* **29**, 877–896 (2005).
- 541 29. Wu, H. *et al.* Functional characterization of a Con7-related transcription factor in
542 *Coprinopsis cinerea* indicates evolutionary conservation of morphogenetic roles.
543 *Microbiological Research* **306**, 128462 (2026).
- 544 30. Michaelis, S. & Barrowman, J. Biogenesis of the *Saccharomyces cerevisiae*
545 pheromone α -factor, from yeast mating to human disease. *Microbiol. Mol. Biol. Rev.* **76**,
546 626–651 (2012).
- 547 31. van der Valk, P. & Marchant, R. Hyphal ultrastructure in fruit-body primordia of the
548 basidiomycetes *Schizophyllum commune* and *Coprinus cinereus*. *Protoplasma* **95**, 57–72
549 (1978).
- 550 32. Bao, D. *et al.* Sequencing and Comparative Analysis of the Straw Mushroom
551 (*Volvariella volvacea*) Genome. *PLoS One* **8**, e58294 (2013).
- 552 33. Foulongne-Oriol, M. *et al.* Mating-Type Locus Organization and Mating-Type
553 Chromosome Differentiation in the Bipolar Edible Button Mushroom *Agaricus bisporus*.
554 *Genes (Basel)* **12**, 1079 (2021).
- 555 34. Mantica, F. *et al.* Evolution of tissue-specific expression of ancestral genes across
556 vertebrates and insects. *Nat Ecol Evol* **8**, 1140–1153 (2024).
- 557 35. Jones, S. K., Jr & Bennett, R. J. Fungal mating pheromones: choreographing the
558 dating game. *Fungal Genet. Biol.* **48**, 668–676 (2011).

559 36. Kothe, E. Mating Types and Pheromone Recognition in the Homobasidiomycete
560 Schizophyllum commune. *Fungal Genetics and Biology* **27**, 146–152 (1999).

561 37. Mignerot, L. & Coelho, S. M. The origin and evolution of the sexes: Novel insights
562 from a distant eukaryotic lineage. *C R Biol* **339**, 252–257 (2016).

563 38. Luthringer, R. *et al.* Repeated co-option of HMG-box genes for sex determination in
564 brown algae and animals. *Science* **383**, eadk5466 (2024).

565 39. Vonk, P. J. & Ohm, R. A. The role of homeodomain transcription factors in fungal
566 development. *Fungal Biology Reviews* **32**, 219–230 (2018).

567 40. Carson, F. L. *Histotechnology: A Self-Instructional Text*. (American Society of Clinical
568 Pathologists Press, Chicago, 2001).

569 41. Espina, V. *et al.* Laser-capture microdissection. *Nat Protoc* **1**, 586–603 (2006).

570 42. Thakare, D. *et al.* RNA-Seq analysis of laser-capture microdissected cells of the
571 developing central starchy endosperm of maize. *Genomics Data* **2**, 242–245 (2014).

572 43. Teichert, I., Wolff, G., Kück, U. & Nowrousian, M. Combining laser microdissection
573 and RNA-seq to chart the transcriptional landscape of fungal development. *BMC*
574 *Genomics* **13**, 511 (2012).

575 44. Baker, S. C. *et al.* The External RNA Controls Consortium: a progress report. *Nat*
576 *Methods* **2**, 731–734 (2005).

577 45. External RNA Controls Consortium. Proposed methods for testing and selecting the
578 ERCC external RNA controls. *BMC Genomics* **6**, 150 (2005).

579 46. Picelli, S. *et al.* Smart-seq2 for sensitive full-length transcriptome profiling in single
580 cells. *Nat Methods* **10**, 1096–1098 (2013).

581 47. Picelli, S. *et al.* Full-length RNA-seq from single cells using Smart-seq2. *Nat Protoc*
582 **9**, 171–181 (2014).

583 48. Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput
584 Sequence Data. <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.

49. Ewels, P., Magnusson, M., Lundin, S. & Käller, M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**, 3047–3048 (2016).
50. Bushnell, B. BBMap: A Fast, Accurate, Splice-Aware Aligner. in (Ernest Orlando Lawrence Berkeley National Laboratory, Berkeley, CA (US), 2014).
51. Bushnell, B. BBMap - Official BBTools Suite. <https://bbmap.org/>.
52. Bushnell, B., Rood, J. & Singer, E. BBMerge – Accurate paired shotgun read merging via overlap. *PLOS ONE* **12**, e0185056 (2017).
53. Hegedüs, B. *et al.* Morphogenesis, starvation, and light responses in a mushroom-forming fungus revealed by long-read sequencing and extensive expression profiling. *Cell Genomics* **5**, 100853 (2025).
54. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
55. Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *GigaScience* **10**, giab008 (2021).
56. Lawrence, M. *et al.* Software for Computing and Annotating Genomic Ranges. *PLOS Computational Biology* **9**, e1003118 (2013).
57. Bullard, J. H., Purdom, E., Hansen, K. D. & Dudoit, S. Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. *BMC Bioinformatics* **11**, 94 (2010).
58. Pertea, G. & Pertea, M. GFF Utilities: GffRead and GffCompare. Preprint at <https://doi.org/10.12688/f1000research.23297.1> (2020).
59. Charif, D. & Lobry, J. R. SeqinR 1.0-2: A Contributed Package to the R Project for Statistical Computing Devoted to Biological Sequences Retrieval and Analysis. in *Structural Approaches to Sequence Evolution: Molecules, Networks, Populations* (eds Bastolla, U., Porto, M., Roman, H. E. & Vendruscolo, M.) 207–232 (Springer, Berlin, Heidelberg, 2007). doi:10.1007/978-3-540-35306-5_10.

612 60. Risso, D., Schwartz, K., Sherlock, G. & Dudoit, S. GC-content normalization for RNA-
613 Seq data. *BMC Bioinformatics* **12**, 480 (2011).

614 61. Zhao, Y., Wong, L. & Goh, W. W. B. How to do quantile normalization correctly for
615 gene expression data analyses. *Sci Rep* **10**, 15534 (2020).

616 62. Hicks, S. C. & Irizarry, R. A. quantro: a data-driven approach to guide the choice of
617 an appropriate normalization method. *Genome Biol* **16**, 117 (2015).

618 63. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and
619 dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).

620 64. Hicks, S. C. *et al.* Smooth quantile normalization. *Biostatistics* **19**, 185–198 (2018).

621 65. R: The R Project for Statistical Computing. <https://www.r-project.org/>.

622 66. Suzuki, R. & Shimodaira, H. Pvclust: an R package for assessing the uncertainty in
623 hierarchical clustering. *Bioinformatics* **22**, 1540–1542 (2006).

624 67. Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M. & Hornik, K. *Cluster: Cluster*
625 *Analysis Basics and Extensions*. (2026).

626 68. Hoffman, G. E. & Schadt, E. E. variancePartition: interpreting drivers of variation in
627 complex gene expression studies. *BMC Bioinformatics* **17**, 483 (2016).

628 69. Ignatiadis, N., Klaus, B., Zaugg, J. B. & Huber, W. Data-driven hypothesis weighting
629 increases detection power in genome-scale multiple testing. *Nat Methods* **13**, 577–580
630 (2016).

631 70. Stephens, M. False discovery rates: a new deal. *Biostatistics* **18**, 275–294 (2017).

632 71. Hausser, J. & Strimmer, K. Entropy inference and the James-Stein estimator, with
633 application to nonlinear gene association networks. Preprint at
634 <https://doi.org/10.48550/arXiv.0811.3579> (2009).

635 72. Christopoulos, D. T. Developing methods for identifying the inflection point of a
636 convex/concave curve. Preprint at <https://doi.org/10.48550/arXiv.1206.5478> (2014).

637 73. Yanai, I. *et al.* Genome-wide midrange transcription profiles reveal expression level
638 relationships in human tissue specification. *Bioinformatics* **21**, 650–659 (2005).

639 74. Jones, P. *et al.* InterProScan 5: genome-scale protein function classification.
640 *Bioinformatics* **30**, 1236–1240 (2014).

641 75. Alexa, A. & Rahnenfuhrer, J. topGO: Enrichment Analysis for Gene Ontology.
642 Bioconductor version: Release (3.18) <https://doi.org/10.18129/B9.bioc.topGO> (2023).

643 76. Beier, S., Bolger, A. M., Bolger, M. E., Schwacke, R. & Usadel, B. Trimmomatic: a
644 decade of feature-rich, high-performance NGS read preprocessing. *Bioinformatics* **42**,
645 btag331 (2026).

646 77. Zhang, K., Tappero, R., Ruytinx, J., Branco, S. & Liao, H.-L. Disentangling the role of
647 ectomycorrhizal fungi in plant nutrient acquisition along a Zn gradient using X-ray
648 imaging. *Science of The Total Environment* **801**, 149481 (2021).

649 78. Merényi, Z. *et al.* Gene age shapes the transcriptional landscape of sexual
650 morphogenesis in mushroom-forming fungi (Agaricomycetes). *eLife* **11**, e71348 (2022).

651 79. Ke, H.-M. *et al.* *Mycena* genomes resolve the evolution of fungal bioluminescence.
652 *Proc. Natl. Acad. Sci. U.S.A.* **117**, 31267–31277 (2020).

653 80. Sipos, G. *et al.* Genome expansion and lineage-specific genetic innovations in the
654 forest pathogenic fungi *Armillaria*. *Nat Ecol Evol* **1**, 1931–1941 (2017).

655 81. Ruytinx, J. *et al.* A transcriptomic atlas of the ectomycorrhizal fungus *Laccaria*
656 *bicolor*. *Microorganisms* **9**, 2612 (2021).

657 82. Steinegger, M. & Söding, J. Clustering huge protein sequence sets in linear time. *Nat*
658 *Commun* **9**, 2542 (2018).

659 83. Demichev, V., Messner, C. B., Vernardis, S. I., Lilley, K. S. & Ralser, M. DIA-NN:
660 neural networks and interference correction enable deep proteome coverage in high
661 throughput. *Nat Methods* **17**, 41–44 (2020).

662 84. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-
663 sequencing and microarray studies. *Nucleic Acids Res* **43**, e47 (2015).

664 85. Pareek, M. *et al.* Preassembled Cas9 Ribonucleoprotein-Mediated Gene Deletion
665 Identifies the Carbon Catabolite Repressor and Its Target Genes in *Coprinopsis cinerea*.
666 *Appl Environ Microbiol* **88**, e0094022 (2022).

86. Schmieder, S. S. *et al.* Bidirectional Propagation of Signals and Nutrients in Fungal Networks via Specialized Hyphae. *Current Biology* **29**, 217-228.e4 (2019).
87. Virágh, M. *et al.* Basidiomycete fruiting body morphology correlates with the convergent loss of photolyase-related proteins. *iScience* **29**, (2026).
88. Emms, D. M. & Kelly, S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol* **16**, 157 (2015).
89. Turek, C. *et al.* mulea: An R package for enrichment analysis using multiple ontologies and empirical false discovery rate. *BMC Bioinformatics* **25**, 334 (2024).
90. Drost, H.-G., Gabel, A., Liu, J., Quint, M. & Grosse, I. myTAI: evolutionary transcriptomics with R. *Bioinformatics* **34**, 1589–1590 (2018).
91. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* **30**, 772–780 (2013).
92. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
93. Chen, K., Durand, D. & Farach-Colton, M. NOTUNG: a program for dating gene duplications and optimizing gene family trees. *J Comput Biol* **7**, 429–447 (2000).
94. Lin, Z. *et al.* Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science* **379**, 1123–1130 (2023).
95. Berend, G. Sparsity Makes Sense: Word Sense Disambiguation Using Sparse Contextualized Word Representations. in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (eds Webber, B., Cohn, T., He, Y. & Liu, Y.) 8498–8508 (Association for Computational Linguistics, Online, 2020). doi:10.18653/v1/2020.emnlp-main.683.
96. Rajković, A. *et al.* HMMER web server: 2026 update. *Nucleic Acids Research* **54**, W272–W278 (2026).

Supplementary table legends

Supplementary Table 1. List of histology samples and related primordium features, RNA extraction statistics, spike-in statistics and the results of the outlier detection process based on spike-in statistics, PCA, hierarchical clustering and histology images.

Supplementary Table 2. RNAseq library statistics. Library features before and after quality trimming, trimming statistics, STAR mapping and read count statistics.

Supplementary Table 3. The list of GC and qsmooth normalised expression values and the results of the deviation contrast-based differential expression analyses: log2 fold change values, false discovery rates and tissue-enhanced or -repressed categories.

Supplementary Table 4. The results of bulk proteomic analyses of apical and basal regions of stage 2 primordia of *C. cinerea*.

Supplementary Table 5. Detailed results of Gene Ontology enrichment analyses across nine major tissue types and the three main ontologies.

Supplementary Table 6. Enrichment analyses of transcription factor families among tissue-enhanced and -repressed genes, by family.

Supplementary Table 7. Gene names, protein IDs, expression levels and tissue-enhanced/repressed patterns of ICMT, pheromone precursor and GPCR genes of *C. cinerea* in our ultralow-input RNA-seq dataset.

Supplementary Table 8. Expression levels of annotated and newly predicted tissue-enhanced or -repressed CAAX proteins and putative CAAX peptides across our RNA-seq dataset.

Supplementary Table 9. Expression patterns of pheromone-precursor genes during *C. cinerea* early development.

Supplementary Table 10. Tissue-specific expression of ICMT, pheromone precursor and mating GPCR-encoding genes across bulk transcriptomes analysed in this study.

Supplementary Table 11. Orthology and expression patterns of cell fusion genes in *C. cinerea*.

Supplementary Table 12. Transcriptome and differential expression data of wild-type and four KO mutant primordium stage 2 samples dissected into apical and basal tissues.

Supplementary Table 13. Detailed results of Gene Ontology enrichment analyses of differentially expressed genes between wild-type and knock-out mutant apical and basal tissues.

Supplementary Table 14. Enrichment of tissue-enhanced and -repressed genes among differentially expressed genes obtained in mutant RNA-Seq analyses.

Supplementary Table 15. Predicted copy numbers and sequences of mating pheromones across the comparative genomic dataset of 187 fungal species. For complete species names, see Hou et al 2025 Curr.Biol.

Supplementary Table 16. Evolutionary conservation of tissue-enhanced gene expression across bulk transcriptomes of six Agaricomycetes species.

Supplementary Table 17. Evolutionary conservation of tissue-enhanced expression in InterPro groups across six species, listing 186 IPR terms comprising $\geq 40\%$ tissue enhanced/repressed genes (column AC).

Supplementary Data 1. Representative images of whole fruiting bodies and cross-sections for each of the mutants generated in this study.

Methods

Histology

Strain and culture

Coprinopsis cinerea strain A43mutB43mut pab1-1 was grown on YMG-Trp media (0.4 m/m% yeast extract, 1 m/m% maltose, 0.4 m/m% glucose, 0.01 m/m% DL Tryptophan, 1 m/m% agar) for 5 days at 37 °C in the dark and then at 25 °C in a 12-hour-dark/12-hour-light regime. We sampled primordia within defined size ranges on specific days to ensure consistency, as follows. Hyphal knot stages (up to ~0.5 mm in diameter) were gathered on the first two days of the dark/light regime, intermediate stages were sampled on days 2-4, and primordia bigger than 1.5-2 mm were picked after the 4th day. We performed sampling within a four-hour window after the end of the dark period each day.

Resin-based embedding

The harvesting process involved excising an agar cube bearing hyphal knot or primordia, and subsequently transferring it to a Half-strength Karnovsky's Fixative consisting of 1% paraformaldehyde and 2.5% glutaraldehyde. Samples were placed under the vacuum three time then incubated overnight at 4 °C with rotational agitation. For embedding, the samples were dehydrated by incubating in an ascending ethanol series with 10 %, 30 %, 50 %, 70%, 90% and finally 100 % ethanol, respectively with rotational agitation at room temperature. Following this, the samples were embedded in TV7100 resin (Heraus Kulzer, Wehrheim, Germany). The blocks were sectioned at 10 μ m thickness by a rotary microtome (Leica HistoCore Autocut R) equipped with a tungsten carbide knife (16 cm, profile-d). The sections were stained with 0.01% w/v methylene blue (Sigma-Aldrich) and analyzed using a Leica DM 2500 Microscope with a camera attached.

778 Transcriptomics

779 Cryo-embedding and sectioning

780 We developed a histological workflow for transcriptomics following general⁴⁰ and specialised
781 histological methods^{41, 42}). To preserve both RNA and tissue integrity, the primordia were placed
782 directly in a fixative and kept in solutions containing 2-mercaptoethanol throughout the whole process,
783 which was finished within 3-4 hours by cryo-embedding samples. Briefly, 1.8 ml Farmer's solution with
784 20 µl/ml 2-mercaptoethanol was used to fixate primordia for 30 minutes while the air stuck in tissues
785 was removed by applying a vacuum three times. Then, the samples were infiltrated with five
786 increasing concentrations of sucrose solutions starting with 10m/v% and finishing at 30m/v%. The
787 sucrose solutions also contained 20 µl/ml 2-mercaptoethanol and the samples were rotated for 15
788 minutes each step. Specimens were incubated three times in the 30m/v% sucrose solution, ensuring
789 complete infiltration. Samples were cryo-embedded in a cryo-mount (Optimal Cutting Temperature
790 (O.C.T.) compound) using dry ice. The embedded materials were kept at -70°C until cryo-sectioning
791 and laser-capture microdissection (LCM). Cryo-sectioning was carried out at -20°C with a cryo-
792 microtome (Leica CM1860 UV) and PEN covered & poly-L-lysine-coated slides were used to mount
793 the 10-13 µm thick sections. The cryo-mount was washed out and the sections were dried via a series
794 of alcohol solutions.

795 Laser-capture microdissection coupled RNA-sequencing

796 LCM was performed using a Zeiss PALM MicroBeam microscope and the PALM RoboSoftware v.4.6.
797 First, we screened the dehydrated histological sections by using a 10x objective (Numerical aperture
798 = 0.3; 100x magnification) to record the size dimensions of the primordia and mark the different tissue
799 types based on the shape and the structure of different regions. The 'joint contour' was used in the
800 RoboSoftware to select and mark tissue of interest, providing a narrow tissue (5 µm) attached to the
801 whole section. By doing this, we prevented the disposition of dissected tissues as the microscope
802 stage moved, as well as the tissue of interest was protected by applying the catapult laser beam only
803 on the joint tissue area. We attempted to maintain at least a 5 µm safety zone from the surrounding
804 tissues to minimize the possible contamination with other tissues. After the tissues of interest were
805 marked, laser dissection was performed with 60x objective (Numerical aperture = 0.75, 600 x
806 magnification), ensuring a thinner laser beam and thus destroying a smaller part of the tissue. Laser
807 pressure catapulting was performed using a 10x objective targeting the joint region of the dissected
808 tissue. The cap of the 0.5 ml microcentrifuge tubes (Eppendorf, DNA LoBind Tubes) prepared in
809 advance was placed as close as possible above the tissues (roughly less than a millimetre space left).
810 Laser-dissected tissues were collected into the cap of the tubes, which contained 20 µl extraction
811 buffer (PicoPure™ RNA Isolation Kit). In each tube, multiple samples were gathered from the same
812 tissue to reach a higher yield of RNA (Supplementary Table 1). The locked tubes were placed into
813 liquid nitrogen until the end of the LCM procedure. We used a heat-shock cell disruption procedure
814 similar to that used by Teichert et al., 2012⁴³, but we kept the samples in the cap of the tubes
815 throughout the whole procedure and did not use vortexing. RNA extraction was performed by using
816 the PicoPure™ RNA isolation kit following the manufacturer's protocol. We checked the quality
817 (electrochromatogram and RNA Integrity Number (RIN) score) and the quantity of RNA extracts with a
818 2100 Bioanalyzer Instrument (Agilent Technologies). We considered the RNA integrity number (RIN)
819 higher than 7.5 or RIN between 6 and 7.5 as excellent or sufficient quality for RNA sequencing,
820 respectively. Prior to the reverse transcription, cDNA amplification and sequencing, we added an
821 external RNA control (ERCC RNA Spike-In Mix, Invitrogen™) into the samples to be able to detect
822 biases that could happen through downstream procedures (Baker et al., 2005⁴⁴; Reid, 2005⁴⁵). The
823 samples were transferred to LC Sciences LLC, where the reverse transcription and cDNA
824 amplification were done by using the SmartSeq2 protocol (Picelli et al.⁴⁶, 2013, 2014⁴⁷) and paired-
825 end (150 bp) sequencing was performed on an Illumina HiSeq 4000 instrument.

826 Read processing and mapping

827 The quality of mRNA libraries was assessed by using FastQC v.0.12.0⁴⁸ and MultiQC v.1.35⁴⁹.
828 Paired-end reads were merged and quality-trimmed using BBDMap (Bushnell et al., 2014^{50,51}; Bushnell
829 et al., 2017⁵²). Reads were quality trimmed at both sides by setting the minimum Phred quality score
830 to 25 and reads shorter than 40 base pairs or with a 23-mer match to adaptor sequences were
831 discarded. Quality-filtered and trimmed reads were mapped to the chromosome level genome of
832 *Coprinopsis cinerea Amut1Bmut1 #326* strain (Hegedüs et al. 2025⁵³) by using STAR v. 020201
833 (Dobin et al., 2013⁵⁴) with the following parameters: sjdbOverhang = 50, alignIntronMax = 3,000,
834 alignMatesGapMax = 3,000, outFilterScoreMinOverLread = 0.5, outFilterMatchNminOverLread = 0.5.
835 Samtools v. 1.14 (Danecek et al., 2021⁵⁵) was used to sort and transform SAM files into BAM format.
836 We counted mapped reads at the transcript level using the summarizeOverlaps function of the
837 GenomicAlignments v.1.26.0 R package (Lawrence et al., 2013⁵⁶) by allowing to count reads if
838 transcripts were overlapping (mode="IntersectionNotEmpty"). For subsequent analyses, we analysed
839 only genes that showed expression greater than 0.5 CPM in at least half of the biological replicates for
840 a given tissue.

841 Quality check and normalisation of the transcriptome

842

843 Normalisation and transformation

844 We evaluated three aspects of read normalisation and transformation of the RNA-seq libraries: (1)
845 whether GC bias was introduced during the cDNA amplification and sequencing library construction
846 step, (2) whether global normalisation can be applied and (3) whether the regularized logarithm (rlog)
847 or the variance stabilizing transformation (VST) of the counts suits best to our data.
848 We visually checked read- and gene-level GC effects on read counts and expected similar
849 distributions of reads and gene counts as a function of GC content between different libraries if no
850 bias was introduced (Bullard et al., 2010⁵⁷). We found that GC bias could have been introduced in our
851 libraries; thus, we applied a GC content normalisation. To do that, we first extracted spliced exon
852 sequences for each transcript from the genome using GffRead v.0.11.6. program (Pertea and Pertea,
853 2020⁵⁸), then the GC content of each gene was calculated by the "GC" function of the seqinr v.4.2-8 R
854 package (Charif and Lobry, 2007⁵⁹). The GC normalisation was performed using the
855 *withinLaneNormalization* function of the EDASeq v.2.24.0 R package (Risso et al., 2011⁶⁰). Next, we
856 also assessed whether the assumption of global normalisation, that is, no global differences in the
857 distributions, is applicable to our data (Zhao et al. 2020⁶¹) by using the quantro function of the quantro
858 v.1.24.0 R package (Hicks and Irizarry 2015⁶²). Finally, we compared the "rlog" and "VST"
859 transformation methods to generate the most accurately transformed and normalised counts. To do
860 this, we assumed that a transformation should give the best correlation between the ERCC molecule
861 concentration and its expression level. We tested the performance of the rlog and vst transformations
862 along eight different normalisation methods (Median ratio, GC + median, GC + median ratio, GC + full
863 quantile, GC + qsmooth, GC + qsmooth + upper quantile, and CPM) to ensure that the normalisation
864 does not affect our decision and to assess the behaviour of alternative normalisation methods.
865 Based on the above-mentioned methods, we decided to apply GC normalisation coupled with smooth
866 quantile normalisation (qsmooth) and use the rlog transformation of the DESeq2 package 1.48.1 (Love
867 et al., 2014⁶³). To create a normalisation factor for DESeq2, we first calculated a GC-normalised
868 dataset described above and applied the qsmooth function of the qsmooth v. 1.6.0. R package (Hicks
869 et al. 2018⁶⁴). Then, we followed the standard EDASeq procedure as follows. We calculated an offset
870 by subtracting the natural logarithm of the raw data plus 0.1 from that of the normalised data plus 0.1.
871 This offset was transformed using the following formula to obtain a normalisation matrix: $\exp(-1 \times$
872 $\text{offset})$.

873

874 Outlier detection

We did a thorough outlier detection by visually assessing principal component analyses (PCA) and hierarchical clustering of the transcriptome libraries. We further evaluated the ERCC spike-in statistics, number of zero-count genes and Cook's distance statistics to filter out outlier libraries (see below). PCAs were generated using the `prcomp` function of the `stats v.3.6.3` R package (Core Team R, 2020⁶⁵) based on log2-normalised counts for all samples. We investigated three different ERCC spike-in statistics: (1) the percentage of the detected spike-in molecules (2) the Pearson correlation coefficient between the input ERCC spike-in concentration and the normalised read counts, and (3) the lowest spike-in control concentration, which could be detected in a library (i.e., the lower-level detection of molecules). The Cook's distance statistics were calculated in a negative binomial generalised linear model using the `DESeq2` package by a likelihood ratio test comparing the full model ($\sim$ tissues) and the reduced model ($\sim$ 1). By inspecting the distribution of Cook's distance values of samples, we expected more distorting genes in an outlier sample than in others. We summarised all the above-mentioned statistics to consider the exclusion of a sample from the final dataset.

Defining tissue types

Because progenitor and intermediate tissues have never been examined at this level of resolution, we wanted to check that the histology-based observations align well with the transcriptomic signatures. To achieve this, we performed hierarchical clustering coupled with bootstrap analyses, and we conducted a stepwise forward selection of how we group tissue samples into tissue types across developmental stages using a Mantel test. Hierarchical clustering was performed using the `pvcust` function of the `pvcust v. 2.2-0` R package⁶⁶, with the distance method defined as one minus the Spearman correlation coefficient, the average agglomerative method, and 1,000 bootstrap replicates. We measured the global performance of a clustering by calculating the average Silhouette value and performing a Mantel test between the transcriptome-based distance and a binary distance calculated from alternative clusterings of tissues, where the distance was zero if they belonged to the same group (tissue type) and one if they were in a different group. With Mantel statistics (Mantel r) and the average Silhouette value, we could measure how well an alternative clustering fits the transcriptomic data. The Silhouette value was calculated by using the `silhouette` function of the `cluster v.2.1.3` R package⁶⁷. For stepwise forward selection, we began with a broad clustering that best fit our data, based on the Mantel statistic and the average Silhouette value. Then we split the clusters into smaller, histology- and hierarchical-clustering-supported groups. Then, we compared whether the splitting of a cluster of tissues into smaller groups has a significantly worse fit to the transcriptome relative to the best-fit broad clustering, based on the Mantel r . We calculated the significance of the difference in Mantel r (Δr) using a permutation test, in which we permuted the clustering labels 1000 times and estimated the null distribution of Δr . If an alternative clustering is significantly worse than the best-fit grouping of tissues, we expect a significantly high Δr ($p < 0.05$). We calculated the separate contributions of developmental stages and tissues to the gene expression variance using the `fitExtractVarPartModel` function of the `variancePartition v. 3.21` package (Hoffman and Schadt 2016⁶⁸).

Detecting tissue-enhanced and repressed genes

After defining the main tissue types that are in concordance with both histological observations and transcriptomic signatures, we performed a differential-expression analysis by performing a deviation contrast across all developmental stages. In this deviation contrast, the difference between the mean of all samples (grand mean) and the mean of the tissue of interest is tested, which is similar to how tissue-specificity was calculated across human tissues (Sonawane et al. 2017²³), but we utilised the statistical framework of the `DESeq2` package. Accordingly, we built a generalised linear model (GLM) in which the deviation contrast was calculated across tissues, while accounting for the developmental stage effect ($\sim$ stage + tissue). To achieve a full rank model matrix (i.e., stage and tissue variables are not linear combinations of each other), we grouped H1-small and vegetative mycelium into one stage

and H1-big and H2 samples into another stage, which is also in concordance with the hierarchical clustering above. Differential expression analysis was performed by using a normalisation factor including GC and qsmooth normalisation described above, estimating dispersion with “parametric fit type”, defining deviation contrast using a model matrix and applying a Wald significance test coupled with independent hypothesis weighting (Ignatiadis et al. 2016⁶⁹). The shrunken log fold change (LFC), used in subsequent analyses, was calculated using the “ashr” shrinkage estimator (Stephens et al., 2016⁷⁰). To delineate tissue-enhanced and tissue-repressed genes, we set a 2.5-fold change threshold (equal to 1.32 LFC) that results in homogeneous gene sets but does not narrow the number of genes too much. This threshold was supported by inflexion points calculated along different LFC thresholds and the Shannon entropy of tissue-enhanced or repressed genes. Shannon entropy was calculated using the entropy.empirical function of the entropy v. 1.3.2 package (Hausser and Strimmer 2009⁷¹) and inflection points were detected by using the findiplist function of the inflection v.1.3.7 package (Christopoulos 2014⁷²). Finally, we also calculated stage enhanced and repressed genes while the tissue effect was controlled in the GLM model using the ~ tissue + stage formula and the same parameters as above. To verify our calculations, we also computed TAU, a widely used gene expression tissue-specificity index. We calculated the index as described in the original article (Yanai et al., 2005⁷³) using the mean expression values of the main tissue types and compared the overlap with the tissue-enhanced genes identified above.

Functional enrichment analysis

We performed a gene ontology (GO) enrichment analysis based on the tissue enhanced and repressed gene sets using a GO annotation generated by InterProScan v. 5.73-104 (Jones et al. 2014⁷⁴). The runTest and GenTable functions of the topGO v2.64.0 R package (Alexa and Rahnenführer 2026⁷⁵) were used to perform enrichment analyses with the following parameters: statistic = “fisher”, algorithm = “weight01”, nodeSize = 5. Significant enrichment of a term was detected at a weighted Fisher $P < 0.05$. Functional divergence based on GO terms was determined by using fold enrichment scores (observed genes / expected number of genes) on a log2 scale. Functional distance between all pairs of tissues was calculated as 1 minus the Pearson correlation coefficient.

Differential expression analyses of KO mutants

The RNA-seq libraries were handled in the same way as above, but Trimmomatic v.0.41 (Beier et al. 2026⁷⁶) was used to quality-trim the reads. Read counts were normalised by the DESeq2 default median ratio normalisation. Differential expression analyses were performed between the corresponding WT and KO mutant tissues (e.g., WT-cap vs KO-cap) by using the ashr shrinkage estimator as above and performing independent hypothesis weighting. Significant differential expression was detected at a two-fold change ($LFC > 1$ or $LFC < -1$) and $P < 0.05$.

Enhanced genes in other species

To evaluate the conservation of tissue-enhanced expression, we identified tissue-enhanced genes from good quality bulk transcriptome data of seven previously published Agaricomycetes species: *C. cinerea* (Krizsán et al. 2019²²), *Lentinula edodes* (Zhang et al. 2021⁷⁷), *Pleurotus ostreatus* (Jacq.) P. Kumm. (Merényi et al. 2022⁷⁸), *Lentinus tigrinus* (Bull.) Fr. (Krizsán et al. 2019²²), *Mycena kentingensis* Y.S. Shih, Chi Y. (Ke et al. 2020⁷⁹), *Armillaria ostoyae* (Romagn.) Herink (Sipos et al. 2017⁸⁰) and *Laccaria bicolor* (Maire) P.D. Orton (Ruytinx et al. 2021⁸¹). Most of these transcriptomes contain similar tissue types, such as vegetative mycelia (VM), cap (C), stipe (S), or lamellae (L), and were usually sampled at one or two stages (typically young and mature fruiting bodies) during fruiting body development. Differential expression analysis of bulk transcriptome samples was performed using a deviation contrast across developmental stages in the DESeq2 framework, comparing gene

expression in each sample to the grand mean, similarly to our approach for the ultralow-input data. Homologous tissues sampled at different developmental stages were switched off in the contrast matrix to prevent reciprocal bias. Genes were considered significantly tissue-enhanced if they showed a log2 fold change > 1.5 and an adjusted P < 0.01 toward that tissue type. To identify orthologs between these seven species, we performed a Reciprocal Best Hit (RBH) approach using the output of an MMSeqs2 (R15-6f462) (Steinegger & Söding 2018⁸²) search between CopciAB (<https://mushroomdb.brc.hu>) as a query and the seven species as subjects. Proteins that formed an RBH with CopciAB proteins across the seven species were grouped into an RBH group, whereas proteins with no RBH connection to any CopciAB protein were classified as an orphan group. These RBH groups were considered as one-to-one orthogroups and were used to quantify cross-species conservation of tissue-enhanced expression in both tissue-specific and tissue-agnostic contexts.

Proteomics

Protein extracts were prepared from harvested and flash-frozen stipe and cap samples by grinding with a mortar and pestle in liquid nitrogen, supplemented with sterile quartz sand. The powdered samples were further extracted in mortars using RIPA buffer (50 mM Tris pH 7.4, 150 mM NaCl, 1 mM DTT, 1% Triton X-100, 0.5% DOC, 0.1% SDS and 25 µM MG132) and incubated for 5 min at room temperature. Subsequently, high-purity sodium dodecyl sulfate (SDS) was added to a 5% final concentration, and the mixtures were incubated for 10 min at 90 °C, followed by centrifugation at room temperature (RT) for 5 min at 17,000 × g. The supernatant was filtered through glass wool and subjected to MS analysis. Protein extracts (three biological replicates each; 20 µL containing 50 - 80 µg total protein) from stipe and cap tissues were digested using the S-Trap micro approach (<https://protifi.com/wp-content/uploads/2025/12/PROTOCOL-S-Trap-Micro-Spin-Column-B.pdf>). Briefly, disulfide bonds were reduced with TCEP, free thiols blocked with MMTS, and proteins digested using sequencing-grade trypsin (3 µg/sample; incubation for 2h at 47 °C). Dried peptides were resuspended in 0.1% formic acid (FA) in water. Aliquots equivalent to ~1 µg of starting protein was loaded onto Evotips and analyzed on an Evosep One system coupled to a FAIMS Pro-equipped Orbitrap Fusion Lumos Tribrid mass spectrometer. Peptides were separated using the standard 88-min Evosep method. Data-independent acquisition (DIA) was performed with a single FAIMS compensation voltage of -45 V. MS1 and MS2 scans were acquired in the Orbitrap at resolutions of 60000 and 15000, respectively. Raw data were processed in DIA-NN v2.5.1 Academia (Demichev et al., 2020⁸³) using the library-free mode against a database of 15196 *Coprinopsis cinerea* sequences, supplemented with a predicted in silico library. Search parameters included: trypsin specificity with maximum of 1 missed cleavage; fixed modification: methylthio (C). Variable modifications (allowed up to 2 per peptide): Met oxidation, protein N-terminal acetylation/Met excision, and N-terminal Gln-to-pyroGlu conversion. Mass tolerances were set to automatic inference. Protein assembly and FDR filtering were restricted to ≤1% at both precursor and protein levels. Differential expression analysis was performed using the limma R package (Ritchie et al., 2015⁸⁴), applying thresholds of log2 fold change ≥ 1 and Benjamini-Hochberg adjusted P ≤ 0.05.

Reverse genetics

We used a ribonucleoprotein-mediated approach for CRISPR/Cas9 modification of target genes⁸⁵, by using a selectable marker, following the strategy of Wu et al 2026²⁹. Briefly, we employed the plasmid including the pUC19 backbone and the selection marker para-aminobenzoic acid (PABA) fragment amplified from the pMA412 vector⁸⁶. In complementation experiments the hygromycin B phosphotransferase gene (*hyg*) was used as a selection marker, with the promoter and terminator sequences of the tubulin-encoding gene in *C. cinerea*.

We investigated whether altering tissue specific expression patterns of pheromones via promoter exchange results in developmental alterations. For promoter exchange the donor cassette was designed to replace the native promoter of the target gene with the promoter of another pheromone-encoding gene through homology-directed repair (HDR). The positive selection marker gene and the target promoter insert were flanked by the upstream homologous arm of the targeted gene on one side, and the open reading frame (ORF) followed by the downstream homologous arms of the target gene on the other side. We replaced the native promoter of CopciAB_1001716, which is normally highly expressed in the cap, with the CopciAB_492306 promoter to reduce its expression across tissues, particularly in the cap. Conversely, we placed CopciAB_492306 under the control of the strong CopciAB_1001716 promoter to increase its expression, with the aim of achieving high expression levels in the cap. For CopciAB_1000575, which shows minimal to non-detectable basal expression under native conditions, the CopciAB_1001716 promoter was used to induce expression across tissues, particularly in the cap, gill, and stipe. Finally, for CopciAB_1002222, which has its lowest native expression levels in the cap and gill, we substituted its native promoter with the CopciAB_1002223 promoter to increase expression across tissues, particularly in the cap and gill.

Comparative genomics

Phylostratigraphy

To evaluate whether tissue-enhanced and tissue-repressed genes are enriched in specific evolutionary ages (phylostratum)—potentially reflecting the emergence of particular tissue types during fungal evolution—we applied a phylostratigraphic approach. For this analysis, we used a previously published dataset containing representatives of all Holomycota phyla from the Opisthokont ancestor, with a special focus on Dikarya. This dataset of 655 genomes (G655⁸⁷) provides a solid basis for high-resolution phylostratigraphic analysis.

Phylostrata was estimated using a reciprocal best-hit (RBH) approach. Specifically, we performed homology searches between the CopciAB genome and all genomes in the G655 dataset, in both directions. Sequence similarity searches were conducted with MMseqs2 (R15-6f462) using three iterations at a sensitivity of 5.7, retrieving up to 5,000 target sequences (E-value < 0.0001) and requiring a minimum coverage of 20%. In order to exclude potential paralogs and thus obtain the gene age corresponding to the last duplication, we followed the logic of OrthoFinder⁸⁸ and excluded hits with an e-value higher than that of the second-best hit of the *CopciAB* query for each query protein. Further, to exclude potential spurious hits that could overestimate the age of a gene, we required at least 50% presence or at least three hits in each clade that could modify the inferred age of the gene (along the backbone from the root to *CopciAB*). Among the retained clades, the most ancient one determined the age of the gene. To assess the robustness of the results, we generated six increasingly stringent datasets for gene age estimation by applying bidirectional coverage filters of 20%, 35%, 50%, 65%, 80% and 95% to the homology search results. From these, we retained only RBH connections, defined as cases where gene A from species 1 has its best match in species 2 as gene B, and gene B's best match in species 1 is also gene A. For each CopciAB gene, we then calculated its “birth node” in the G655 species tree, defined as the most recent common ancestor of all species with an RBH connection to that gene. Genes with no RBH connections to any other genome were considered strain-specific and were assigned the most recent age. Each node from the root node to the tip of *C. cinerea* corresponds to a phylostratum from PS1 to PS35, respectively. Over-representation analysis was performed using the ora function of the mulea package⁸⁹. Input gene sets, such as tissue-enhanced and tissue-repressed genes from different tissues, were tested against the background distribution of gene ages. Significance was assessed with empirical false discovery rate (eFDR) correction based on 10,000 permutations.

Transcriptome age index

We calculated the transcriptome age index (TAI) from phylostrata assignments made above (bidirectional coverage filter = 40%) and normalised log-transformed count data using the *as_BulkPhyloExpressionSet* function of the myTAI v.2.2.0.9006 R package⁹⁰. Flat line, reductive hourglass, reverse hourglass, early or late conservation, and pairwise statistical tests were performed using the *stat_flatline_test*, *stat_reductive_hourglass_test*, *stat_reverse_hourglass_test*, *stat_early_conservation_test*, and *stat_late_conservation_test* and BulkPhyloExpressionSet functions, respectively. We defined early, mid and late stages as Vegetative mycelium-Primary hyphalknot (VM-H1), Secondary hyphal knot-Stage 0 primordium (H2-P0) and Stage 1 - Stage 2 primordia (P1-P2). In order to identify the minimum number of genes able to generate a non-significant TAI profile of our observed TAI pattern, we employed a multi-objective optimization procedure which optimized for candidate gene selection using the following two objectives: 1) minimise the number of genes contained in the selection set and 2) increase the p-value of the flat-line test implemented in the myTAI R package to up to a p-value >0.75 ("destroyed TAI pattern"). We employed a Genetic Algorithm to achieve this multi-objective optimization outcome. Since this optimization procedure is non-deterministic, we performed 20 independent runs and only retained the intersection set over these runs.

Phylogenomics

Candidate ICMT and GPCR gene sequences as well as homologs of the CAAX protein maturation pathway were collected based on Interpro domains. Sequences were aligned using the L-INS-i algorithm in MAFFT v7.526⁹¹. Three distant outgroup sequences, Rhimi59_2_1898070, sp_O60725 and sp_Q9EQK7 (Uniprot IDs), poorly aligned sequences and ragged termini were removed using Trim-AI (-gt 40). Gene tree inference was performed by RAxML v8.2.13⁹² in MPI mode under the PROTGAMMAWAG evolutionary model, with branch support evaluated with 1000 bootstrap replicates. The gene tree was then reconciled with the species tree using Notung v2.9.1.5 with an edge-weight threshold of 0.9 to rearrange its topology and analyze the loss and gain events throughout evolutionary history⁹³. Inferred gene copy numbers were plotted as bubbles on each node of the species tree. The above analysis was repeated with *Sphaerobolus stellatus*, a species with a highly fragmented genome removed, to address the robustness of the results.

Prediction of novel pheromone-precursors

Pheromone prediction using binary supervised classifiers

We trained multiple supervised binary classifiers to identify pheromone protein sequences within the genomes of 186 fungal species. We assembled a list of known pheromone sequences consisting of 628 entries pooled over different Agaricomycetes. To avoid data leakage, pheromone protein sequences originating from the genome under evaluation were excluded from the training set. The number of predicted short ORFs over the 186 species surpassed 145 million entries, with the average and median number of ORFs per species being 780,837.05 and 603,676.5, respectively.

A specificity of the genome sequences was that – apart from the previously mentioned 628 ones – they were unlabeled. As such, we faced a classification problem, where only a tiny fraction of the total instances belonged to the positive class. To obtain classification instances of the negative (i.e., non-pheromone) class, we treated those sequences not ending with the CAAX motif as negative instances. This resulted in a distant supervision setting, with a high imbalance towards the negative class instances. Based on the above created datasets, we trained a series of binary classification models and an ensemble classifier that aggregated the outputs of the individual classifiers. In total, we trained 49 independent models and their ensemble. Our individual models ranged from traditional statistics-based approaches to k-nearest neighbor classifiers and more sophisticated neural network-based ones. In the following, we provide details on the 49 classifiers that we trained for each species.

The traditional models we trained were multinomial naive Bayes classifiers using a vector

space representation of the genomic sequences. We trained six different naive Bayes classifier models with protein n-grams serving as the basis of modeling, with $1 \leq n \leq 6$.

We also exploited modern transformer-based protein language models by using facebook/esm2_t6_8M_UR50D as our backbone model⁹⁴. We performed fine-tuning of this 6-layered, 8 million parameter transformer model for each species. Our hyperparameters for fine-tuning followed common practices, i.e., a learning rate of $2e-5$, a batch size of 32 and a regularization of 0.01. We used the AdamW optimizer for updating the model and conducted fine-tuning over three epochs. We conducted preliminary experiments for assessing the sensitivity of the hyperparameter choices, and we found that the predictions of the models trained by differently chosen hyperparameters remained generally stable, to which end we used the same hyperparameters for all the species.

We also used our transformer backbone model as a feature extractor. In this case, we did not train any of the weights of the backbone model. Instead, we trained a simple logistic regression classifier based on the internal hidden representations produced by the intact pre-trained model. The backbone model assigns a special [CLS] token for each sequence, which we used as an input to the logistic regression classifier. As our backbone model has six layers, we trained six different such models for each species, with each model relying on the hidden representations of the backbone model originating from a different layer.

Besides the logistic regression classifier, we also trained a 1-nearest neighbor (1-NN) classifier based on the hidden representations determined by our backbone model for the individual sequences. Whenever the nearest training data point of a candidate sequence – measured via the cosine distance of their corresponding hidden representation – was labeled as a pheromone, we predicted it to also belong to the positive class. A 1-NN classifier was trained based on all the six layers of the backbone model, resulting in 6 additional models.

We earlier showed that sparse coding can transform the continuous hidden representations of transformer-based language models into a set of interpretable quasi-symbolic features⁹⁵. Following this approach, we applied sparse coding to the hidden representations produced by our backbone model and used the resulting sparse vectors as input features for logistic regression and naive Bayes classifiers. When obtaining the sparse vectorial feature representations, we transformed the originally 320-dimensional dense internal vectors into 1024-dimensional sparse vectors. The inventory of semantic atoms was learned through an unsupervised dictionary learning procedure that optimized the reconstruction of the dense hidden states as sparse linear combinations of 1024 dictionary atoms. We trained six groups of such models, i.e., one for each of the internal layers of the backbone model. Within each group, we had 5 different variants of the classifiers that relied on the sparse coded representations, yielding a total of 30 models per species. The 5 variants differed on whether the sparse coefficients were binarized or not, whether the underlying classifier model was a naive Bayes or a logistic regression classifier, and whether the presence (non-zerosness) of the individual quasi-symbolic features per input sequences were considered as separate features.

Source code for the experiments above is publicly accessible at <https://github.com/szegedai/sysbiol/>.

Pheromone prediction using custom HMMs

To construct a custom HMM for pheromone precursors, high-confidence seed sequences were first obtained by scanning 5 kb genomic regions flanking mating GPCR loci. STE3 receptor homologs were detected in predicted proteomes using the PF02076 STE3 HMM with hmmsearch ($E \leq 1e-5$). STE3 hits were mapped back to genome using coordinates of the corresponding GFF3 annotations, and 5 kb flanking genomic regions around each STE3 locus were extracted. These flanking regions were scanned for CAAX-anchored small ORFs, as follows. ORFs between 20 and 100 amino acids were retained if they terminated immediately before a stop codon and contained a C-terminal CAAX-like motif. The motif was defined as a cysteine followed by two aliphatic residues and one terminal residue, with A/G/I/L/V/M/T/P allowed at the two aliphatic positions. Candidate ORFs sharing the same genomic stop coordinate and strand were collapsed.

Candidate seed peptides were filtered using two independent criteria. First, candidates were searched against the fungal mating-type pheromone HMM PF08015 using hmmsearch⁹⁶ ($E \leq 1e-2$).

1163 Second, candidates were evaluated with a rule-based pheromone precursor filter requiring a terminal
1164 CAAX motif, an upstream processing motif, sufficient mature peptide hydrophobicity, and an N-
1165 terminal acidic residue. Candidates passing both the PF08015 HMM filter and the rule-based filter
1166 were retained as cross-validated seed sequences. In total, 329 raw seed sequences from 85 species
1167 were obtained. Exact duplicates were removed, followed by clustering at 95% identity, resulting in 194
1168 nonredundant training seed sequences. These sequences were aligned using MAFFT L-INS-i (--
1169 localpair --maxiterate 1000 --thread 1). The alignment was trimmed using TrimAl, and multiple gap-
1170 threshold settings were evaluated. The model generated from the -gt 0.6 trimmed alignment was
1171 selected for downstream analyses because it retained sufficient alignment length while reducing gap-
1172 rich columns. The final custom profile HMM was built using HMMER⁹⁶ hmmbuild --amino and indexed
1173 using hmmcompress.

1174 The new custom HMM was used to identify candidate pheromone precursors from CAAX-
1175 anchored ORF peptide FASTA files obtained above. Candidate ORF peptide sets were searched with
1176 HMMER hmmssearch. For each search, hit identifiers and the corresponding peptide sequences were
1177 written to separate output files for downstream comparison and manual inspection.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryTable.2.xlsx](#)
- [SupplementaryTable.6.xlsx](#)
- [SupplementaryTable.5.xlsx](#)
- [SupplementaryTable.11.xlsx](#)
- [SupplementaryTable.15.xlsx](#)
- [SupplementaryTable.9.xlsx](#)
- [SupplementaryTable.7.xlsx](#)
- [SupplementaryTable.8.xlsx](#)
- [SupplementaryTable.16.xlsx](#)
- [SupplementaryTable.10.xlsx](#)
- [SupplementaryTable.4.xlsx](#)
- [SupplementaryTable.17.xlsx](#)
- [SupplementaryTable.13.xlsx](#)
- [SupplementaryTable.12.xlsx](#)
- [SupplementaryData.1.SupplDatasetMUT.pdf](#)
- [Supplementarymaterial.pdf](#)
- [SupplementaryTable.3.xlsx](#)
- [SupplementaryTable.1.xlsx](#)
- [SupplementaryTable.14.xlsx](#)