

Vertical and horizontal gene transfer shaped plant colonization and biomass degradation in the fungal genus *Armillaria*

Received: 10 November 2022

Accepted: 11 July 2023

Published online: 7 August 2023

 Check for updates

Neha Sahu^{1,2}, Boris Indic³, Johanna Wong-Bajracharya^{4,5}, Zsolt Merényi¹, Huei-Mien Ke^{6,21}, Steven Ahrendt⁷, Tori-Lee Monk⁴, Sándor Kocsubé^{8,9}, Elodie Drula^{10,11}, Anna Lipzen⁷, Balázs Bálint¹, Bernard Henrissat^{12,13}, Bill Andreopoulos⁷, Francis M. Martin¹⁴, Christoffer Bugge Harder^{15,16}, Daniel Rigling¹⁷, Kathryn L. Ford¹⁸, Gary D. Foster¹⁸, Jasmyn Pangilinan⁷, Alexie Papanicolaou⁴, Kerrie Barry⁷, Kurt LaButti⁷, Máté Virágh¹, Maxim Koriabine⁷, Mi Yan⁷, Robert Riley⁷, Simang Champramary^{2,3}, Krista L. Plett⁵, Igor V. Grigoriev^{7,19}, Isheng Jason Tsai⁶, Jason Slot²⁰, György Sipos³, Jonathan Plett⁴ & László G. Nagy¹✉

The fungal genus *Armillaria* contains necrotrophic pathogens and some of the largest terrestrial organisms that cause tremendous losses in diverse ecosystems, yet how they evolved pathogenicity in a clade of dominantly non-pathogenic wood degraders remains elusive. Here we show that *Armillaria* species, in addition to gene duplications and de novo gene origins, acquired at least 1,025 genes via 124 horizontal gene transfer events, primarily from Ascomycota. Horizontal gene transfer might have affected plant biomass degrading and virulence abilities of *Armillaria*, and provides an explanation for their unusual, soft rot-like wood decay strategy. Combined multi-species expression data revealed extensive regulation of horizontally acquired and wood-decay related genes, putative virulence factors and two novel conserved pathogenicity-induced small secreted proteins, which induced necrosis in planta. Overall, this study details how evolution knitted together horizontally and vertically inherited genes in complex adaptive traits of plant biomass degradation and pathogenicity in important fungal pathogens.

Plant pathogenic fungi cause serious economic losses worldwide in a wide variety of plant species, including forest trees. Among tree pathogens, the genus *Armillaria* (Basidiomycota, Agaricales and Physalacriaceae) stands out as one of the most important in temperate systems, responsible for great losses in both natural and planted stands of woody plants^{1–3}. At the genus level, they are known to cause ‘*Armillaria* root-rot disease’^{3,4}. The most common pathogenic species *A. mellea sensu lato* has been reported to infect >500 plant species¹ and is solely responsible for up to 40% annual loss of vinegrape in

California⁵. The gymnosperm-specific *A. ostoyae* is responsible for considerable losses in conifer forests¹.

Armillaria species evolved a range of features exceptional or rare among fungi, which have conceivably all emerged in its most recent common ancestor (MRCA). These include a very low mutation rate, extreme longevity and immense colony sizes (>2,500 years, >900 ha (refs. 1–3)), diploidy, bioluminescence, specialized underground structures known as rhizomorphs¹ and potential to facilitate atmospheric N₂ fixation⁶. Perhaps the most economically important aspect of this

A full list of affiliations appears at the end of the paper. ✉e-mail: lnagy@fungenomelab.com

genus is the ability to infect and kill woody plants^{1,3}. Most *Armillaria* species are broad-host-range necrotrophs^{1,3}. After infection through the roots, they colonize and kill the cambium of mostly weakened trees, causing the death of the plant and enabling the fungus to transition to its necrotrophic phase^{1,7}. Despite a well-documented epidemiology and aetiology^{2,7,8}, molecular aspects of the infection process are poorly known. Recent studies and genome sequencing efforts highlighted certain secondary metabolites, plant cell wall-degrading enzymes (PCWDEs), chitin-binding proteins and expanded protein-coding repertoires enriched in putative pathogenicity-related genes, among others^{1,6,9}. It is likely that these previous findings only cover a fraction of the virulence genes of *Armillaria*, leaving much of its pathogenic arsenal yet to be characterized.

As a result, it is unknown how infection models developed on the basis of better-studied necrotrophic fungi (for example, *Sclerotinia sclerotiorum*¹⁰, or *Colletotrichum* spp.¹¹) are applicable for *Armillaria*, if at all, and what traits the latter evolved for plant infection. In other necrotrophs, broad roles of tissue acidification, tolerance towards and detoxification of plant secondary metabolites and reactive oxygen bursts¹² as well as secretion of diverse PCWDEs^{11,13} and effectors¹⁴ have been established as key assets of infection. However, given that most relatives of *Armillaria* are non-pathogenic, it can be inferred that it evolved necrotrophy independently, which might have resulted in unique infection mechanisms, our understanding of which remains limited.

In this Article, we sequenced eight new genomes and report transcriptome data from new in planta and in vitro pathosystems, which enabled us to explore genome evolution and key aspects of the necrotrophic lifestyle of *Armillaria* spp. We infer that gene duplications, genus-specific gene families and horizontal gene transfer (HGT) have shaped the genomic toolkit available for plant infection and biomass degradation. RNA sequencing (RNA-seq) data from six experiments and four *Armillaria* species, including new in planta time series and fresh tree stem invasion experiments, allowed us to decipher gene expression patterns specific for these processes in virulent and non-virulent strains. Experimental validation of predicted pathogenicity-induced small secreted proteins (PiSSPs) revealed potential conserved virulence factors in *Armillaria*. Overall, our phylogenomic and gene expression studies elucidate how vertically and horizontally acquired genes became integrated into complex adaptive traits of a globally important fungal group.

New *Armillaria* genomes

We report the high-quality annotated de novo genomes of eight *Armillaria* species (Fig. 1 and Supplementary Table 1). The new genomes were assembled to 33–864 scaffolds comprising 40–79 Mbp haploid size, with 12,228–19,984 predicted gene models and benchmarking universal single-copy orthologue (BUSCO (fungi)) completeness of 97.7–99.7%.

We sampled all major clades recognized currently¹⁵, including the Northern Hemisphere, the Australasian/Southern American, African and melleoid clades (Fig. 1). We also included two species from subgenus *Desarmillaria* (*A. tabescens* and *A. ectypa*), of which the moss-associated *A. ectypa* had the smallest genome among *Armillaria* spp., as well as *Guyanagaster necrorhizus*⁶, which is the sister genus of *Armillaria*¹⁵.

Genetic innovations in *Armillaria* clade

To analyse the genomic innovations associated with the emergence of *Armillaria*, we used genomes of 15 *Armillaria* species and 5 outgroups from the Physalacriaceae. We combined these with other Agaricales exhibiting a range of lifestyles, resulting in a dataset of 66 species (Supplementary Table 1).

Reconstruction of genome-wide gene gain/loss patterns revealed genome expansion in *Armillaria*. We estimated 9,929 ancestral genes

at the root node, suggesting an early origin for most genes (Extended Data Fig. 1a). The *Armillaria* genus showed a net genome expansion with 18,662 protein-coding genes inferred for the MRCA of *Armillaria* (2,913 duplications and 189 losses), as opposed to 15,938 for that of *Armillaria* and *Guyanagaster* and 18,155 for the MRCA of *Armillaria* and *Hymenopellis* (Fig. 1 and Extended Data Fig. 1). The MRCA of the Northern Hemisphere clade was inferred to have further expanded to 23,756 genes. These data indicate that the large protein-coding repertoires of *Armillaria* spp. can be explained by genus-specific gene duplications, as suggested before⁹.

The 2,913 duplications in the mrca of *Armillaria* happened in a total of 1,473 orthogroups (OGs). Gene ontology (GO) and InterPro enrichment analyses showed significant over-representation of 55 molecular functions, 18 biological processes and 3 cellular component terms ($P < 0.05$) (Extended Data Fig. 2) and 733 InterPro terms (Supplementary Table 2). These included functions related to plant biomass utilization, such as pectin degradation (pectate lyases, pectin lyases, esterases and GH28), cellulose binding, and putative extracellular and aromatic compound breakdown (for example, intradiol ring-cleavage dioxygenase and multi-copper oxidases) (Supplementary Table 2 and Extended Data Fig. 2). Duplicated genes were also enriched in putative pathogenesis-related gene families, including genes encoding deuterolysins, aspartic peptidases, chitin deacetylases and SCP-like Golgi-associated pathogenesis-related proteins. Cerato-platanins and LysM domains, which were reported to assist infection in pathogenic fungi^{16,17}, were also enriched in *Armillaria*.

Small secreted proteins (SSPs), which include cysteine-rich proteins involved in, among others, host colonization¹⁸, were found in 270–507 copies in *Armillaria*, with *A. fuscipes* having the fewest and *A. luteobubalina* having the most. Of these, 45–57% had no known functional domains, which we hereafter refer to as unannotated SSPs. Typically, this latter class of SSPs are called candidate-secreted effector proteins in pathogenic fungi^{11,19}.

Gene families that arose within and are conserved in most *Armillaria* spp. may be particularly relevant for explaining *Armillaria*-specific innovations. We found 212 such families (Supplementary Table 2 and Supplementary Fig. 1); of which 116 consisted of proteins with no known functional annotations. The remaining OGs were dominated by F-box domains, Leucine-rich repeats, Cytochrome P450s, Zinc-finger C2H2 type transcription factors, protein kinases and other fast-evolving gene families.

Armillaria spp. are among the few bioluminescent fungi. Ke et al. reported the luciferase gene cluster in *Armillaria*, comprising five genes²⁰. We found the luciferase cluster to be conserved and highly syntenic in *Armillaria* genomes (Supplementary Fig. 2); however, the cluster was missing in other Physalacriaceae, suggesting it was lost in those, or gained in the common ancestor of *Armillaria* and *Guyanagaster*.

Plant biomass degradation by *Armillaria*

Armillaria species are reported to be facultative necrotrophs that first kill the host, then utilize its biomass during the saprotrophic phase. Using the new genomes, we analysed the wood decay strategy of *Armillaria*, relative to other fungi based on their PCWDE gene repertoires. Similar to white rot fungi and necrotrophs²¹, *Armillaria* species possess the complete enzymatic repertoire for degrading woody plant biomass (Fig. 1). We generated phylogenetic principal components analysis (PCAs) for *Armillaria* species, other Physalacriaceae, white rot and litter decomposer species (Supplementary Table 1) based on PCWDE copy numbers acting on cellulose, hemicellulose, pectin and lignin. Cellulases and pectinases clearly separated *Armillaria* spp. and other Physalacriaceae from white rot and litter decomposer fungi (Fig. 2 and Extended Data Fig. 3), whereas PCAs of hemicellulase and ligninase grouped them together (Fig. 2 and Supplementary Figs. 3–5). Cellulase loading factors indicate that this separation was mainly driven by expansins, the AA16, AA8, and AA3_1, GH1 and GH45 families.

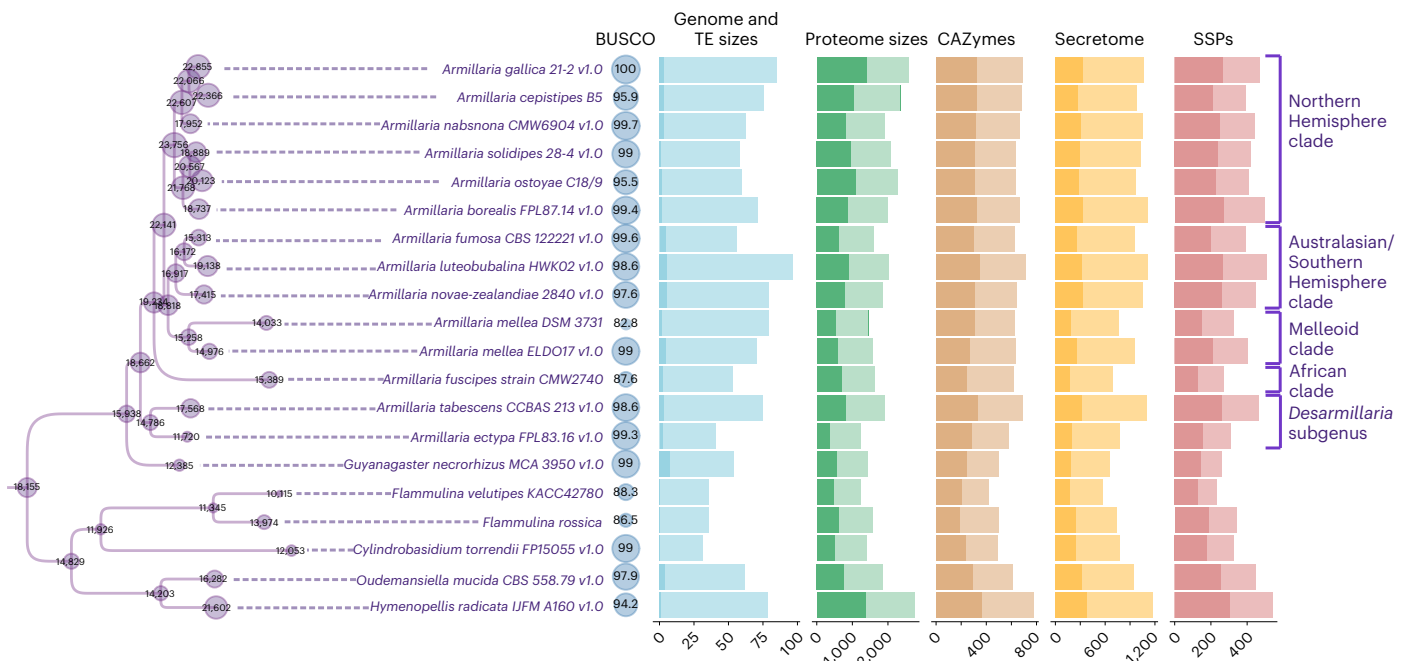


Fig. 1 | Genome statistics and reconstruction of ancestral genome sizes for 15 *Armillaria* species and 5 Physalacriaceae outgroups. The numbers at nodes represent ancestral proteome sizes in the Physalacriaceae tree. Purple circles correspond to sizes for each node in the tree (for gene gains and losses at each node, and for the complete species tree, see Extended Data Fig. 1a). Blue circles represent BUSCO scores for each species. For genome sizes, the darker region shows TE sizes (Mbp) from the primary scaffolds, and the lighter colour shows

the genome assembly sizes in Mbp (for TE categories, see Extended Data Fig. 1b). For proteome sizes, the darker colour shows proteins with no known functional domains (unannotated proteins). For CAZymes, the darker colour shows PCWDEs and the lighter colour shows other CAZymes. For secretomes and SSPs, the darker colour shows unannotated proteins and the lighter colour shows proteins with known functional domains (for general genome statistics of all 66 species in Dataset 1, see Supplementary Table 1). Clade names are based on Koch et al.^{15,41}.

In line with loading factors, *Armillaria* and other Physalacriaceae have more AA3_1, GH1 and GH45 genes than white rot and litter decomposer fungi (Supplementary Fig. 6). On the other hand, they were depleted in CBM1 and GH5_5 with an average of 12 and 1.5 genes, respectively, while white rotters and litter decomposers have 20–60 CBM1 and 4–10 GH5_5 genes. Notably, a few brown rot and ectomycorrhizal fungi had more gene copies of CBM1 and GH5_5 than *Armillaria* (Supplementary Fig. 6). The GH44 family, which is specific to Basidiomycota²², is missing in the Physalacriaceae and *Armillaria*, indicating that in specific cases, gene losses also drove trophic mode evolution of these fungi towards an Ascomycota-like lifestyle (see below).

In the pectinase PCA, the highest loading family was CBM67, which is enriched in *Armillaria* species but absent in most white rot species in our dataset (Supplementary Fig. 7). CBM67s are frequently associated with GH78 and PL1 proteins and are putatively binding rhamnose, a pectin component²³. The PL1 family is present in all Physalacriaceae species, but is depleted or missing in many white rotters and litter decomposers (Supplementary Fig. 7). Other pectin-acting families, such as GH28, GH53, GH88, CE8 and CE12, were present in higher numbers in *Armillaria* and other Physalacriaceae than in white rot and litter decomposer fungi, whereas PL1_7, PL26, PL3_2 and PL4_1 were abundant in *Armillaria* but absent in most other decomposer species.

These analyses portray *Armillaria* and the Physalacriaceae as versatile wood decayers that are distinct from white rotters, despite previous classifications as such^{3,9,24}. This is consistent with microscopy, chemical and transcriptomic analyses^{25,26}, which indicated that their decay is similar to soft rot^{27,28}, a decay mode known only in the Ascomycota. This apparent discrepancy prompted us to systematically look for similarities between the Physalacriaceae and the Ascomycota. We found 16 PCWDE OGs that were significantly over-represented in both groups with respect to white rotters and litter decomposers (Benjamini–Hochberg-corrected *P* value > 0.05, Fisher's exact test,

Fig. 2 and Supplementary Table 3). These included several of the high-loading families from the PCAs, as well as other PCWDEs acting on cellulose (AA3_1, AA8 and CBM1), pectin (PL3_2, PL1_7 and PL9_3), cellulose/chitin (AA16) and hemicellulose (GH31, GH43, GH93 and CE4). These families could either be the result of co-expansion in both the Ascomycota and the Physalacriaceae or represent HGT events. Blast searches with *Armillaria* GH28 genes suggested the latter scenario to be more likely, which led us to systematically evaluate the role of HGTs.

Widespread horizontal transfer of genes from Ascomycota

To identify horizontally transferred (HT) genes, we first screened candidates based on the Alien index (AI)²⁹ across a broad set of 942 species of fungi, plants and bacteria (Supplementary Table 1, Dataset 3 and Supplementary Table 4). Then, we validated each candidate HT gene by phylogenetic analyses ($\geq 70\%$ bootstrap in maximum likelihood gene trees) and in-depth similarity searches (UniRef100 database) (Methods; gene trees available at Figshare https://figshare.com/articles/dataset/Gene_trees/22730534, Supplementary Table 1, Dataset 3). Overall we recovered 124 strongly supported horizontal transfer events (Fig. 3a) into Physalacriaceae, corresponding to 1,387 individual genes, in 110 OGs. We identified 37–107 HT genes per species in Physalacriaceae, with *Hymenopellis radicata* having the most and *Cylindrobasidium torrendii* the fewest. Among *Armillaria*, *A. tabescens* had the highest estimated number of HT genes (83), and *G. necrorhizus* the lowest (51) (Fig. 3a and Supplementary Table 4).

Multiple internal nodes of the Physalacriaceae tree were identified as putative recipients. One hundred four HGT events were confidently associated with Ascomycota as donors, in particular the Sordariomycetes and Dothideomycetes (34 and 31 events, respectively). We found that ~40% of HT genes have undergone duplications, indicating that they were probably integrated into the life history of Physalacriaceae.

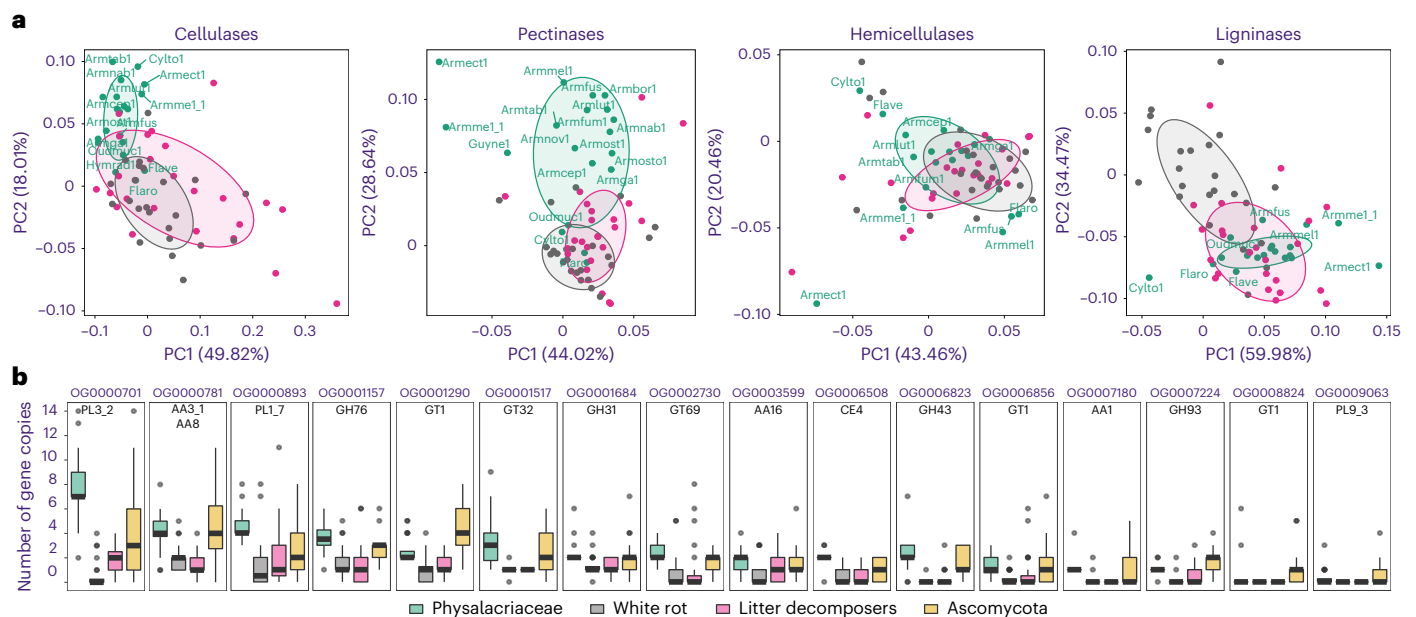


Fig. 2 | Plant biomass degradation related genes in *Armillaria*. **a**, Phylogenetic PCAs for PCWDE gene families. Species abbreviations are shown only for Physalacriaceae species (all species names and PCA loadings are given in Extended Data Fig. 3 and Supplementary Table 3). **b**, Box plot of copy numbers of 16 CAZy OGs co-enriched in Physalacriaceae ($n = 20$) and in Ascomycota ($n = 21$) with respect to white rot ($n = 23$) and litter decomposer fungi ($n = 24$).

The box plots show the median and interquartile range, with the upper whiskers extending to the largest value from the 75th percentile, and lower whiskers extending to the smallest value from the 25th percentile. Scale limits for the box plot were set to 14, losing one sample point (Conioc1 in OG0000781 with 18 genes). Lifestyles of the species used are denoted by colour.

Expression levels of HT genes support their functionality (Fig. 3c and Extended Data Fig. 4).

Among the phylogenetically validated HT genes, we found 164 Carbohydrate Active enzymes (CAZymes) (Supplementary Table 4), of which 117 belonged to families co-enriched in Ascomycota and Physalacriaceae (see above), suggesting that the co-enrichment signal was probably created by HGT. In addition to CAZymes, HGT affected intradiol ring-cleavage dioxygenases, CAP domain proteins and Pyr1-like SCP domains, as well as cytochrome P450, peptidases, transporters and transcription factors (Supplementary Table 4).

Taken together, we find that horizontal transfer affected several gene families associated with wood decay (for example, AA3_1, GH43 and PL3_2), as well as plant–fungal interactions (for example, CAP domain proteins and peptidases), which suggests that it might have shaped the plant biomass degrading and pathogenic abilities of the genus. Given that Physalacriaceae species have been reported to cause soft-rot-like decay, which is classically restricted to the Ascomycota³⁰, we hypothesize that HGT contributed to the evolution of the plant biomass degrading ability of *Armillaria*. We speculate that the large number of putative HGT events from Sordario- and Dothideomycetes might stem from the extensive contact of *Armillaria* ssp. with plant pathogens in these classes and/or their peculiar niche and longevity. Their rhizomorphs can reach several metres in length and can contact numerous other soil microbes, perhaps providing time windows for gene exchange. Additionally, other idiosyncrasies of *Armillaria*, such as diploidy, could also be factors promoting HGT.

Gene expression profile in plant colonization and wood decay

Armillaria enters the host through the roots, colonizes and kills the cambium, leading to death of the plant and the onset of the necrotrophic phase¹. Molecular aspects associated with this process are hardly known, with most information available on the wood-decaying phase^{9,26}. To obtain a molecular perspective on these strategies, we produced new RNA-seq data for two in vitro pathosystems and re-analysed

data from two published studies on wood decay as well as rhizomorph and fruiting body development^{9,26}. New data were generated from an in planta time-series experiment of *A. luteobubalina* infecting *Eucalyptus grandis* seedlings (Extended Data Fig. 5a) and for in vitro fresh tree stem invasion assays with highly and less virulent isolates of both *A. ostoyae* and *A. borealis* (Extended Data Fig. 5b), emulating the cambium-killing phase of the fungus.

We independently analysed differentially expressed gene (DEG) lists from each experiment and calculated DEG enrichment ratios in each of the 24 gene groups we defined (Methods). These ratios reflect how enriched DEGs are in a given gene group in a given experiment and are shown as a heat map on Fig. 4 (and Extended Data Fig. 6). The different experiments showed distinct enrichment patterns. Cellulose-, hemicellulose-, pectin- and lignin-related PCWDE genes were enriched in stem invasion and wood-decay experiments, as expected^{9,31}. On fresh stems, pectinase-encoding genes were most dominant, possibly enabling the fungus to break carbohydrate–lignin bonds and spread between the bark and the sapwood. At the same time, PCWDEs were depleted during the infection phase in the in planta time series experiment, indicating that *A. luteobubalina* did not induce these genes during the infection phase. This is consistent with most necrotrophic fungi expressing a limited set of PCWDEs for plant penetration and a larger battery of enzymes during necrotrophy^{11,13}. Other virulence-related genes include cerato-platanins¹⁶, which were enriched at 24 h and 48 h in *A. luteobubalina* and in stem invasion by *A. borealis*. HT genes showed an enrichment in wood-decay and stem invasion experiments. Bioluminescence genes were enriched in root-invading mycelium and rhizomorph (*A. ostoyae*) samples. SSPs, expanded, novel and stress-related genes were enriched in various fruiting body developmental stages (Extended Data Fig. 6). Genes related to oxidative stress were upregulated in various stages of plant infection and in stem invasion assays (Fig. 4 and Extended Data Fig. 7). Specifically, genes encoding superoxide dismutases, catalases and members of the glutathione system and ergothioneine pathway were enriched among DEGs in the in planta invasion assays, whereas genes

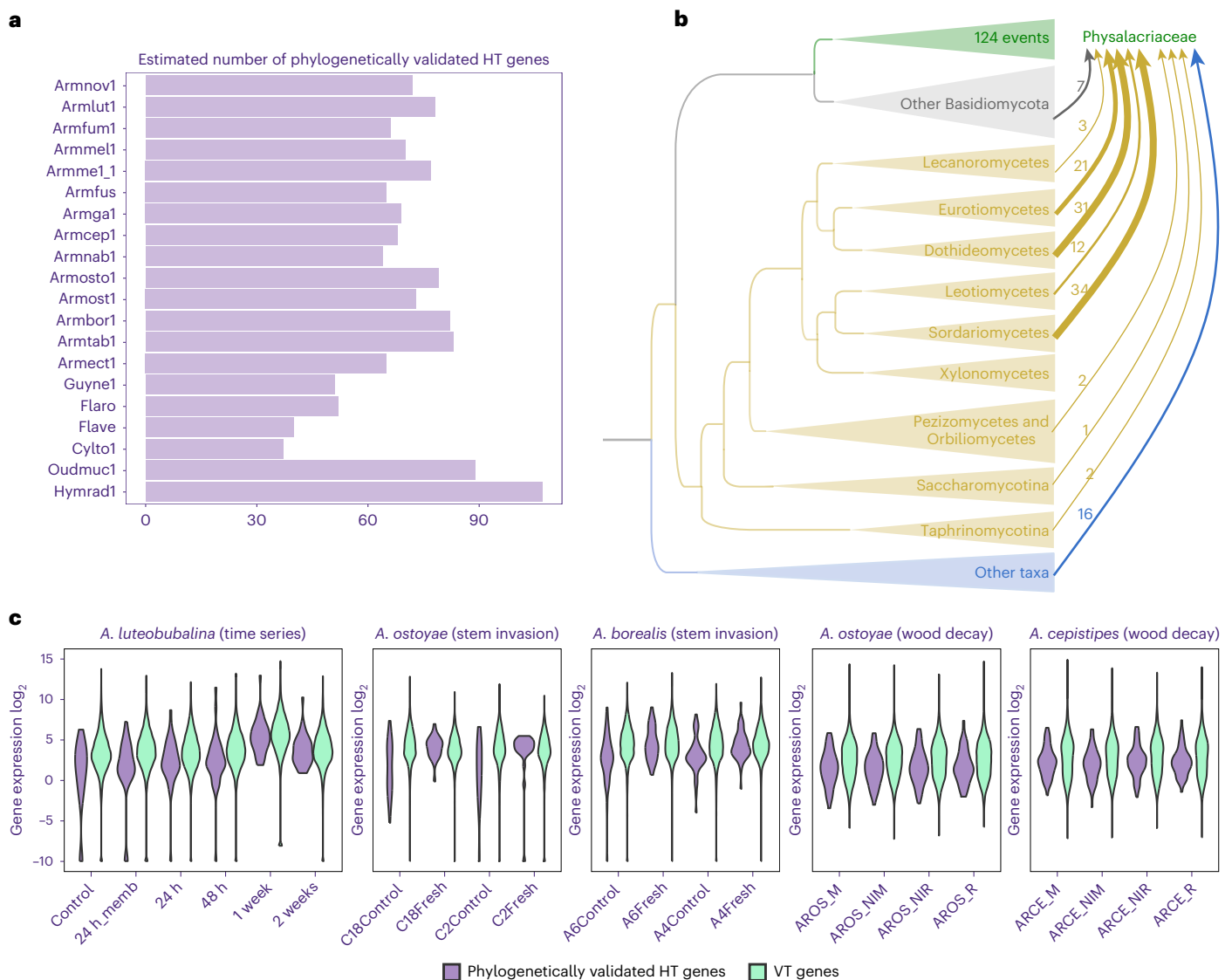


Fig. 3 | HGTs into *Armillaria* and the Physalacriaceae family. a, Bar plot showing the number of phylogenetically validated HT genes identified in each genome. **b**, Schematic representation of donor and recipient relationships for HT events after phylogenetic validation. The size of the arrows are proportional to the number of HGT events inferred from all the nodes belonging to a specific

donor group. The number of events for each donor group are listed along the arrows. **c**, Violin plots showing the expression dynamics of phylogenetically validated HT genes and vertically transferred (VT) genes. The x axis shows the sample names and the y axis shows $\log_2$ FPKMs (for the expression dynamics of HT and VT genes in the developmental dataset, see Extended Data Fig. 4).

encoding glutathione-S-transferases and catalases were enriched in wood-decay experiments.

Figure 4 also shows species- and strain-specific enrichment patterns among up- and downregulated genes. For example, the virulent *A. borealis* A6 strain had a unique enrichment of pathogenicity-related LysM proteins and cerato-platanins, which was not seen in the less-virulent A4 strain or in *A. ostoyae*. Nevertheless, in both *A. borealis* and *A. ostoyae*, the virulent strains (A6 and C18, respectively) showed more enrichment across most gene groups tested. The species- and strain-specific enrichment patterns also suggest that each species mount a different response under the examined conditions, which might correlate with differences in lifestyle or virulence.

Early stages of live host colonization by *Armillaria*

To understand the early stages of *Armillaria* infection, we analysed gene expression of *A. luteobubalina* colonizing *E. grandis* roots across five timepoints (pre-symbiosis, 24 h, 48 h, 1 week and 2 weeks) in four

biological replicates and compared these with non-symbiotic samples (Supplementary Fig. 8 and Supplementary Table 5). SuperSeq³² revealed that we could detect 77–94% of DEGs in *A. luteobubalina* across different timepoints, whereas, in *E. grandis*, we could variably detect 25–70% of DEGs (Supplementary Fig. 9). DEGs were arranged into modules based on expression similarity using the Short Time-series Expression Miner (STEM)³³. This yielded eight and four significant modules in the fungus and plant, respectively. Five fungal modules showed gradual increase in expression through time (Supplementary Fig. 10). Genes in these were enriched for GO terms often associated with pathogenic interactions³⁴, with the strongest signal for (oxidative) stress (Supplementary Fig. 10 and Extended Data Fig. 7). We identified *A. luteobubalina* DEGs related to four pathogenic processes: host immune suppression, oxidative stress, detoxification and cytotoxicity (for example, cerato-platanins).

Genes encoding glutathione-S-transferases were almost uniformly upregulated in plant-associated samples, compared with free-living mycelium, whereas other genes related to oxidative stress response

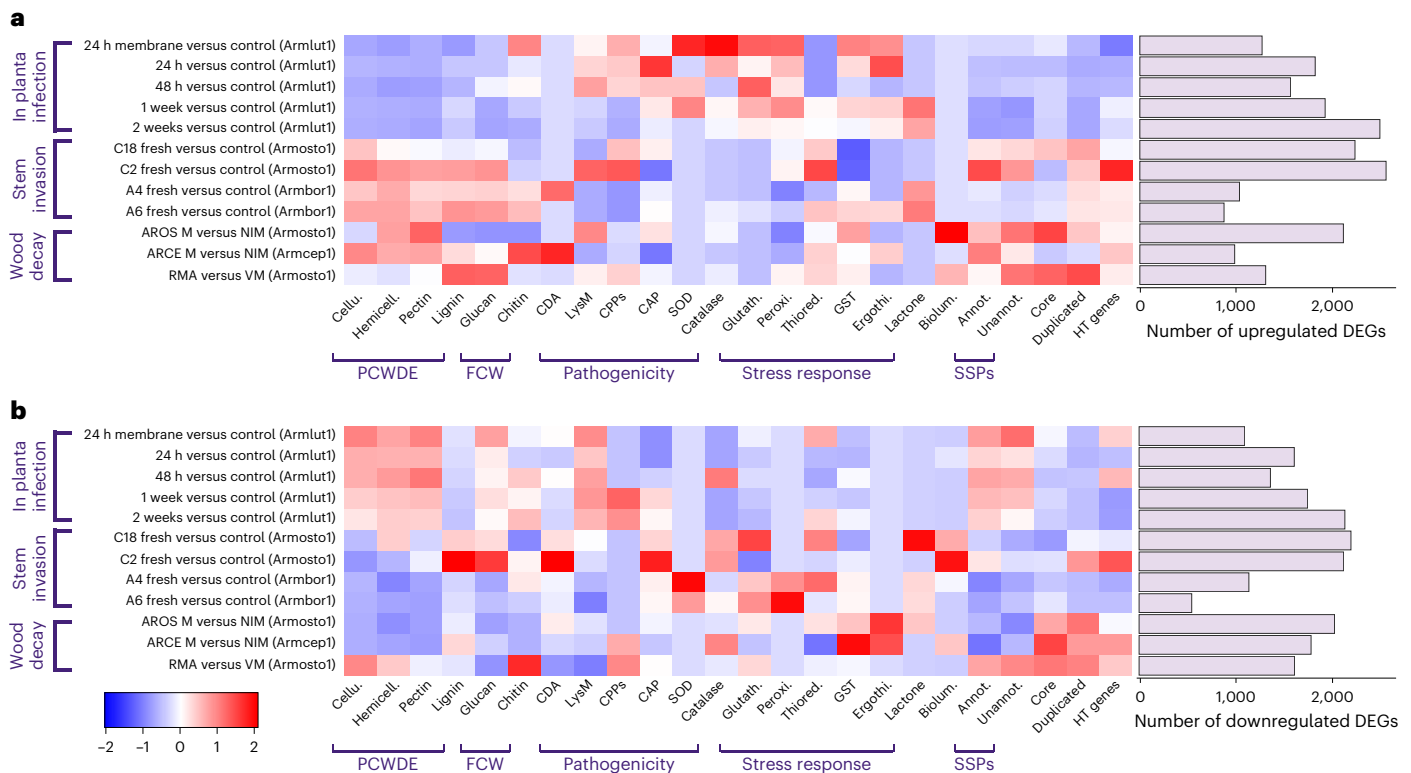


Fig. 4 | Enrichment of DEGs of wood-decay, pathogenicity, stress response and other gene families in six RNA-seq datasets. a, b. The heat maps show enrichment ratios for 24 gene groups (x axis) from aggregated differential gene expression data across six experiments (upregulated genes in **a** and downregulated genes in **b**). Enrichment ratios were calculated by considering the number of DEGs in a given gene group in the given sample, all DEGs in that sample as well as all genes in the genome and all genes in that gene group. The total number of DEGs in each experiment is shown as a bar plot on the right. In

the heat maps, warmer colours mean higher enrichment ratios (for a complete list of enrichment ratios, see Supplementary Table 5). Cellu., cellulases; Hemicell., hemicellulases; PCWDE, plant cell wall degrading enzymes; FCW, fungal cell wall; CDA, chitin deacetylase; LysM, lysine binding module; CPPs, ceratoplatans; SOD, superoxide dismutase; Glutath., glutathione system; Peroxi., peroxidases; Thioered, thioredoxins; GST, glutathione-S-transferase; Ergothi., ergothioneine system; Biolum, bioluminescence; Annot., annotated; Unannot., unannotated; SSPs, small secreted proteins; HT, horizontally transferred.

(for example, superoxide dismutases) were mostly upregulated at 1 and 2 weeks (Extended Data Fig. 7). Toxin efflux systems, which are involved in the tolerance against plant secondary and defence metabolites (for example, phytoalexins and phytoanticipins), have been reported to be transcriptionally regulated in necrotrophs³⁵. Membrane transporters, cytochrome P450 monooxygenases and laccases, which are often associated with detoxification or plant metabolites^{35,36}, showed extensive regulation during the experiment (Extended Data Fig. 7). On the other hand, we did not observe extensive regulation of PCWDEs (Fig. 4).

We observed an upregulation of HT genes at certain stages of in planta colonization. Of the 78 *A. luteobubalina* HT genes, 29 were differentially expressed at some timepoint (Supplementary Table 5). We identified two horizontally acquired CAP domain/pathogenicity-related-1 (PR-1) genes that were strongly induced across the time series and three LysM domain proteins (CBM50) that were upregulated at 24 and 48 h. Both PR-1 and LysM genes are known to be involved in fungal pathogenicity, by facilitating the transport of fatty acids and sterols³⁷, and masking the presence of chitin residues from the plant immune system, respectively³⁸.

The four significant expression modules of *E. grandis* were enriched in defence-related terms, including chitinase activity, chitin binding, defence response to fungus and defence response to bacterium (Supplementary Fig. 11), suggesting an activation of plant immunity. A finer-level analysis of the plant response identified 95 DEGs upregulated across the whole time course and 59 DEGs repressed in the same series (Supplementary Fig. 12). Within the upregulated genes, we found enrichment for phosphoprotein phosphatase activity due

to upregulation of PP2 family genes related to effector immunity and ABA-JA cross-talk³⁹. Other GO terms enriched in this gene set included unfolded protein binding/chaperone activity, heat shock and jasmonic acid response due to PR4 and MYB108 gene induction⁴⁰. Of the core repressed genes, we found GO enrichment for copper ion binding proteins, peroxidase activity, carboxylic acid binding and lipid transfer. Reduction of copper ion binding proteins, peroxidase activity and lipid transfer would probably lead to a delayed hypersensitive response as copper ions are necessary for ethylene production and repression of the ABA biosynthesis⁴¹, while peroxidase activity directly leads to plant cell death and the lipid-associated genes encode DIR-like proteins responsible for long-range immune signalling⁴². The genes associated to carboxylic acid binding and hydrolase activity, meanwhile probably affect pathogen nutrition more directly as the repressed genes associated to the former are involved in biosynthesis of L-serine, a lack of which would reduce the nutrition value of *E. grandis* tissues, while the latter are involved in detoxification (that is, cyanoalanine nitrilase), which could lead to a toxic build up of 3-cyano-L-alanine, thereby inhibiting pathogen growth⁴³.

PiSSPs

Phytopathogenic fungi often utilize effectors that suppress or manipulate host defence responses^{14,44}. As many effectors are short proteins, we scrutinized the 39 and 38 annotated and unannotated SSPs, respectively, of *A. luteobubalina* that were upregulated in at least one timepoint (Fig. 5 and Extended Data Fig. 8). Annotated SSPs with highest fold changes (FCs) included PrY1-like, SCP- or CAP-, Ricin-B-like lectin- and

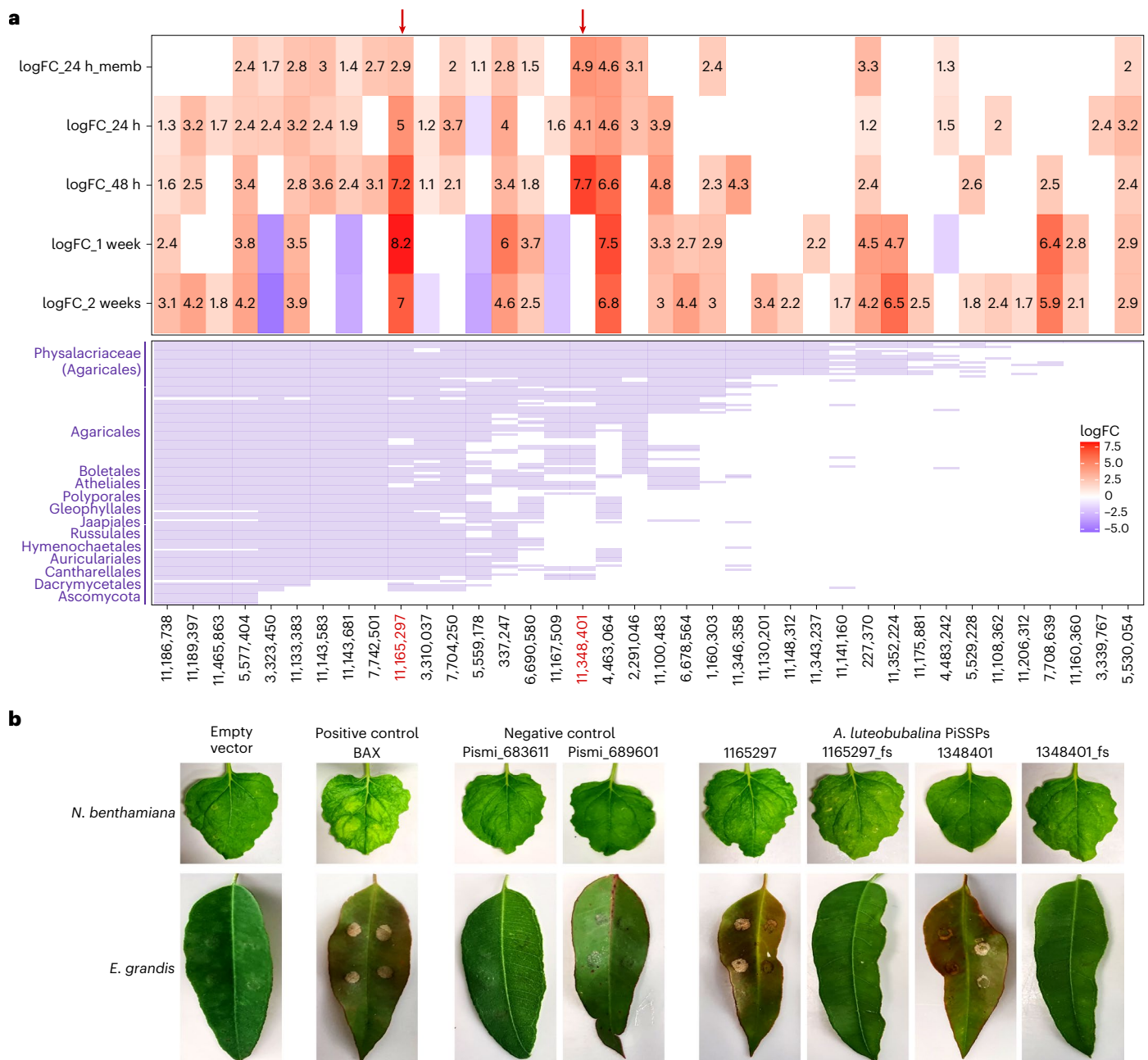


Fig. 5 | PiSSPs of *A. luteobubalina* induce cell death in host plants. a, Heat map shows log₂ FCs for SSPs significantly upregulated in at least one timepoint. Red shows higher and blue depicts lower logFC, followed by presence/absence matrix of homologues of unannotated SSPs in 131 species (Dataset 2). The two experimentally validated PiSSPs are shown by red arrows. **b**, Transient transformation of the non-host *N. benthamiana* and the host *E. grandis* with

negative controls (empty vector; *Pisolithus microsporus* proteins *Pismi_683611* and *Pismi_689601*), with a positive control (BAX) or with an in planta expression vector encoding *Armlut_1165297* or *Armlut_1348401*. As further controls, frameshift versions of the *A. luteobubalina* sequences were included (*Armlut_1165297_fs* or *Armlut_1348401_fs*).

LysM domain-containing genes (Supplementary Table 6). Two unannotated SSPs (1165297 and 1348401) had the highest FC values among all genes, so we selected these for experimental validation and call them PiSSPs; (Fig. 5a, red arrows). Gene 1348401 had a peak expression at 48 h, whereas 1165297 peaked at 1 week. The encoded proteins contain no known conserved domain and have no predicted function.

We cloned both PiSSPs (1165297 and 1348401) and expressed them transiently in the non-host *Nicotiana benthamiana* and the host *E. grandis* (Fig. 5b). The positive control expression of BAX led to cell death in *N. benthamiana* as well as in *E. grandis*. The expression of the two *A. luteobubalina* PiSSPs in *E. grandis* leaves, likewise, led to rapid

cell death (Fig. 5b), indicating a specific interaction with plant tissue. This cell death phenotype was not seen in heterologous expression of these proteins with an induced frameshift mutation, nor in leaves expressing secreted proteins from mutualistic fungi⁴⁵, nor from empty vector transformants. These data indicate that these *A. luteobubalina* PiSSPs can cause necrotic lesions in host plant tissue.

Orthology searches revealed that the two PiSSPs were surprisingly conserved (Fig. 5 and Extended Data Figs. 9 and 10). Homologues of 1348401 are present throughout the Agaricomycetes and the family strongly expanded in *Armillaria*, with on average 7 genes per species compared with 1.8 in other Agaricales (Extended Data Fig. 9). The

non-pathogenic *A. ectypa* had the fewest genes (three) in this family. The evolutionary conservation of these two proteins outside *Armillaria* is somewhat surprising and is in contrast with observations on other (mostly Ascomycota) effectors, which are believed to evolve fast and are often species specific^{18,19}. Whether they trigger cell death via direct (as true effectors) or indirect routes remains to be established. Nevertheless, these data raise the possibility that *Armillaria* use specific mechanisms for necrotrophy, as postulated by the gene-for-gene model¹⁴.

Potential virulence factors in *Armillaria* spp

In vitro invasion of freshly cut conifer stems was used to explore gene expression in highly and less virulent isolates^{46,47} of the conifer-specific *A. ostoyae* and *A. borealis* (Extended Data Fig. 5). These experiments emulated cambium colonization and killing, and virulent strains consistently colonized the stems faster. We performed differential gene expression analyses between control and invasive mycelia, as well as between virulent and less virulent strains of the same species (Supplementary Table 5). According to SuperSeq³², we detected 89.4–98.6% of all DEGs (Supplementary Fig. 9).

To uncover cambium colonization strategies, we scrutinized genes upregulated in invasive mycelia, relative to control mycelia (Supplementary Note 1). We identified upregulation of PCWDEs, in particular pectinases, as a major transcriptional response to fresh stems in both species. Enrichment ratios confirmed that pectin-related genes are induced in the largest proportions, followed by hemicellulases, cellulases and lignin-related genes (Fig. 4). HT intradiol ring-cleavage dioxygenases were also upregulated in these experiments.

We also aimed to identify genes encoding potential virulence factors by comparing highly and less virulent isolates (Supplementary Note 1). On the basis of manual curation and sequence-based comparison of upregulated genes against PHI-base, we identified 303 and 56 candidate virulence factor genes from *A. borealis* and *A. ostoyae* (Supplementary Table 6). We further determined the secreted proteins to predict potential plant interacting factors. In *A. borealis*, 40 secreted proteins, including 9 SSPs, 5 CAZymes, 2> peptidases, 2 intradiol ring-cleavage dioxygenases and 1 hydrophobin-related gene were identified. In *A. ostoyae* we found 18 genes, including 3 SSPs, 5 CAZymes and 1-1 genes representing hydrophobins, intradiol ring-cleavage dioxygenases and tyrosinases (Supplementary Table 6). Notably, highly virulent strains of both species upregulated a pair of polygalacturonase (GH28) and cellulase (GH3) genes, both homologous to virulence effectors reported from Ascomycota (Supplementary Table 6).

Discussion

The genus *Armillaria* is a globally important group of primarily tree pathogens that evolved a range of unique or rare traits, from extreme colony sizes to bioluminescence, and represent an independent origin of broad-host-range necrotrophy. In search of the genetic underpinnings of the unique biology of these fungi, in this study, we combined new genomes with transcriptomic profiling of in planta and in vitro pathosystems. Comparative genomic analyses support the view that *Armillaria* spp. have an expanded protein-coding repertoire and possess a complete set of CAZymes for degrading lignocellulosic plant biomass (including lignin)^{3,9,24,48}. However, we also detected a considerable influx of Ascomycete genes through HGT which, based on the expression data, seem to have influenced both the plant biomass degrading ability and pathogenic attributes of *Armillaria*. Previous studies showed that wood decay by *Armillaria* is reminiscent of soft rot^{26–28,49}, a lifestyle known only in Ascomycota. Congruently, our phylogenetic PCA analyses separated *Armillaria* from white rot species, which is paradoxical given that white rot is the dominant wood-decay strategy in the Agaricales³¹. Alien CAZyme genes acquired through HGT from Ascomycota may explain this controversy and establish HGT as a

source of novelty in the evolution of fungal plant biomass degrading systems in the Basidiomycota.

Armillaria are necrotrophic pathogens that first kill, then feed on their hosts^{1,9}. Our data provided the first insights into the molecular mechanisms of early and late stages of colonization of host plants. We detected genes related to four major necrotrophic processes: host immune evasion and suppression, oxidative stress, detoxification and cytotoxicity, as well as specific regulation of HT genes. The presence of PiSSPs and the lack of broad plant biomass-degrading CAZyme gene upregulation during early colonization supports the emerging view that, instead of a non-specific, brute-force attack with a broad battery of CAZymes, necrotrophs use specific effectors for modulating the plant immune system¹⁴. The expression patterns of the two PiSSPs are consistent with the gene-for-gene theory of pathogenicity, which has traditionally been discussed for (hemi-)biotrophs⁵⁰, but may apply to necrotrophs as well¹⁴. Determining whether the two PiSSPs are truly specific effectors will require uncovering the mechanisms of their action, especially in light of their broad conservation, but these data provide an entry point into investigating fine details of *Armillaria*–plant interactions. In contrast to early stage infection, cambium colonization was characterized by a broad upregulation of plant cell wall-degrading CAZymes. Comparisons of low and highly virulent strains in these assays further provided insights into genes associated with virulence, such as certain polygalacturonase- or cellulase-encoding genes.

Overall, this study revealed the molecular underpinnings of the lifestyle of a group of widespread pathogens of woody plants, which, combined with existing and emerging experimental tools, will facilitate research on the molecular mechanisms of plant colonization. We propose that genome evolution in *Armillaria* relied on combined effects of multiple types of genetic innovation, including HGT, and that genes gained early during the evolution of the genus were integrated into cellular regulatory networks of plant biomass degradation and pathogenicity.

Methods

Strains for sequencing, assembly and annotation

Armillaria strains (Supplementary Table 1) were inoculated on malt extract agar and incubated at 25 °C in the dark for 7–10 days. The identity of each strain was confirmed by amplifying and sequencing the fungal internal transcribed spacer (ITS1-5.8S-ITS2) region and bacterial contamination was checked using universal 16S primers (Integrated DNA Technologies). Strains were inoculated in malt extract broth in 500 ml Erlenmeyer flasks and incubated at 25 °C in the dark for 4–5 weeks until a substantial quantity of fungal biomass was grown. Fungal mass was stored at –80 °C until DNA and RNA extraction. Before nucleic acid extraction, fungal tissues were homogenized using liquid nitrogen in a mortar and pestle. DNA extraction was performed using the Blood and Cell Culture DNA Maxi Kit (Qiagen) and RNA extraction was performed using the RNeasy Midi Kit (Qiagen) as per the manufacturer's instructions. Using the DNA, we again confirmed strain identity by sequencing the internal transcribed spacer region and nucleic acid quantity was measured using a Qubit (ThermoFisher) according to the manufacturer's instructions.

Genomic and transcriptomic library preparation, sequencing, assembly and annotation. For *A. nabsnona*, *A. mellea*, and *A. ectypa*, 5 µg of genomic DNA was sheared to approximately 15–20 kb using Megaruptor3 (Diagenode). The sheared DNA was treated with DNA Prep to remove single-stranded ends, and DNA damage repair mix followed by end repair, A-tail and ligation of PacBio overhang adaptors using SMRTbell Express Template Prep 2.0 Kit (Pacific Biosciences). The final libraries were size selected with BluePippin (Sage Science) at 10 kb cutoff size. For *A. borealis*, *A. fumosa*, *A. novae-zelandiae* and *A. tabescens*, 5 µg of genomic DNA was left unshredded due to marginal high molecular weight (HMW) DNA quality. The unshredded gDNA was

treated with exonuclease to remove single-stranded ends and DNA damage repair mix followed by end repair and ligation of blunt adaptors using SMRTbell Template Prep Kit 1.0 (Pacific Biosciences). All libraries were purified with AMPure PB beads.

The PacBio Sequencing primers were then annealed to the SMRTbell template libraries and sequencing polymerase was bound to them using the Sequel Binding kit 2.0 (*A. nabsnana*, *A. mellea*, *A. tabescens* and *A. ectypa*), 2.1 (*A. borealis*) or 3.0 (*A. fumosa* and *A. novae-zelandiae*). The prepared SMRTbell template libraries were then sequenced on a Pacific Biosciences' Sequel sequencer using v3 sequencing primer, 1 M v2 (*A. borealis*, *A. nabsnana*, *A. mellea*, *A. tabescens* and *A. ectypa*) or v3 (*A. fumosa* and *A. novae-zelandiae*) SMRT cells, and Version 2.1 (*A. borealis*, *A. nabsnana*, *A. mellea*, *A. tabescens* and *A. ectypa*) or 3.0 (*A. fumosa* and *A. novae-zelandiae*) sequencing chemistry with 1 × 360 and 1 × 600 sequencing movie run times.

Filtered subread data were processed to remove artefacts and assembled together with Falcon (<https://github.com/PacificBiosciences/FALCON>) version 1.8.8 (*A. ectypa*, *A. nabsnana*, *A. tabescens* and *A. borealis*) or version pb-assembly 0.0.2, falcon-kit 1.2.3, pypeflow 2.1.0 (*A. mellea*, *A. novae-zelandiae* and *A. fumosa*) to generate an initial assembly. Mitochondrial genomes were assembled separately from the Falcon pre-assembled reads (preads) using an in-house tool (assemblemito.sh (*A. ectypa*, *A. nabsnana*, *A. tabescens*, *A. borealis* and *A. mellea*) or assemblemito.py (*A. fumosa* and *A. novae-zelandiae*)), used to filter the preads, and polished with Arrow version SMRTLink v5.0.1.9578 (*A. ectypa* and *A. nabsnana*), v5.1.0.26412 (*A. tabescens* and *A. borealis*), v6.0.0.47841 (*A. mellea*) or v7.0.1.66975 (*A. novae-zelandiae* and *A. fumosa*) (<https://github.com/PacificBiosciences/GenomicConsensus>). A secondary Falcon assembly was generated using the mitochondria-filtered preads, improved with finisherSC⁵¹ version 2.0 (*A. ectypa* and *A. nabsnana*) or 2.1 (*A. fumosa*), and polished with Arrow version SMRTLink v5.0.1.9578 (*A. ectypa* and *A. nabsnana*), v5.1.0.26412 (*A. tabescens* and *A. borealis*), v6.0.0.47841 (*A. mellea*) or v7.0.1.66975 (*A. novae-zelandiae* and *A. fumosa*). Contigs less than 1,000 bp were excluded.

Stranded complementary DNA libraries were generated using the Illumina Truseq Stranded mRNA Library Prep kit. Messenger RNA was purified from 200 ng (*A. nabsnana*) or 1 µg (all other species) of total RNA using magnetic beads containing poly-T oligos. For *A. nabsnana*, mRNA was fragmented using divalent cations and high temperature. The fragmented RNA was reverse transcribed using random hexamers and SSII (Invitrogen) followed by second strand synthesis. For all other species, mRNA was fragmented and reverse transcribed using random hexamers and SSII (Invitrogen) followed by second strand synthesis. The fragmented cDNA was treated with end-pair, A-tailing, adaptor ligation and ten (*A. nabsnana*) or eight (all other species) cycles of PCR.

All prepared libraries were quantified using KAPA Biosystems' next-generation sequencing library qPCR kit and run on a Roche Light-Cycler 480 real-time PCR instrument. Sequencing was performed on the Illumina NovaSeq sequencer using NovaSeq XP V1 reagent kits, S4 flowcell (*A. fumosa*, *A. novae-zelandiae* and *A. mellea*) or on the Illumina HiSeq2500 sequencer using TruSeq paired-end cluster kits v4 and TruSeq SBS sequencing kits v4 (*A. nabsnana*, *A. borealis*, *A. tabescens* and *A. ectypa*), following a 2 × 150 indexed run recipe.

Raw reads were evaluated with BBDuk (<https://sourceforge.net/projects/bbmap/>) for artefact sequence by kmer matching (kmer = 25), allowing one mismatch and detected artefacts were trimmed from the 3' end of the reads. RNA spike-in reads, PhiX reads and reads containing any Ns were removed. Quality trimming was performed using the phred trimming method set at Q6. Following trimming, reads under the length threshold were removed (minimum length 25 bases or one-third of the original read length, whichever is longer). Filtered reads were assembled into consensus sequences using Trinity v. 2.3.2 (ref. 52) with the `-normalize_reads` and `-jaccard_clip` options.

Genomes were annotated with the support of their corresponding transcriptomes using the JGI Annotation Pipeline. Assemblies and annotations are available from the JGI Genome Portal MycoCosm (<https://mycocosm.jgi.doe.gov>)^{53,54}.

Taxon sampling and functional annotations

We used three datasets in this study. Dataset 1 comprises 66 species (64 Agaricales, 2 Boletales outgroups), which were used to analyse gene family evolution in *Armillaria* (Supplementary Table 1, Dataset 1). This was extended into a phylogenetically more diverse Dataset 2, which was used to perform gene copy number analysis of wood-decay gene repertoires and comprised 131 fungal species from both Basidiomycota and Ascomycota, including species with various lifestyles such as white rotters, brown rotters, litter decomposers, ectomycorrhizal (ECM) fungi, pathogens and species with uncertain life history strategies (Supplementary Table 1, Dataset 2). Dataset 3 was used to identify HT genes and comprised 942 species, including species ranging from fungi, bacteria as well as plants (Supplementary Table 1, Dataset 3). All the species used in these datasets are published and publicly available sequences from JGI MycoCosm (except a few that were used with the Author's permission). The version numbers of the genomes are provided in Supplementary Table 1 for reference. For transposable element (TE) classification, Repeatmasker v4.0.3 (ref. 55) was run (with the parameters `-nolow -no_is -s`) on the *Armillaria* and Physalacriaceae outgroup species genome assemblies using a database composed of the Repbase repeat library 25.03 (ref. 56) and additional rDNA sequences from MycoCosm. For genomes identified as diploid, only primary scaffolds were considered. For *A. fuscipes*, *F. velutipes* and *F. rossica*, only scaffolds >1,000 bp were considered. Proteomes of all the species used in this manuscript were functionally annotated by using InterPro (IPR) Scan v.5.46-81.0 (ref. 57). Secreted and small secreted (at least 300 amino acids) proteins were identified as described previously⁵⁸. The luciferase cluster and its synteny were identified in *Armillaria* species and five Physalacriaceae outgroups as described previously²⁰. Orthologous groups from these genomes were inferred using OrthoFinder v2.5.4 (refs. 59,60) and synteny around the luciferase cluster was visualized based on ordering and orienting scaffolds using the *Armillaria ostoyae* C18/9 genome and using genoPlotR⁶¹ package. The hispidin synthase and cytochrome P450 gene were previously missing and re-annotated on the scaffold of NODE_104435 in *Armillaria mellea* DSM 3731 using predictions of *Armillaria mellea* ELDO17 v1.0.

Genome evolution in *Armillaria* spp

Reconstruction of genome-wide gene family evolution was performed using Dataset 1. Predicted protein sequences were clustered into OGs using OrthoFinder v2.3.1 (refs. 59,60) with default settings. For the species tree, multiple sequence alignment for 81 single-copy OGs (SCOGs) were inferred using MAFFT v7.453 (-auto)⁶², followed by removal of gapped regions from the alignments using TrimAl v1.2 (ref. 63) with a gap threshold of 0.9. Trimmed alignments were then concatenated into a supermatrix using an in-house R script. The species tree was reconstructed in RAxML v8.2.12 (ref. 64) with the option of rapid bootstrap analysis and search for the best-scoring maximum-likelihood (ML) tree under the PROTGAMMAWAG model of protein evolution. The tree was rooted using FigTree and is provided in Supplementary Table 1.

We analysed the genome-wide duplications and losses across gene families using the COMPARE method⁶⁵, which uses reconciled gene trees to perform Dollo parsimony mapping and orthologue coding to tell apart duplications from speciation events. To obtain the reconciled gene trees for COMPARE, we first inferred multiple sequence alignments for the OrthoFinder OGs with at least four proteins using MAFFT v7.453 (-auto)⁶². Aligned clusters were trimmed to remove spurious regions using TrimAl v1.2 (ref. 63) with -strict parameter. Clusters for which TrimAl resulted in blank alignments (due to -strict parameter)

were used in their non-trimmed form. In total, 16,936 trimmed and 819 non-trimmed clusters were used in RAXML v8.2.12 (ref. 64) for gene tree inference, followed by Shimodaira–Hasegawa-like branch support calculation, also in RAXML under the PROTGAMMAWAG substitution model. The 17,755 gene trees with their Shimodaira–Hasegawa-like support values, along with the species tree were used for rerooting, followed by gene tree species tree reconciliation in Notung v2.9 run in batch mode with the edge weight threshold set to 70. Reconciled gene trees were used to reconstruct the gene duplication/loss scenarios in OGs along the species tree using the COMPARE pipeline (available at <https://github.com/laszloagnagy/COMPARE>). For each gene family, the number of gains, losses, net gains (sum of gains and losses) and ancestral copy numbers were obtained and their summary was mapped back onto the species tree. GO terms significantly enriched (P values < 0.05) among the duplications at *Armillaria* MRCA were identified by topGO⁶⁶ using the weight01 algorithm and Fisher testing.

Analyses of CAZymes

Dataset 2 was used to analyse where Physalacriaceae (including *Armillaria*) fits in terms of their wood-decay repertoires. Annotation of genes encoding CAZymes for the required species were downloaded from JGI Mycocosm. Species whose CAZymes were not present in JGI Mycocosm were annotated using the CAZy annotation pipeline⁶⁷. On the basis of previously published reports, annotated CAZymes were further categorized on the basis of their putative plant cell wall-degrading preferences (Supplementary Table 3) into those acting on cellulose, hemicellulose, pectin and lignin. We subjected the proteomes of these 131 species to OrthoFinder v2.3.1 (refs. 59,60) with default settings, giving us a total of 41,205 OGs. For inferring the species tree, we generated SCOGs using a set of in-house R scripts, which gave us 548 OGs in which the duplications were caused by gains at species levels (or terminal gains). From these, we omitted proteins showing less similarity based on amino acid distance against the other members of the cluster. In this step, we used at least 40% of all species for each cluster. Finally, the resulting 514 SCOGs were concatenated (minimum 60 length of AA, and at least 66 species) into one supermatrix and followed by species tree inference in IQ-Tree v1.6.12 (ref. 68).

We generated phylogenetic PCA using the *phyl.pca*⁶⁹ function from Phytools⁷⁰ for Physalacriaceae along with white rot and litter decomposer fungi. We compared *Armillaria* and other Physalacriaceae to only white rot and litter decomposer fungi because including brown rot and mycorrhizal species, which have huge differences in gene content, reduced the resolution of the plots. We provided the function with a gene copy number matrix normalized according to proteome sizes for each substrate category and the above described ML species tree (subsampled to white rot, litter decomposer and Physalacriaceae species) as inputs. Independent contrasts were calculated under the Brownian motion model and the parameter mode = 'cov'. The substrate-wise gene copy numbers and their resulting phylogenetic PCA loadings are provided in Supplementary Table 3. We also plotted their proteome-size normalized copy numbers into bar plots for visualization of copy numbers across the phylogeny (Supplementary Figs. 3–7). We further investigated whether Physalacriaceae members have CAZymes shared with Ascomycota. For checking this hypothesis, we first identified the CAZy OGs from the 131 species, by selecting OGs with at least 50% of the proteins annotated as CAZymes, and at least 5 proteins. This gave us 401 CAZy OGs, which were used to perform Fisher's exact test-based enrichment for identifying co-enriched CAZy families in Physalacriaceae and Ascomycota with respect to white rot and litter decomposers (Supplementary Table 3).

Analyses of HT genes

To check whether Physalacriaceae species are more similar to Ascomycota than expected by chance, which could either come from HGT or long-retained ancestral genes, we compiled the

Dataset 3. This comprised a taxonomically diverse dataset with a total of 942 species, ranging from 110 fungi from Basidiomycota (including 20 Physalacriaceae), 741 from Ascomycota, 26 from Mucoromycota, 10 from Zoopagomycota, 13 from Chytridiomycota and 17 other early diverging fungi, as well as 15 bacterial and 10 plant species.

We identified candidate HT genes using the AI, for which we ran MMseqs (easy-search, e-value 0.001) (ref. 71) using all Physalacriaceae proteomes merged together as our query, against a database of proteomes of all the other species as our subject. Using an in-house R script, we parsed the output to retain only the top hits (based on e-values) for each taxonomic group. AI scores were calculated as $(\log_{10}(\text{best hit to species within the group lineage} + 1 \times 10^{-200}) - \log_{10}(\text{best hit to species outside the group lineage} + 1 \times 10^{-200}))$ (ref. 72).

Further, for phylogenetic validation of HT events, we used a subset of 329 species from Dataset 3 based on AI of top HT donors. In this case, we restricted our donor list to major HT contributors only, viz. Ascomycota, followed by Mucoromycota and Zoopagomycota as outgroups. The proteomes of these 329 species were subjected to OrthoFinder v2.5.4 (refs. 59,60) clustering with default settings. From these, we fished out the OGs with at least one AI-based HT gene and also the OGs containing CAZymes. Further, we retained OGs having at least one Physalacriaceae and one Ascomycota/Mucoromycota/Zoopagomycota species, resulting in a total of 675 OGs.

Proteins from these 675 OGs were aligned using MAFFT v7.453 (-auto)⁶² Trimal v1.2 (-strict)⁶³. Trimmed alignments were used to infer gene trees using IQ-Tree v1.6.12 (ref. 68) with ultrafast bootstrap (1,000 replicates) under the WAG + G substitution model. Using an in-house R-script, we identified potential HT events by extracting clades from the gene trees based on support values (>70%) and taxon occupancy (receiver clade with >70% Physalacriaceae species, donor clade with >70% Ascomycota and finally sister clade with >70% Ascomycota to ensure the direction of gene transfer). Physalacriaceae proteins from these putative HT clades were used as a query against the UniRef100 database⁷³ from the UniProt Reference Cluster using MMseqs easy-search, e-value 0.001 (ref. 71). To remove the low sequence similarity matches and false hits, we retained hits with percent identity and bidirectional coverage of more than 45%. We parsed the filtered hits to retain only the top 100 hits (based on percent identity) for each query, and the taxonomic category contributing predominantly among the top 100 hits was assigned as the putative donor. Further, we classified the assigned donor into two categories: type A, where the top 100 hits were dominated and confidently associated with Ascomycota, and type B, where the top 100 hits were dominated by non-Agaricomycetes species, however, were ambiguously distributed among different taxa (Supplementary Table 4).

Live plant and stem invasion pathosystems

To understand gene expression in *Armillaria* and its host during the early stages of host infection, we performed live in planta assays between *Eucalyptus grandis* and *Armillaria luteobubalina* (Extended Data Fig. 5b). *A. luteobubalina* was cultured onto half-strength modified Melin–Norkrans (MMN) media (pH 5.5; 1 g l⁻¹ glucose) and grown for 1 month in the dark. *E. grandis* seeds were sterilized for 10 min in 30% hydrogen peroxide (H₂O₂, vol/vol) and germinated on 1% (wt/vol) water agar for 1 month (25 °C; 16 h light cycle). Four weeks before contact with the fungus, seedlings were transferred onto half-strength MMN medium (pH 5.5; 1 g l⁻¹ glucose) and grown at 22–30 °C night/day temperature with a 16 h light cycle. Once the *E. grandis* seedlings were 2 months old and the fungal cultures had grown for 1 month, the plants were separated into one of three treatment categories: (1) fungal-free controls whereby the seedlings were transferred onto new half-strength MMN medium, (2) 'pre-symbiosis', which involved the transfer of seedlings onto new half-strength MMN medium in indirect contact with fungal mycelium for 24 h by separating the two organisms by a permeable cellophane membrane, and (3) 'physical contact' seedlings were transferred onto new half-strength MMN medium

and then placed into direct contact with the active growing edge of a fungal colony and then samples were collected at 24 h, 48 h, 1 week and 2 weeks post-contact. These plates were then closed using micropore tape to allow for gas exchange with the external environment. Four biological replicates per treatment and timepoint were generated and collected at the described timepoint into liquid nitrogen and stored at -80°C until RNA extraction.

For stem invasion assays, high and low virulent isolates of conifer-specific *A. ostoyae* (C18, highly virulent; C2, low virulent) and *A. borealis* (A6, highly virulent; A4, low virulent)^{46,47} were maintained in the dark on Petri dishes on RS medium (40 g malt extract, 20 g dextrose, 5 g bacto peptone, 19 g agar/1 l) at 24°C and 4°C , respectively (Extended Data Fig. 5a). Before the start of the experiment, fresh cultures of all isolates were set up. The system for growing subcortical mycelial fans in the laboratory consisted of plastic jars containing a layer of moistened and inoculated rice, sawdust, tomato and orange (RSTO) medium⁹. After approximately 10 days of incubation in dark at 24°C , 10-cm-long, freshly cut stem segments of Norway spruce were placed on top of this growing mycelial lawn, allowing *Armillaria* to invade the stem segments. The timing and effectiveness of the advancing invasive mycelial fans were followed by their arrival in small monitoring ‘windows’ (1×1.5 cm) cut into the bark halfway up to the top of the stems. The samples were collected soon after the mycelial front line appeared in the cutouts. *Armillaria* vegetative mycelia collected from the jars without spruce stems were used as controls.

RNA isolation and sequencing

For in planta assays, live tissues and frozen samples were collected and extracted using the ISOLATE II Plant miRNA kit (Bioline) as per the manufacturer’s instructions. Following extraction, the RNA samples were sequenced at the Beijing Genome Institute. For the stem invasion assays, sections of the mycelial fan that were collected from under the bark of infected colonized stems were frozen in liquid nitrogen and stored at -80°C . RNA was isolated from the frozen samples using the RNeasy Plant Midi Kit (Qiagen) according to the manufacturer’s protocol. Before the RNA isolation, fungal tissues were homogenized with the help of liquid nitrogen, and a mortar and pestle. RNA quantity was measured using a Qubit (ThermoFisher) according to the manufacturer’s protocol. Biological triplicates were analysed for all sample types. The libraries for Illumina sequencing were prepared using NEBNext Ultra II Directional RNA Library Prep Kit for Illumina (NEB). Briefly, 100 ng RNA was enriched using RiboCop rRNA Depletion Kits (Lexogen). Thereafter, the RNA was fragmented, end prepped and adaptor ligated. Finally, the libraries were amplified according to the manufacturer’s instructions. The quality of the libraries was checked on Agilent 4200 TapeStation System using D1000 Screen Tape (Agilent Technologies), the quantity was measured on Qubit 3.0. Illumina sequencing was performed on the NovaSeq 6000 instrument (Illumina) with 2×151 run configuration. Raw RNA-seq reads were aligned against the *A. ostoyae* (National Center for Biotechnology Information genome GCA_900157425.1 version 2) and *A. borealis* (JGI: *Armillaria borealis* FPL87.14 v1.0) genomes using STAR v2.7.5a (ref. 74). After alignment, the level of expression was estimated using RSEM v1.3.1 (ref. 75).

Analysis of expression data

For the stem invasion assay, the estimated counts matrix was used for differential expression analysis using Limma-Voom⁷⁶. Before running the differential expression analysis, counts were normalized using the trimmed mean of M values using edgeR v3.38.1 (ref. 77) and lowly expressed genes ($\text{cpm} \leq 10$) were filtered. Differential gene expression analysis for the in planta pathosystem data was carried out as in our previous study²⁶.

In both analyses, genes with at least two-fold expression change and false discovery rate values ≤ 0.05 were considered significant. Multi-dimensional scaling (‘plotMDS’ function in edgeR) was applied

to visualize gene expression profiles. Expression heat maps were generated by hierarchically clustering based on average linkage of fragments per kilobase million (FPKM) values with the heatmap.2 function from the gplots package

To check the reliability of both expression data, we used the superSeq package³², which uses a subsampling approach to simulate and predict the number of DEGs at lower read depths and random sampling points from the original dataset. These subsampled reads are then extrapolated to predict the relationship between statistical power and read depth.

Genes from the time series in planta assay from both fungal and the plant side were clustered into co-expression based modules using STEM v1.3.11 (ref. 33). GO terms significantly enriched (P values < 0.05) among each of the STEM modules were identified by topGO⁶⁶, using the weight01 algorithm and Fisher testing.

Enrichment ratios for specific gene families in each gene expression dataset were calculated using the fisher.test function in R which uses the formula $(AD)/(BC)$, where A represents the number of DEGs in the gene family of interest, B represents the total number of genes in that specific gene family, C represents the total number of DEGs excluding the gene family of interest and D represents the total number of proteins excluding the DEGs, the gene family of interest and the total genes in that gene family. These enrichment or odds ratio for different gene families in each experiment were visualized using the heatmap.2 function from the gplots package. Expression heatmaps for different gene families in the six datasets are available in Figshare (https://figshare.com/articles/figure/Gene_expression_heatmaps/22778477?file=40472333).

Experimental validation of PiSSPs

Aliquots of RNA extracted from the in planta pathosystem assay (6 timepoints, 24 samples) were pooled and used for generation of full-length cDNA using the Tetro cDNA synthesis kit (Bioline) according to the manufacturer’s instructions and using only the oligo-dT primer. Two *A. luteobubalina* HWK02 v1.0 SSP sequences were amplified using the KAPA HiFi Hotstart Readymix (Roche) according to the manufacturer’s instructions using the following primers: *Armlut1_1165297* (5'-GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGTTTCYARYTNYTNTTYGCNGC-3' and 5'-GGGGACCACTTTGTACAAGAAAGCTGGGTGTTACTCTTCTAGCAGCTTCACGTC-3') and *Armlut1_1348401* (5'-GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGTTGTTTCAGTTTCTTCTCTTCTACC-3' and 5'-GGGGACCACTTTGTACAAGAAAGCTGGGTGCTARTCNACRCANGTNGTNSWDATRTCTCNG-3'). As negative controls, we cloned these same genes with a single nucleotide frameshift using the primers *Armlut1_1165297_fs* (5'-GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGTTTCYARYTNYTNTTYGCNGC) and *Armlut1_1348401_fs* (5'-GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGTTTCAGTTTCTTCTCTTCTACC) with the same reverse primer as per the sense constructs. As a further set of negative cell death transformation controls, we also cloned two SSPs lacking the secretion domain from the mutualistic ectomycorrhizal fungus *Pisolithus microcarpus* isolate SI14 to act as negative controls for the induction of cell death. These were Pismi_683611 and Pismi_689601, effector proteins that do not cause visible plant damage when expressed in planta (Pismi_683611 (ref. 45), Pismi_689601 (ref. 78)). Successful amplification was verified using gel electrophoresis, products were gel purified using the Wizard SV Gel and PCR Cleanup kit (Promega). As a positive control for cell death, we synthesized the full-length protein apoptosis regulator BAX isoform 1 from *Mus musculus* (NP_031553.1; MDGSGEQLGSGGPTSSEQIMKTGAFLQGFQIDRAGRMAGETPELTLEQPPQDASTKKLSECLRRIGDELDSNMELQRMADVDTDSPREVFVRVAADMFDGNFNFNWRVVALFYFASKLVKALCTKVPILRTIMGWTLDFLRERLLVWIQDQGGWEGLLSYFGTPTWQTVTIFVAGVLTASLTWKKMG; Twist Biosciences) with codon usage optimized for in planta expression. All of these constructs were ligated into pDONR221-b1b4 (Invitrogen)

using BP ligase (Invitrogen). Ligations were transformed into *Escherichia coli* and positive colonies were selected based on PCR screening. Positive bacterial colonies were grown overnight in Luria-Bertani (LB) + kanamycin (50 mg l⁻¹) and plasmids purified using the PureLink HiPure Plasmid Miniprep kit. Gene inserts were sequence verified using Sanger sequencing at the Hawkesbury Institute for the Environment sequencing platform. Plasmids containing the proper gene insert were used in an LR ligation with pBiFCt-2in1 destination vector with an empty cassette ligated into the b2b3 position. Similarly, as a control for in planta transformations, a version of pBiFCt-2in1 with empty cassettes was generated as an empty vector control. Screening procedures following were as per with pDONR221. Positive plasmids were purified and transformed into *Agrobacterium tumefaciens* GV3101. Expression in *N. benthamiana* was performed as per Plett et al.⁴⁵. Similarly, young leaves of *E. grandis* were agroinfiltrated using the same solutions as for *N. benthamiana* and put into growth chambers at 25 °C constant temperature (16 h light/8 h dark). As opposed to *N. benthamiana* leaves, *E. grandis* leaves were more recalcitrant to infiltration without damaging the leaves, therefore the bacteria was only introduced into a circle the size of the syringe bore. Both plant systems were left until visible symptoms developed in the positive control and the leaves were photographed at the site of infiltration. Leaves from three separate plants were infiltrated for each plant model system, and representative images are presented.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

New genomic assemblies and annotation generated in this study are deposited under the 1000 Fungal Genome Project at JGI Mycocosm (<https://mycocosm.jgi.doe.gov/Armillaria/Armillaria.info.html>) and at DDBJ/EMBL/GenBank under the accession numbers PRJNA463936, PRJNA500536, PRJNA500837, PRJNA519860, PRJNA519861, PRJNA571622, PRJNA677793 and PRJNA677794. New RNA-seq datasets used in this study are deposited in the National Center for Biotechnology Information's Gene Expression Omnibus Archive at <https://www.ncbi.nlm.nih.gov/geo/>. Accession number for the in planta assay between *A. luteobubalina* and *E. grandis* is PRJNA975488, or GSE233220. For the stem invasion assay, the accession numbers are PRJNA972908 for *A. ostoyae* and PRJNA972989 for *A. borealis*. Phylogenetically validated gene trees and gene expression heatmaps for various gene families for the six RNA-seq datasets used in this study can be found in the Figshare repository at https://figshare.com/articles/dataset/Gene_trees/22730534 and https://figshare.com/articles/figure/Gene_expression_heatmaps/22778477?file=40472333 respectively. Source data are provided with this paper.

Code availability

Codes associated with the data analyses and visualization are available at <https://github.com/nehasahu486/Armillaria-phylogenomics/tree/main>.

References

- Baumgartner, K., Coetzee, M. P. A. & Hoffmeister, D. Secrets of the subterranean pathosystem of *Armillaria*: subterranean pathosystem of *Armillaria*. *Mol. Plant Pathol.* **12**, 515–534 (2011).
- Heinzelmann, R. et al. Latest advances and future perspectives in *Armillaria* research. *Can. J. Plant. Pathol.* **41**, 1–23 (2019).
- Sipos, G., Anderson, J. B. & Nagy, L. G. *Armillaria*. *Curr. Biol.* **28**, PR297–R298 (2018).
- Coetzee, M., Wingfield, B. & Wingfield, M. *Armillaria* root-rot pathogens: species boundaries and global distribution. *Pathogens* **7**, 83 (2018).
- Baumgartner, K. Root collar excavation for postinfection control of armillaria root disease of grapevine. *Plant Dis.* **88**, 1235–1240 (2004).
- Koch, R. A. et al. Symbiotic nitrogen fixation in the reproductive structures of a basidiomycete fungus. *Curr. Biol.* **31**, 3905–3914. e6 (2021).
- Ford, K. L., Henricot, B., Baumgartner, K., Bailey, A. M. & Foster, G. D. A faster inoculation assay for *Armillaria* using herbaceous plants. *J. Hort. Sci. Biotechnol.* **92**, 39–47 (2017).
- Devkota, P. & Hammerschmidt, R. The infection process of *Armillaria mellea* and *Armillaria solidipes*. *Physiol. Mol. Plant Pathol.* **112**, 101543 (2020).
- Sipos, G. et al. Genome expansion and lineage-specific genetic innovations in the forest pathogenic fungi *Armillaria*. *Nat. Ecol. Evol.* **1**, 1931–1941 (2017).
- Liang, X. & Rollins, J. A. Mechanisms of broad host range necrotrophic pathogenesis in *Sclerotinia sclerotiorum*. *Phytopathology* **108**, 1128–1140 (2018).
- O'Connell, R. J. et al. Lifestyle transitions in plant pathogenic *Colletotrichum* fungi deciphered by genome and transcriptome analyses. *Nat. Genet.* **44**, 1060–1065 (2012).
- Newman, T. E. & Derbyshire, M. C. The evolutionary and molecular features of broad host-range necrotrophy in plant pathogenic fungi. *Front. Plant Sci.* **11**, 591733 (2020).
- Olson, Å. et al. Insight into trade-off between wood decay and parasitism from the genome of a fungal forest pathogen. *N. Phytol.* **194**, 1001–1013 (2012).
- Shao, D., Smith, D. L., Kabbage, M. & Roth, M. G. Effectors of plant necrotrophic fungi. *Front. Plant Sci.* **12**, 687713 (2021).
- Koch, R. A., Wilson, A. W., Séné, O., Henkel, T. W. & Aime, M. C. Resolved phylogeny and biogeography of the root pathogen *Armillaria* and its gasteroid relative, *Guyanagaster*. *BMC Evol. Biol.* **17**, 33 (2017).
- Baccelli, I. Cerato-platanin family proteins: one function for multiple biological roles? *Front. Plant Sci.* **5**, 769 (2015).
- Muraosa, Y., Toyotome, T., Yahiro, M. & Kamei, K. Characterisation of novel-cell-wall LysM-domain proteins LdpA and LdpB from the human pathogenic fungus *Aspergillus fumigatus*. *Sci. Rep.* **9**, 3345 (2019).
- Plett, J. M. & Plett, K. L. Leveraging genomics to understand the broader role of fungal small secreted proteins in niche colonization and nutrition. *ISME Commun.* **2**, 49 (2022).
- Lo Presti, L. et al. Fungal effectors and plant susceptibility. *Annu. Rev. Plant Biol.* **66**, 513–545 (2015).
- Ke, H.-M. et al. *Mycena* genomes resolve the evolution of fungal bioluminescence. *Proc. Natl Acad. Sci. USA* **117**, 31267–31277 (2020).
- Nagy, L. G. et al. Genetic bases of fungal white rot wood decay predicted by phylogenomic analysis of correlated gene–phenotype evolution. *Mol. Biol. Evol.* **34**, 35–44 (2017).
- Sun, P. et al. Fungal glycoside hydrolase family 44 xyloglucanases are restricted to the phylum Basidiomycota and show a distinct xyloglucan cleavage pattern. *iScience* **25**, 103666 (2022).
- Resl, P. et al. Large differences in carbohydrate degradation and transport potential among lichen fungal symbionts. *Nat. Commun.* **13**, 2634 (2022).
- Collins, C. et al. Genomic and proteomic dissection of the ubiquitous plant pathogen, *Armillaria mellea*: toward a new infection model system. *J. Proteome Res.* **12**, 2552–2570 (2013).
- Daniel, G., Volc, J. & Nilsson, T. Soft rot and multiple T-branching by the basidiomycete *Oudemansiella mucida*. *Mycol. Res.* **96**, 49–54 (1992).
- Sahu, N. et al. Hallmarks of basidiomycete soft- and white-rot in wood-decay -omics data of two armillaria species. *Microorganisms* **9**, 149 (2021).

27. Campbell, W. G. The chemistry of the white rots of wood. *Biochem. J.* **25**, 2023–2027 (1931).
28. Schwarze, F. W. M. R. Wood decay under the microscope. *Fungal Biol. Rev.* **21**, 133–170 (2007).
29. Gladyshev, E. A., Meselson, M. & Arkhipova, I. R. Massive horizontal gene transfer in bdelloid rotifers. *Science* **320**, 1210–1213 (2008).
30. Worrall, J. J., Anagnost, S. E. & Zabel, R. A. Comparison of wood decay among diverse lignicolous fungi. *Mycologia* **89**, 199 (1997).
31. Floudas, D. et al. Evolution of novel wood decay mechanisms in Agaricales revealed by the genome sequences of *Fistulina hepatica* and *Cylindrobasidium torrendii*. *Fungal Genet. Biol.* **76**, 78–92 (2015).
32. Bass, A. J., Robinson, D. G. & Storey, J. D. Determining sufficient sequencing depth in RNA-Seq differential expression studies. Preprint at *bioRxiv* <https://doi.org/10.1101/635623> (2019).
33. Ernst, J. & Bar-Joseph, Z. STEM: a tool for the analysis of short time series gene expression data. *BMC Bioinf.* **7**, 191 (2006).
34. Bautista, D. et al. Comprehensive time-series analysis of the gene expression profile in a susceptible cultivar of tree tomato (*Solanum betaceum*) during the infection of *Phytophthora betacei*. *Front. Plant Sci.* **12**, 730251 (2021).
35. Westrick, N. M., Smith, D. L. & Kabbage, M. Disarming the host: detoxification of plant defense compounds during fungal necrotrophy. *Front. Plant Sci.* **12**, 651716 (2021).
36. Lah, L. et al. The versatility of the fungal cytochrome P450 monooxygenase system is instrumental in xenobiotic detoxification: fungal P450 systems in xenobiotic detoxification. *Mol. Microbiol.* **81**, 1374–1389 (2011).
37. Darwiche, R. et al. Plant pathogenesis-related proteins of the cacao fungal pathogen *Moniliophthora perniciosa* differ in their lipid-binding specificities. *J. Biol. Chem.* **292**, 20558–20569 (2017).
38. Gao, F. et al. Deacetylation of chitin oligomers increases virulence in soil-borne fungal pathogens. *Nat. Plants* **5**, 1167–1176 (2019).
39. Saito, N. et al. Roles of RCN1, regulatory A subunit of protein phosphatase 2A, in methyl jasmonate signaling and signal crosstalk between methyl jasmonate and abscisic acid. *Plant Cell Physiol.* **49**, 1396–1401 (2008).
40. Cui, F., Brosché, M., Sipari, N., Tang, S. & Overmyer, K. Regulation of ABA dependent wound induced spreading cell death by MYB108. *N. Phytol.* **200**, 634–640 (2013).
41. Liu, H. et al. Copper ions suppress abscisic acid biosynthesis to enhance defence against *Phytophthora infestans* in potato. *Mol. Plant Pathol.* **21**, 636–651 (2020).
42. Maldonado, A. M., Doerner, P., Dixon, R. A., Lamb, C. J. & Cameron, R. K. A putative lipid transfer protein involved in systemic resistance signalling in *Arabidopsis*. *Nature* **419**, 399–403 (2002).
43. O’Leary, B., Preston, G. M. & Sweetlove, L. J. Increased β -Cyanoalanine Nitrilase activity improves cyanide tolerance and assimilation in *Arabidopsis*. *Mol. Plant* **7**, 231–243 (2014).
44. Tanaka, S. & Kahmann, R. Cell wall-associated effectors of plant-colonizing fungi. *Mycologia* **113**, 247–260 (2021).
45. Plett, J. M. et al. Mycorrhizal effector PaMiSSP10b alters polyamine biosynthesis in *Eucalyptus* root cells and promotes root colonization. *N. Phytol.* **228**, 712–727 (2020).
46. Heinzelmann, R., Prospero, S. & Rigling, D. Virulence and stump colonization ability of *Armillaria borealis* on Norway spruce seedlings in comparison to sympatric *Armillaria* species. *Plant Dis.* **101**, 470–479 (2017).
47. Prospero, S., Holdenrieder, O. & Rigling, D. Comparison of the virulence of *Armillaria cepistipes* and *Armillaria ostoyae* on four Norway spruce provenances. *For. Pathol.* **34**, 1–14 (2004).
48. Collins, C. et al. Proteomic Characterization of *Armillaria mellea* reveals oxidative stress response mechanisms and altered secondary metabolism profiles. *Microorganisms* **5**, 60 (2017).
49. Campbell, W. G. The chemistry of the white rots of wood. *Biochem. J.* **26**, 1829–1838 (1932).
50. Flor, H. H. Current status of the gene-for-gene concept. *Annu. Rev. Phytopathol.* **9**, 275–296 (1971).
51. Lam, K.-K., LaButti, K., Khalak, A. & Tse, D. FinisherSC: a repeat-aware tool for upgrading *de novo* assembly using long reads. *Bioinformatics* **31**, 3207–3209 (2015).
52. Grabherr, M. G. et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–652 (2011).
53. Grigoriev, I. V. et al. MycoCosm portal: gearing up for 1000 fungal genomes. *Nucleic Acids Res.* **42**, D699–D704 (2014).
54. Kuo, A., Bushnell, B. & Grigoriev, I. V. Fungal Genomics. In *Advances in Botanical Research* vol. 70 1–52 (Elsevier, 2014).
55. Smit, A. F. A., Hubley, R. & Green, P. RepeatMasker Open-4.0. 2013–2015 (2015).
56. Bao, W., Kojima, K. K. & Kohany, O. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob. DNA* **6**, 11 (2015).
57. Jones, P. et al. InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236–1240 (2014).
58. Almási, É. et al. Comparative genomics reveals unique wood-decay strategies and fruiting body development in the Schizophyllaceae. *N. Phytol.* **224**, 902–915 (2019).
59. Emms, D. M. & Kelly, S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* **16**, 157 (2015).
60. Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**, 238 (2019).
61. Guy, L., Roat Kultima, J. & Andersson, S. G. E. genoPlotR: comparative gene and genome visualization in R. *Bioinformatics* **26**, 2334–2335 (2010).
62. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
63. Capella-Gutierrez, S., Silla-Martinez, J. M. & Gabaldon, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009).
64. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
65. Nagy, L. G. et al. Latent homology and convergent regulatory evolution underlies the repeated emergence of yeasts. *Nat. Commun.* **5**, 4471 (2014).
66. Alexa, A. & Rahenfuhrer, J. topGO: enrichment analysis for Gene Ontology. <https://doi.org/10.18129/B9.bioc.topGO> (2016).
67. Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M. & Henrissat, B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* **42**, D490–D495 (2014).
68. Nguyen, L.-T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).
69. Revell, L. J. Size-correction and principal components for interspecific comparative studies. *Evolution* **63**, 3258–3268 (2009).
70. Revell, L. J. phytools: an R package for phylogenetic comparative biology (and other things): phytools: R package. *Methods Ecol. Evol.* **3**, 217–223 (2012).
71. Steinegger, M. & Söding, J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat. Biotechnol.* **35**, 1026–1028 (2017).
72. Alexander, W. G., Wisecaver, J. H., Rokas, A. & Hittinger, C. T. Horizontally acquired genes in early-diverging pathogenic fungi enable the use of host nucleosides and nucleotides. *Proc. Natl Acad. Sci. USA* **113**, 4116–4121 (2016).

73. Suzek, B. E. et al. UniRef clusters: a comprehensive and scalable alternative for improving sequence similarity searches. *Bioinformatics* **31**, 926–932 (2015).
 74. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
 75. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinf.* **12**, 323 (2011).
 76. Ritchie, M. E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
 77. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinform. Oxf. Engl.* **26**, 139–140 (2010).
 78. Plett, J. M. et al. Speciation underpinned by unexpected molecular diversity in the mycorrhizal fungal genus. *Pisolithus*. *Mol. Biol. Evol.* **40**, msad045 (2023).
- B.I., H.-M.K., S.K., E.D., B.B., B.H., M.V., S.C., I.J.T., J.S. and L.G.N. carried out data analysis. E.D. and B.H. annotated CAZymes for the genomes not available in JGI Mycocosm. N.S., J.S., Z.M., S.K. and L.G.N. analysed HGT events. S.A., T.-L.M., A.L., B.A., J.P., A.P., K.B., K.L., M.K., M.Y., R.R. and I.G.V. performed genome sequencing, assembly and annotation. J.P. and K.L.P. performed PiSP experimental validation. K.L.F. and GDF contributed strains to the genome sequencing. C.B.H. contributed genomic data. L.G.N., N.S., J.P., F.M.M., J.S., S.K., G.S. and D.R. wrote the manuscript. All authors reviewed, checked and approved the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41564-023-01448-1>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41564-023-01448-1>.

Correspondence and requests for materials should be addressed to László G. Nagy.

Peer review information *Nature Microbiology* thanks Jonathan Schilling and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature Limited 2023

Acknowledgements

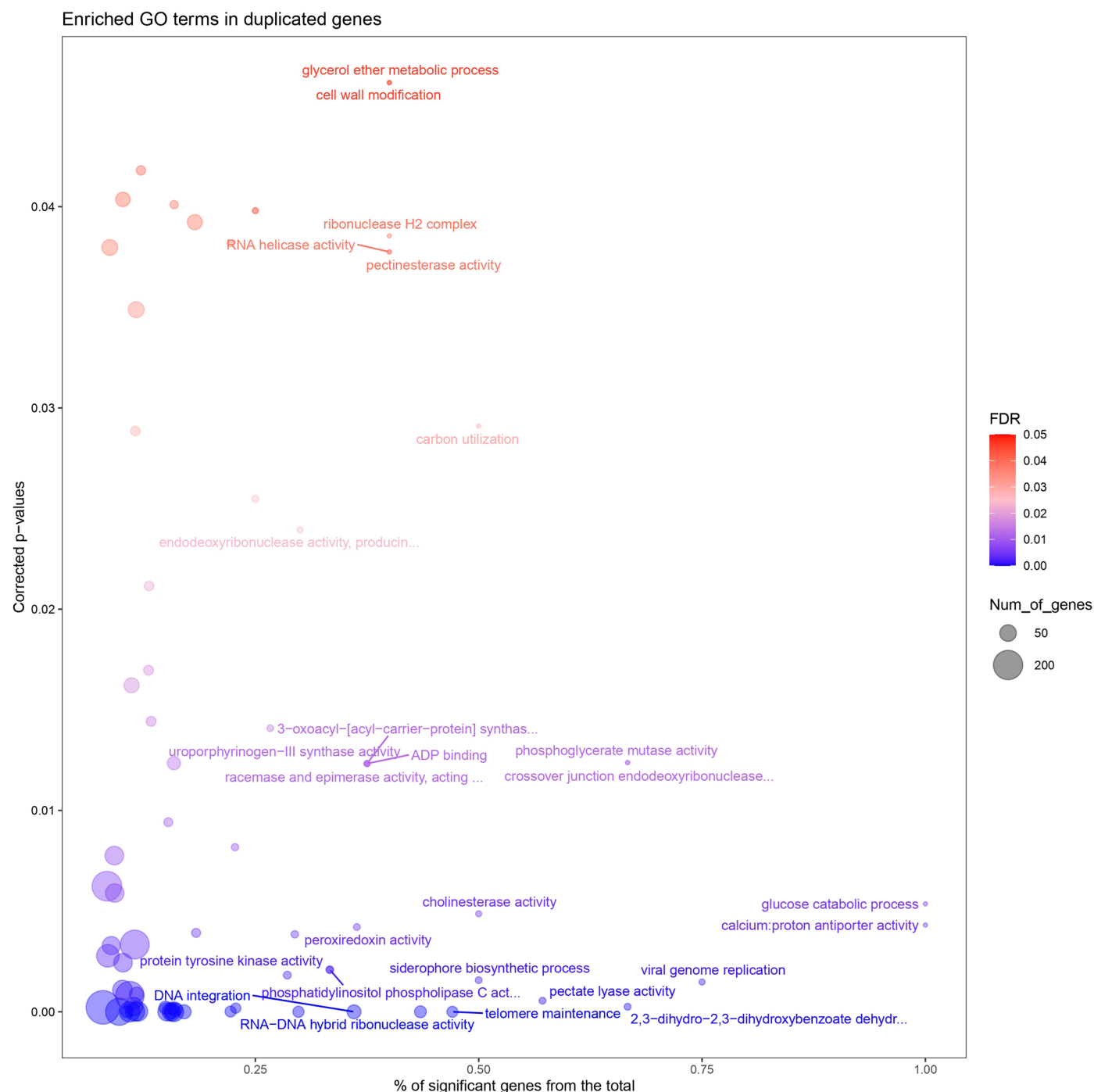
We acknowledge support by the 'Momentum' programme of the Hungarian Academy of Sciences (contract no. LP2019-13/2019 to L.G.N.) the European Research Council (grant no. 758161 to L.G.N.) as well as the Eotvos Lorand Research Network (SA-109/2021). G.S. acknowledges support by the Hungarian National Research, Development, and Innovation Office (GINOP-2.3.2-15-2016-00052). The work (proposals: <https://doi.org/10.46936/10.25585/60001060> and <https://doi.org/10.46936/10.25585/60001019>) conducted by the US Department of Energy (DOE) Joint Genome Institute (<https://ror.org/04xm1d337>), a DOE Office of Science User Facility, is supported by the Office of Science of the US DOE operated under contract no. DE-AC02-05CH11231. The research was performed in collaboration with the Genomics and Bioinformatics Core Facility at the Szentágotthai Research Centre of the University of Pécs. Ian Hood and Pam Taylor (Scion Research, New Zealand Forest Research Institute Ltd.) kindly provided the *A. nova-zealandiae* 2840 strain. D. Lindner (Forest Products Laboratory, USA) kindly shared strains of *A. borealis* and *A. ectypa* for sequencing. We appreciate the permission of G. Bonito for using the genome of *Flagelloscypha* sp.

Author contributions

N.S., L.G.N., J.P. and G.S. conceived the study. N.S., B.I., J.W.-B., Z.M., K.L.P. and J.P. carried out the laboratory experiments, including DNA/RNA isolation for genome and transcriptome sequencing. N.S., Z.M.,

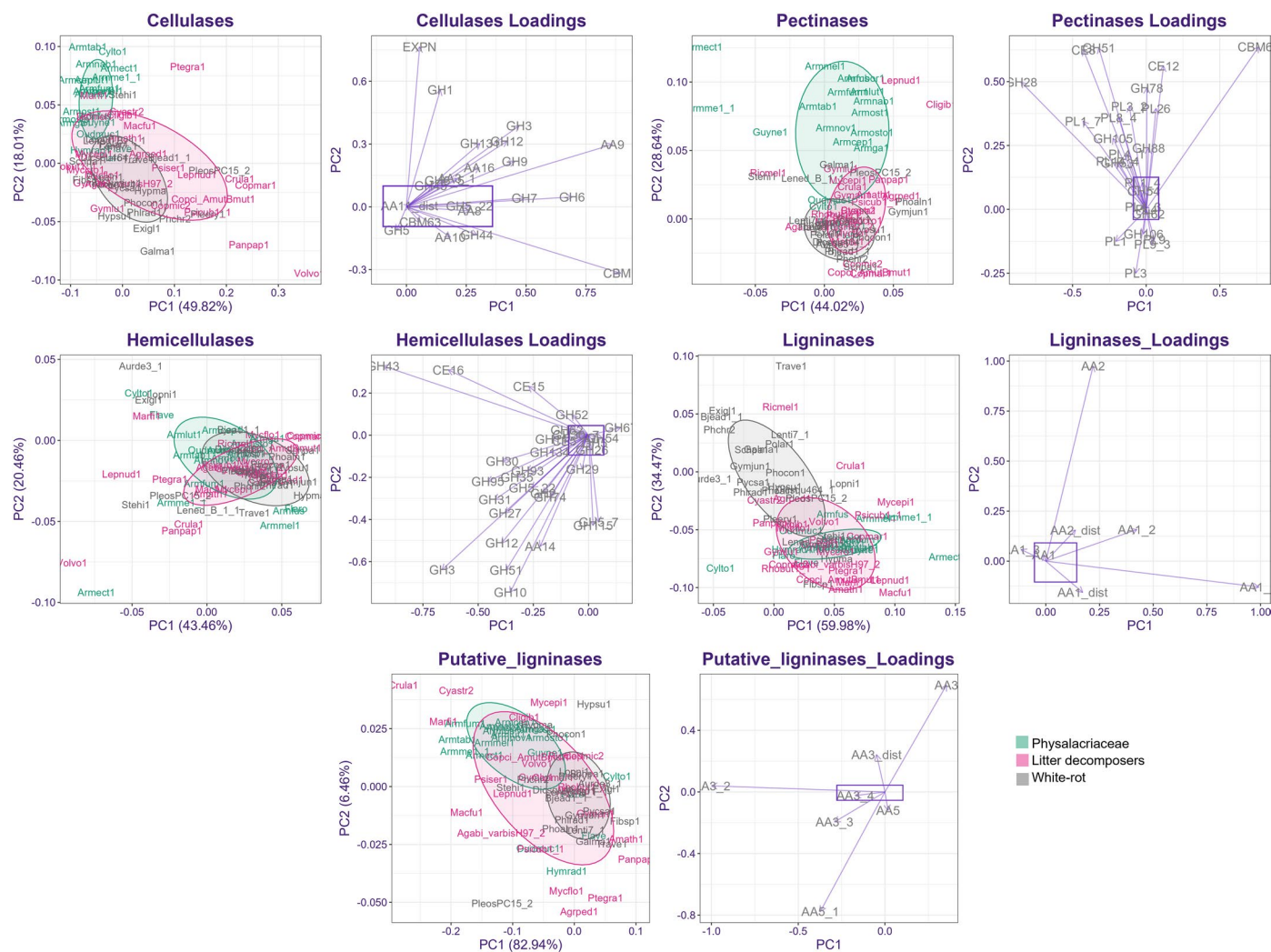
¹Biological Research Center, Synthetic and Systems Biology Unit, Szeged, Hungary. ²Doctoral School of Biology, Faculty of Science and Informatics, University of Szeged, Szeged, Hungary. ³Functional Genomics and Bioinformatics Group, Faculty of Forestry, Institute of Forest and Natural Resource Management, University of Sopron, Sopron, Hungary. ⁴Hawkesbury Institute for the Environment, Western Sydney University, Richmond, New South Wales, Australia. ⁵Elizabeth Macarthur Agricultural Institute, NSW Department of Primary Industries, Menangle, New South Wales, Australia. ⁶Biodiversity Research Center, Academia Sinica, Taipei, Taiwan. ⁷US Department of Energy Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, CA, USA. ⁸Department of Microbiology, Faculty of Science and Informatics, University of Szeged, Szeged, Hungary. ⁹ELKH-SZTE Fungal Pathogenicity Mechanisms Research Group, University of Szeged, Szeged, Hungary. ¹⁰Architecture et Fonction des Macromolécules Biologiques (AFMB), CNRS, Aix-Marseille Université, Marseille, France. ¹¹INRAE, UMR 1163, Biodiversité et Biotechnologie Fongiques, Marseille, France. ¹²DTU Bioengineering, Technical University of Denmark, Kongens Lyngby, Denmark. ¹³Department of Biological Sciences, King Abdulaziz University, Jeddah, Saudi Arabia. ¹⁴Université de Lorraine, INRAE, UMR 1136 'Interactions Arbres/Microorganismes', Centre INRAE Grand Est – Nancy, Champenoux, France. ¹⁵Department of Biology, Section of Terrestrial Ecology, University of Copenhagen, København Ø, Denmark. ¹⁶Department of Biosciences, University of Oslo, Blindern, Oslo, Norway. ¹⁷Swiss Federal Research Institute WSL, Birmensdorf, Switzerland. ¹⁸School of Biological Sciences, Life Sciences Building, University of Bristol, Bristol, UK. ¹⁹Department of Plant and Microbial Biology, University of California Berkeley, Berkeley, CA, USA. ²⁰Department of Plant Pathology, The Ohio State University, Columbus, OH, USA. ²¹Present address: Department of Microbiology, Soochow University, Taipei, Taiwan. ✉e-mail: lnagy@funkenomelab.com



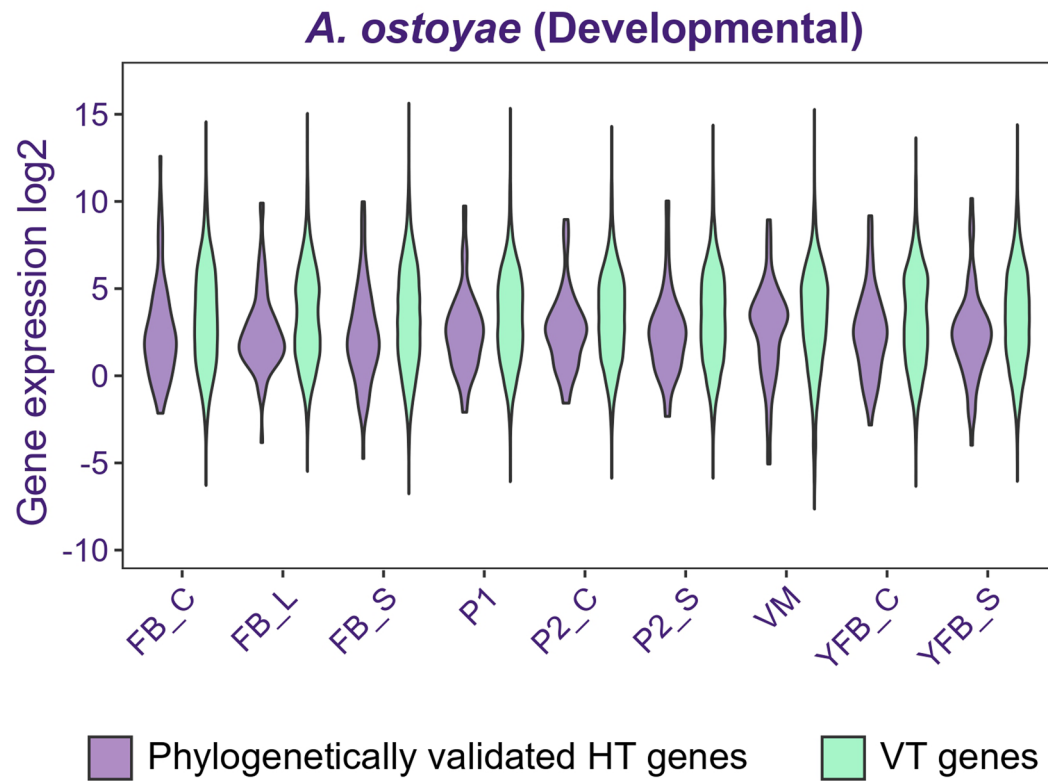


Extended Data Fig. 2 | GO terms enriched in duplicated genes. Significantly enriched GO terms in the 1473 orthogroups, inferred by 2913 duplications at *Armillaria* MRCA. GO enrichment analysis was performed using one-sided Fisher's test with the weight01 algorithm in the topGO package (R), with p-value ≤ 0.05 considered as significant. X-axis shows the percentage of significant genes

from the total number of genes, y-axis shows p-values. Blue shows lower and red shows higher p-values. GO terms that had at least 30% of genes significant from the total number of genes are mentioned on the plot (see Supplementary Table 2 for the complete list of enriched GO terms).



Extended Data Fig. 3 | Plant biomass degradation related genes in *Armillaria*. Phylogenetic PCAs and their respective loading factors for PCWDE gene families. Species abbreviations are colored according to nutritional modes.



Extended Data Fig. 4 | Expression of HT and VT genes in *A. ostoyae* developmental transcriptome. Violin plot showing gene expression of phylogenetically validated HT and VT genes in *A. ostoyae* fruiting body development transcriptome. Y-axis shows log₂ transformed expression values, and x-axis shows the sample comparisons for each experiment.

A**B**

CONTROL

With *Armillaria*

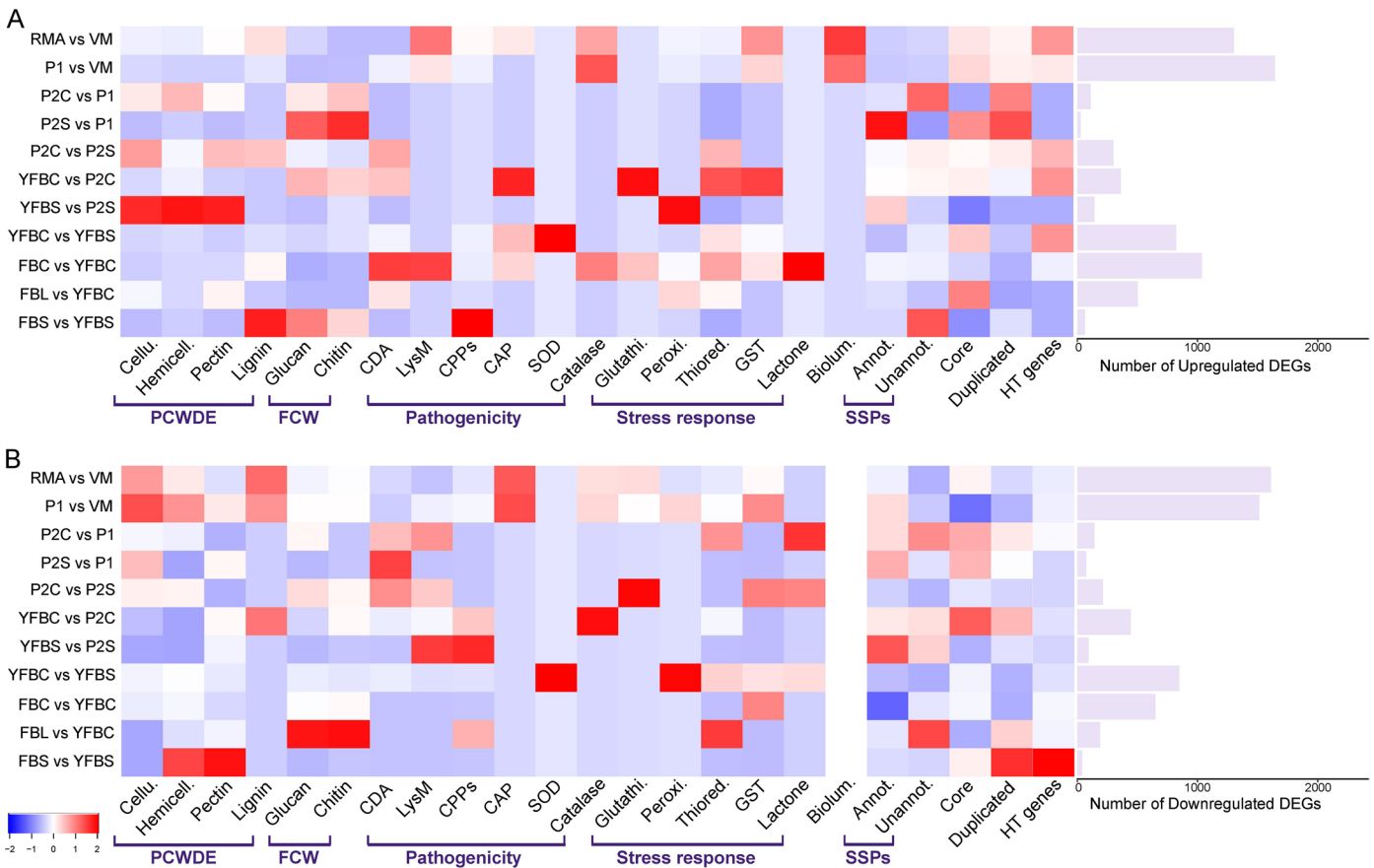
24 hrs

48 hrs

1 week

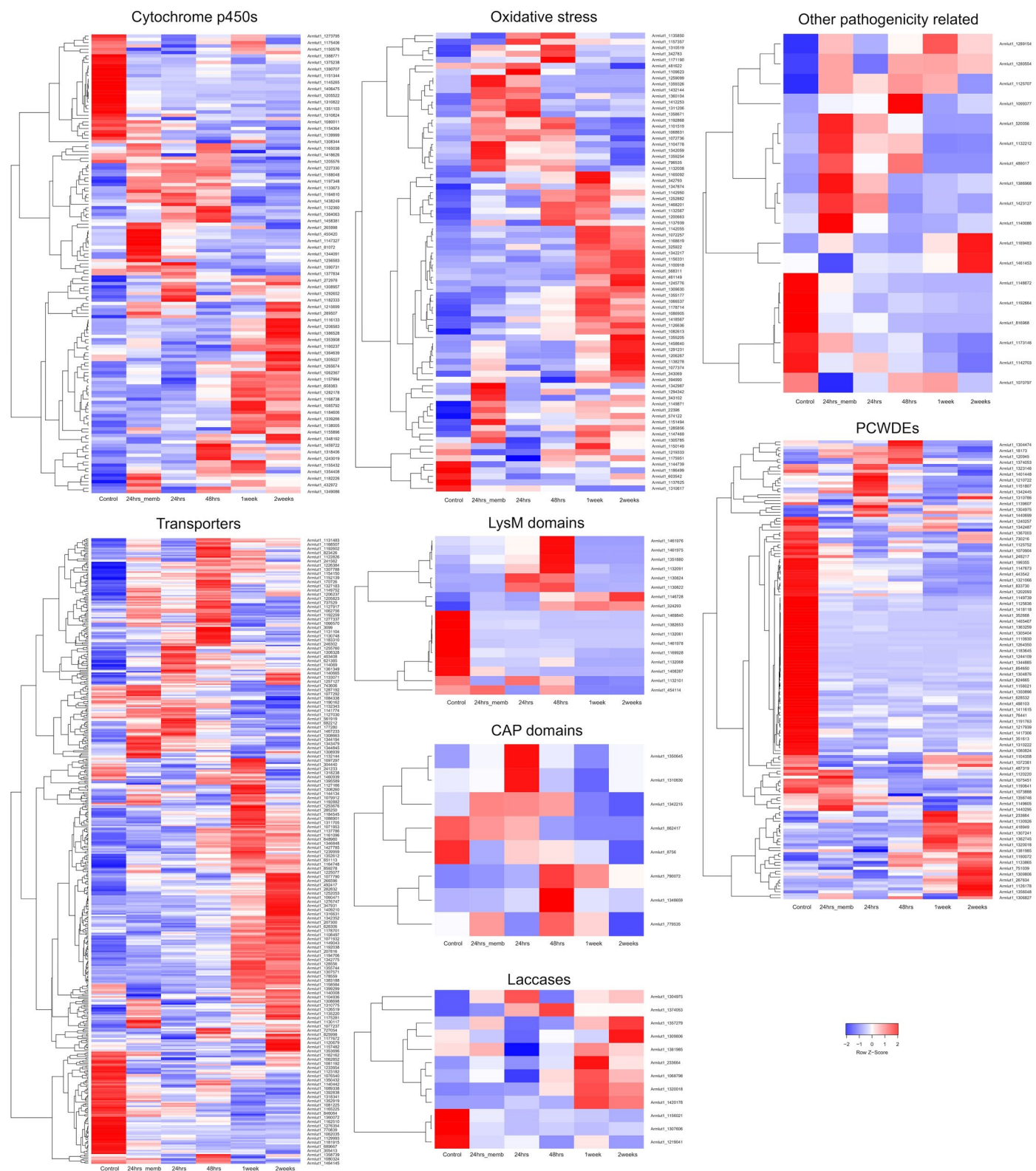
2 weeks

Extended Data Fig. 5 | Experimental setup. Setup for the new RNA Seq experiments used in this study. **a)** Setup for the time-course experiment. **b)** Setup for the stem invasion assay.

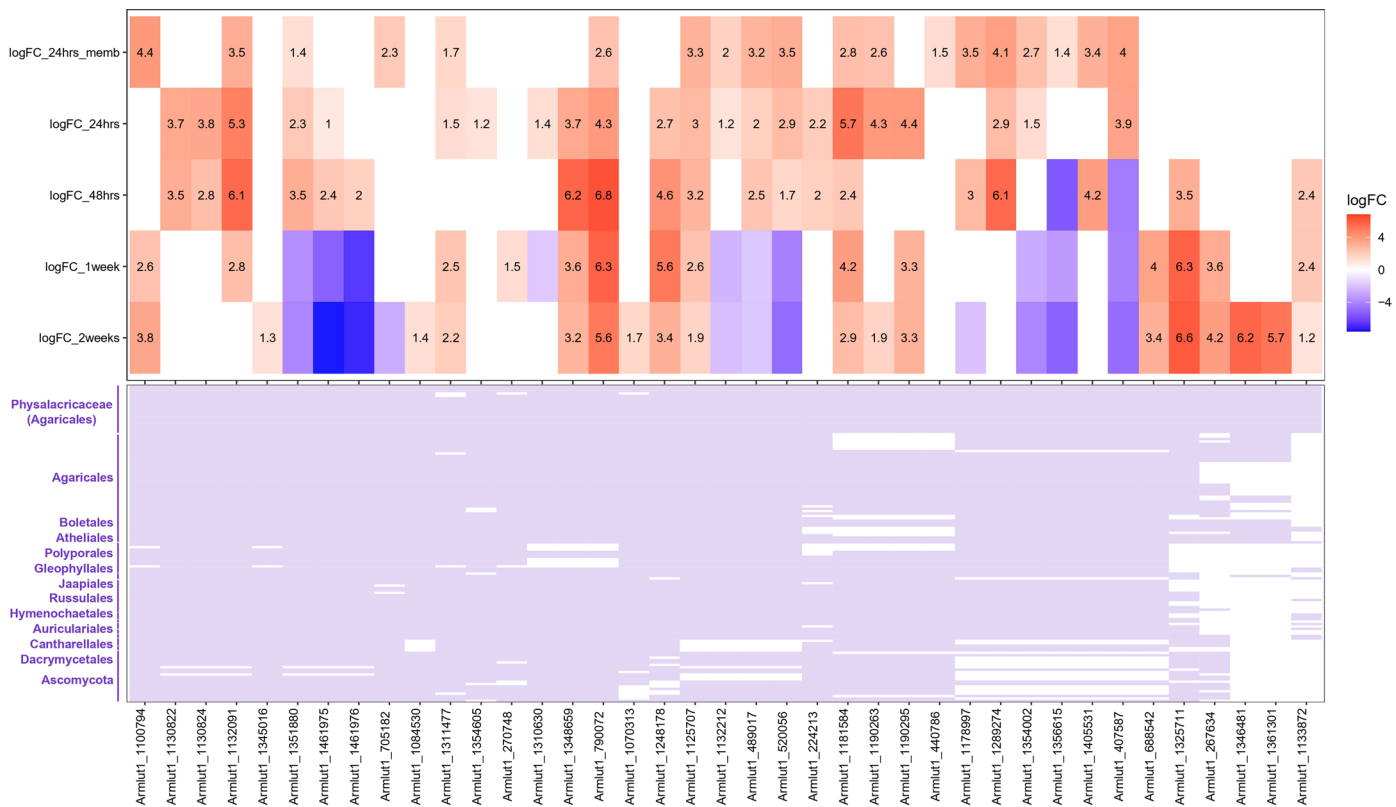


Extended Data Fig. 6 | Enrichment of differentially expressed genes of selected gene families in 6 RNA-Seq datasets. The heatmap shows enrichment ratios for 23 gene groups ('Ergothione: removed due to no enrichment) from aggregated differential gene expression data across 6 experiments

(a - upregulated, b - downregulated genes). Y-axis shows the sample comparison for each dataset, with number of DEGs shown as a barplot at right. In the heatmap, warmer colors mean higher enrichment ratios (for complete list of odds ratios, see Supplementary Table 5).



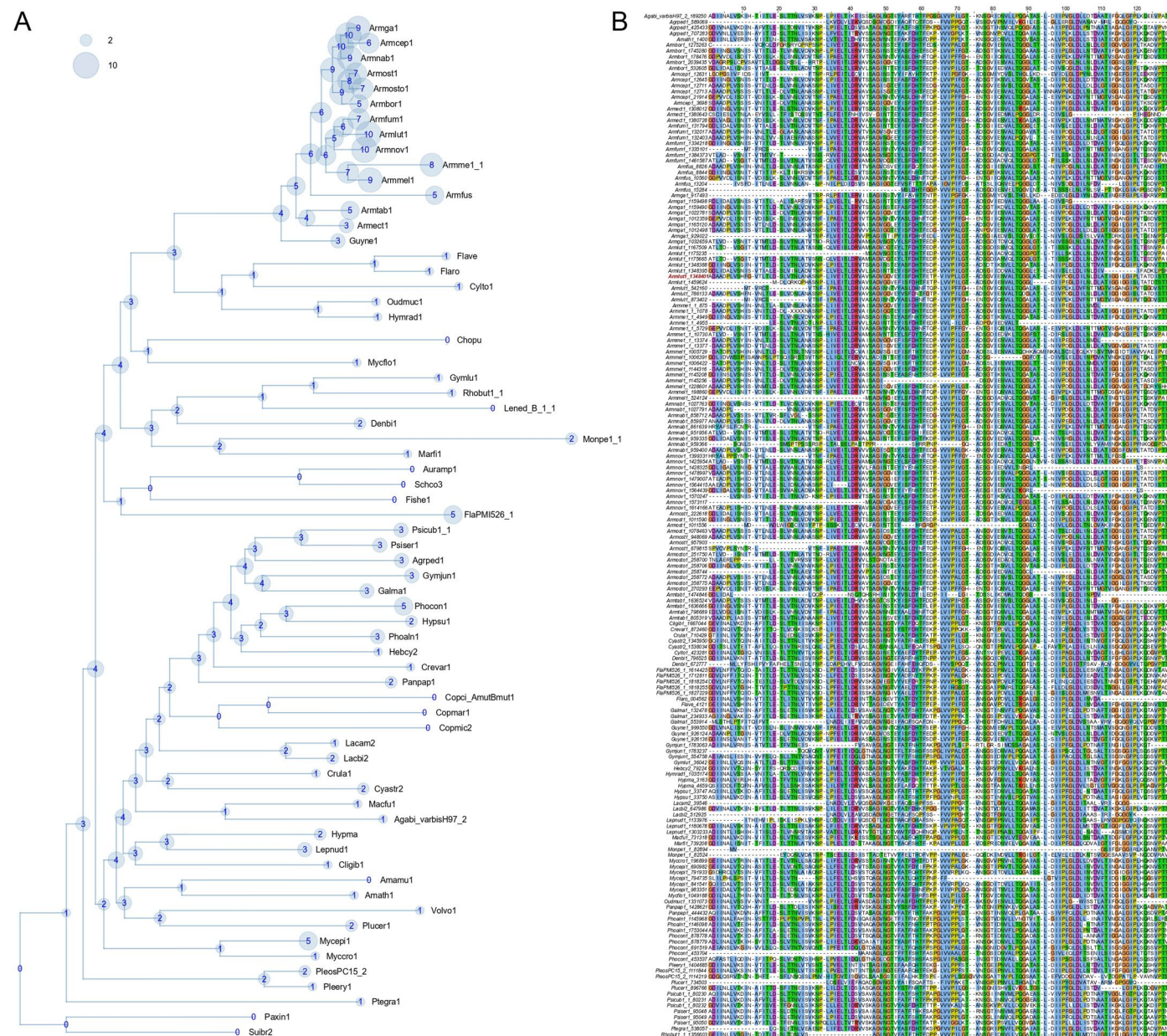
Extended Data Fig. 7 | Expression heatmaps in *A. luteobubalina*. Heatmaps showing gene expression along the time course in *A. luteobubalina* for gene families related to host immune suppression, oxidative stress, detoxification, and cytotoxicity. Warmer color depicts higher expression.

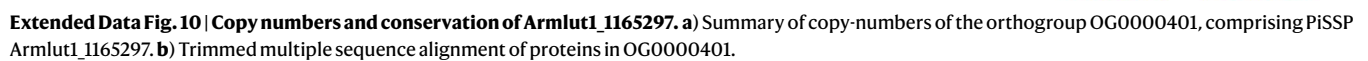


Homologs of 39 annotated SSPs – Upregulated in at least one time point

Extended Data Fig. 8 | Pathogenicity-induced SSP expression and conservation in *A. luteobubalina*. A) Heatmap shows $\log_2$ fold changes for annotated SSPs, upregulated in at least one-time point. Red shows higher and blue depicts lower logFC; followed by presence/absence matrix of homologs in

131 species (Dataset 2). X-axis shows ProteinIDs for both heatmap and presence/absence matrix. Y-axis shows sample comparisons in the heatmap; and species order in presence/absence matrix.





Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- ☐ ☒ The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- ☐ ☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- ☐ ☒ The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- ☒ ☐ A description of all covariates tested
- ☐ ☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- ☐ ☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- ☐ ☒ For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- ☒ ☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- ☒ ☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- ☒ ☐ Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection	InterPro Scan (v.5.46-81.0), OrthoFinder (v2.5.4, v2.3.1), MAFFT (v7.453), TrimAL, RAxML (v8.2.12), genoPlotR (v0.8.11), FigTree (v1.4.4), IQ-Tree v2.0.3, MMseqs2 v13.45111, STEM v1.3.11, Falcon (v1.8.8), pb-assembly (v0.0.2), falcon-kit (v1.2.3), pypeflow (v2.1.0), SMRTLink v5.0.1.9578 (A. ectypa and A. nabsnona), v5.1.0.26412 (A. tabescens and A. borealis), v6.0.0.47841 (A. mellea) or v7.0.1.66975 (A. novae-zelandiae and A. fumosa), finisherSC version 2.0 (A. ectypa and A. nabsnona) or 2.1 (A. fumosa), BBduk (https://sourceforge.net/projects/bbmap/), Trinity (ver. 2.3.2), CAZY database, JGI Mycocosm
Data analysis	COMPARE pipeline (available at https://github.com/laszlognag/COMPARE), R packages used for analysis (topGO, phytools, SuperSeq), custom codes used for the analyses are available at https://github.com/nehasahu486/Armillaria-phylogenomics/tree/main .

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

New genomic assemblies and annotation generated in this study are deposited under the 1000 Fungal Genome Project at JGI Mycocosm (<https://mycocosm.jgi.doe.gov/Armillaria/Armillaria.info.html>) and at DDBJ/EMBL/GenBank under the accession numbers PRJNA463936, PRJNA500536, PRJNA500837, PRJNA519860, PRJNA519861, PRJNA571622, PRJNA677793, PRJNA677794. New RNA-seq datasets used in this study are deposited in the NCBI's Gene Expression Omnibus (GEO) Archive at <https://www.ncbi.nlm.nih.gov/geo/>. Accession number for the in-plant assay between *A. luteobubalina* and *E. grandis* is PRJNA975488, or GSE233220. For the stem-invasion assay, the accession numbers are PRJNA972908 for *A. ostoyae* and PRJNA972989 for *A. borealis*. Phylogenetically validated gene trees and gene expression heatmaps for various gene families for the six RNA-Seq datasets used in this study can be found in the Figshare repository at https://figshare.com/articles/dataset/Gene_trees/22730534 and https://figshare.com/articles/figure/Gene_expression_heatmaps/22778477?file=40472333 respectively.

Human research participants

Policy information about [studies involving human research participants and Sex and Gender in Research](#).

Reporting on sex and gender	NA
Population characteristics	NA
Recruitment	NA
Ethics oversight	NA

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	In this paper we explored the biology of <i>Armillaria</i> species by: (i) phylogenomic comparisons of new and old <i>Armillaria</i> genomes with a diverse fungal dataset (ii) systematic detection of horizontally transferred genes in <i>Armillaria</i> and the Physalacriaceae family (iii) transcriptome profiling for both early and late stage infection in novel pathosystems (iv) experimental validation of two pathogenicity-induced small secreted proteins (PiSSPs)
Research sample	New genomes described in this study: <i>Armillaria nabsnona</i> CMW6904 v1.0, <i>Armillaria borealis</i> FPL87.14 v1.0, <i>Armillaria fumosa</i> CBS 122221 v1.0, <i>Armillaria luteobubalina</i> HWK02 v1.0, <i>Armillaria novae-zelandiae</i> 2840 v1.0, <i>Armillaria mellea</i> ELDO17 v1.0, <i>Armillaria tabescens</i> CCBAS 213 v1.0, <i>Armillaria ectypa</i> FPL83.16 v1.0 Species used for RNASeq in stem-invasion assay: <i>Armillaria ostoyae</i> (C18 - virulent and C2 - less-virulent strain), <i>Armillaria borealis</i> (A6 - virulent and A4 - less virulent strain) Species used for RNASeq in live in-plant assay: <i>Armillaria luteobubalina</i> and <i>Eucalyptus grandis</i>
Sampling strategy	Stem-invasion assay: High and low virulent isolates of conifer-specific <i>A. ostoyae</i> and <i>A. borealis</i> were allowed to invade freshly cut stem segments of Norway spruce and compared with vegetative mycelia collected from the jars without spruce stems as controls Live in-plant assay: <i>A. luteobubalina</i> colonizing <i>E. grandis</i> roots across five time points (pre-symbiosis, 24hrs, 48hrs, 1 week and 2 weeks) and compared these to non-symbiotic samples
Data collection	Samples were collected via microbiological methods:

Data collection	stem invasion assays: G.S. and B.I. live in-planta assay: J.P., J.W-B.
Timing and spatial scale	NA
Data exclusions	In the in-planta experiment, the JW18 sample was removed from the DEG analysis because it was an outlier, leaving us with three remaining replicates
Reproducibility	Three to four biological replicates were used in RNA-Seq experiments.
Randomization	NA
Blinding	Samples processed for RNA-Seq analyses were blinded after the initial sampling step.

Did the study involve field work? ☐ Yes ☒ No

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging