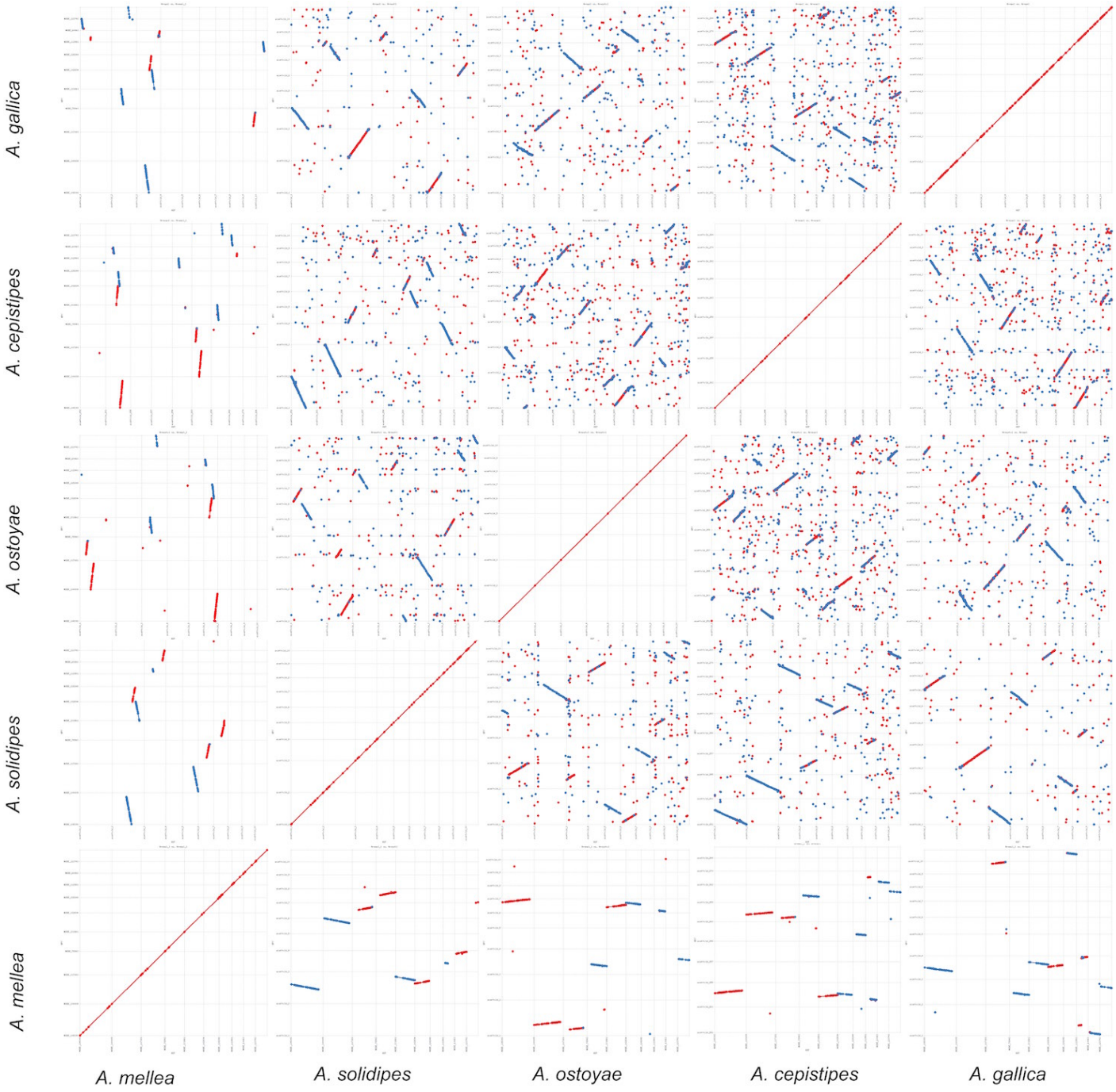
Supplementary Figures

for

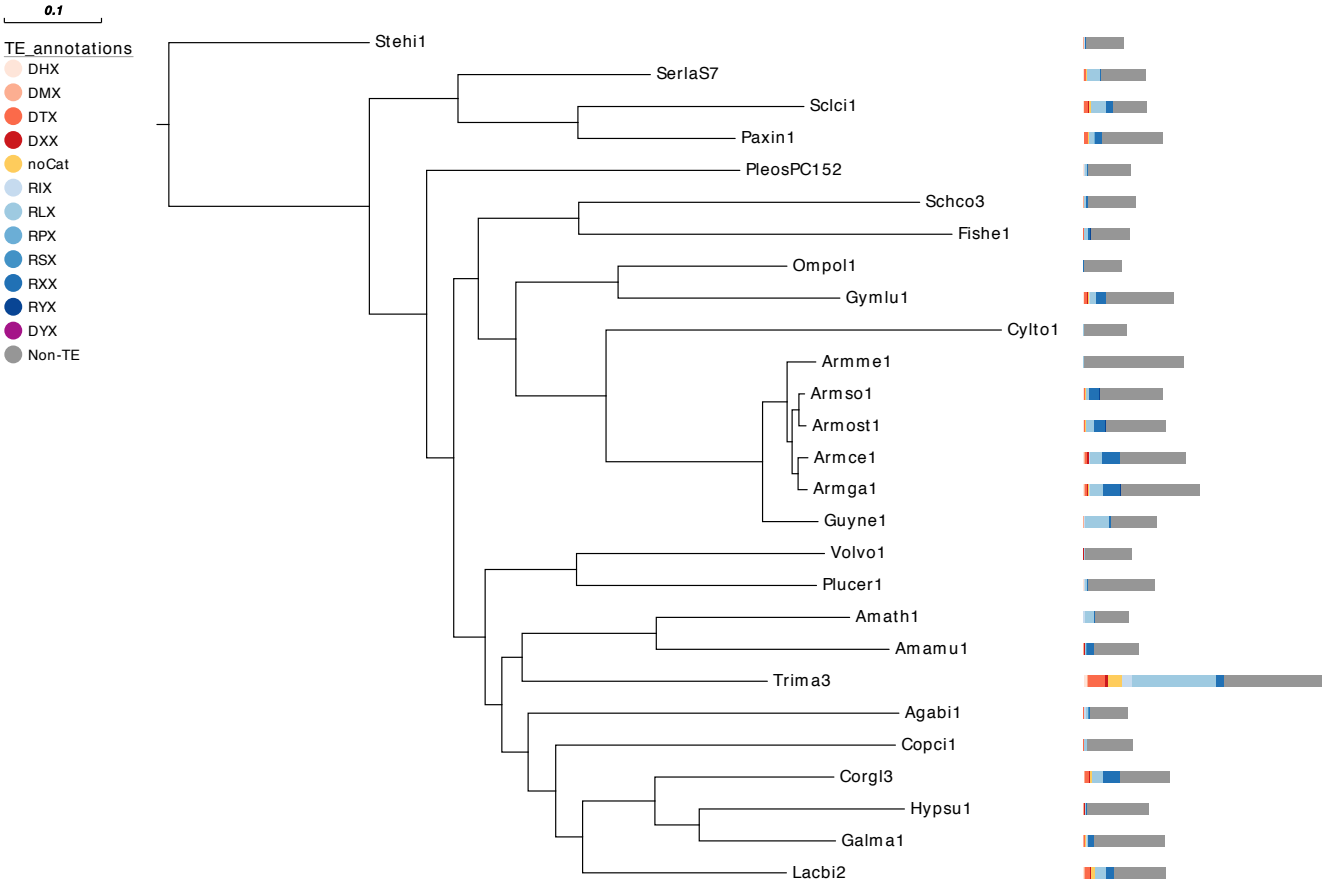
Genome expansion and lineage-specific genetic innovations in the world's largest organisms (*Armillaria*)

Gyorgy Sipos, Arun N. Prasanna, Mathias C. Walter, Eoin O'Connor, Balazs Balint, Krisztina Krizsan, Brigitta Kiss, Jacky Hess, Torda Varga, Jason Slot, Robert Riley, Bettina Boka, Daniel Rigling, Kerrie Barry, Juna Lee, Sirma Mihaltcheva, Kurt LaButti, Anna Lipzen, Rose Waldron, Nicola M. Moloney, Christoph Sperisen, Laszlo Kredics, Csaba Vagvolgyi, Andrea Patrignani, David Fitzpatrick, Istvan Nagy, Sean Doyle, James B. Anderson, Igor V. Grigoriev, Ulrich Güldener, Martin Münsterkötter, Laszlo G. Nagy

Contents:



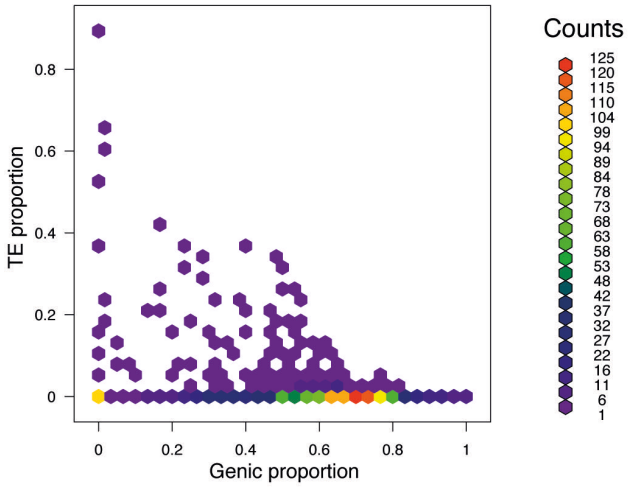
Supplementary Figure 1. Synteny analysis among 5 *Armillaria* species. Pairwise dot plots of the 10 largest scaffold in each genome showing extensive macro- and microsynteny among species. Genomes were aligned using nucmer.



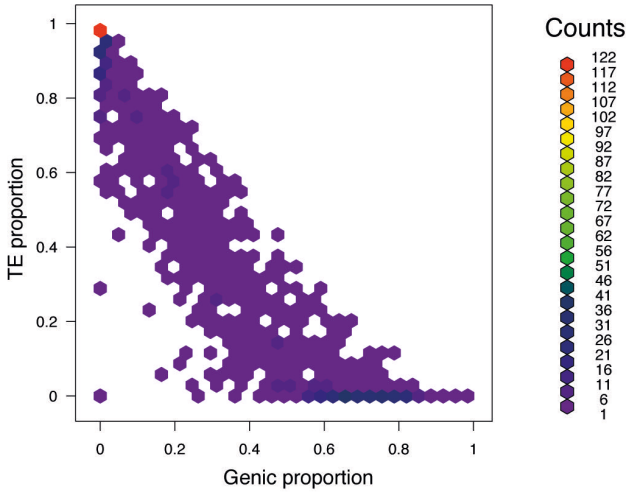
Supplementary Figure 2. Transposable element distribution in 27 Agaricomycetidae species broken down by TE category. Species tree annotated with TE proportions estimated using the REPET packages. TEs were classified according to Wicker et al. 2007 and are separated into RNA-based elements (blue) and DNA-based elements (red). DHX: Helitron, DMX: Maverick, DTX: TIR-transposon, DXX: unknown DNA-based element, noCat: unknown repetitive element, RIX: LINE elements, RLX: LTR elements, RPX: Penelope, RSX: SINE, RYX: DIRS. Transposable element content ranged from 0.9% in *Omphalotus olearius* to 59.2% in *Tricholoma matsutake* which constitute the smallest and the largest genome in the dataset, respectively. Generally, TE content increased with genome size, with the exception of the *A. mellea* genome which among the largest genomes with approximately 79 MB but only 0.94% of the sequence were annotated as TEs. This may in part be due to the removal of short contigs, since TEs often cluster on these in fragmented assemblies but closer inspection of the genome sequence showed evidence that the sequence may have been repeat masked prior to submission to JGI. We therefore excluded the *A. mellea* genome from subsequent analyses. It has to be noted that the accuracy of the proportions of TEs estimated across the different genomes likely varies, since the data comprise a set of heterogeneous sequencing and assembly strategies which may differ in their ability to resolve and hence assemble repeated sequences. As is typical for basidiomycete fungi, the most frequent classes of TEs identified were RNA-based elements (R--) with LTR elements (RLX) in particular found to be most abundant (Castanera et al. 2017). Genome expansion in the *A.* clade may have at least partially driven by expansions in TE repertoires, although gains in non-TE sequence were proportionally greater than increases in TE content. *Cylindrobasidium torrendii,* the closest outgroup is nearly devoid of TEs (1.1% overall TE content), while the genome of *Guyanagaster necrorhiza* has been invaded by LTR-type elements and increased in TE content to 37.7%. While the non-TE proportions of the genome between *C. torrendii* and *G. necrorhiza* are quite similar in size, with appr. 31 and 33 MB, respectively, further evolution of genome architecture in the *A.* clade is characterized by

expansion of non-TE sequence. In contrast TE content in the four *A.* species grew proportionally less (*A. cepistipes* 26 MB, 34.8% TE and *A. gallica* 28 MB, 32.0% TE, or decreased all together (*A. ostoyae*, 16 MB, 26.9% TE and *A. solidipes*12 MB, 20.7% TE) compared to the approx. 20 MB of TE sequence in *G. necrorhiza*. This trend for non-TE genome expansion mirrors that of the whole-genome phylogenomic analysis, which found large scale gene space expansions within the *Armillaria.*

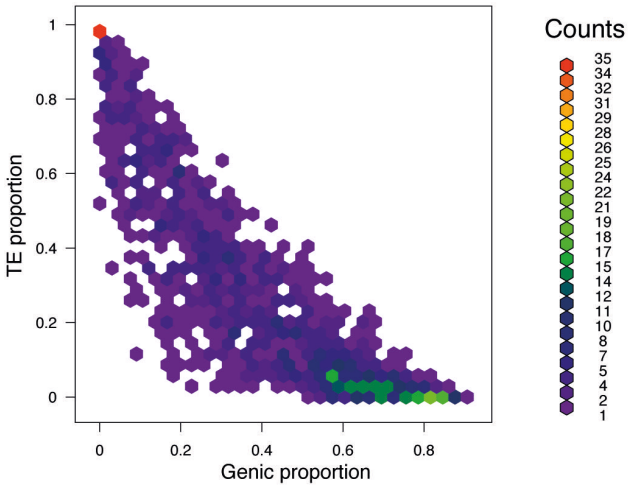
C. torrendii 50kb windows



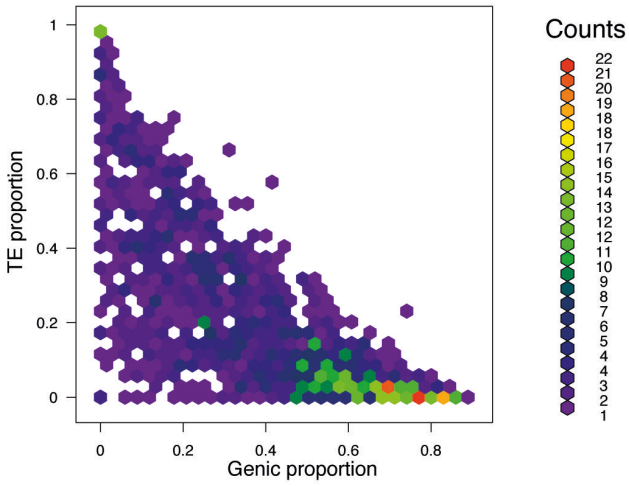
G. necrorhiza 50kb windows



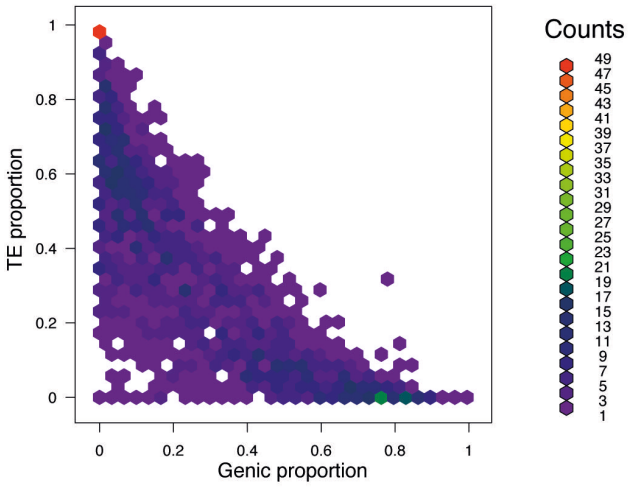
A. ostoyae 50kb windows



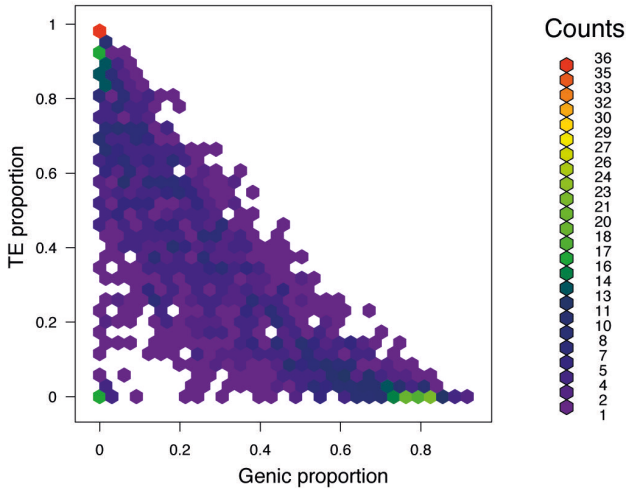
A. solidipes 50kb windows



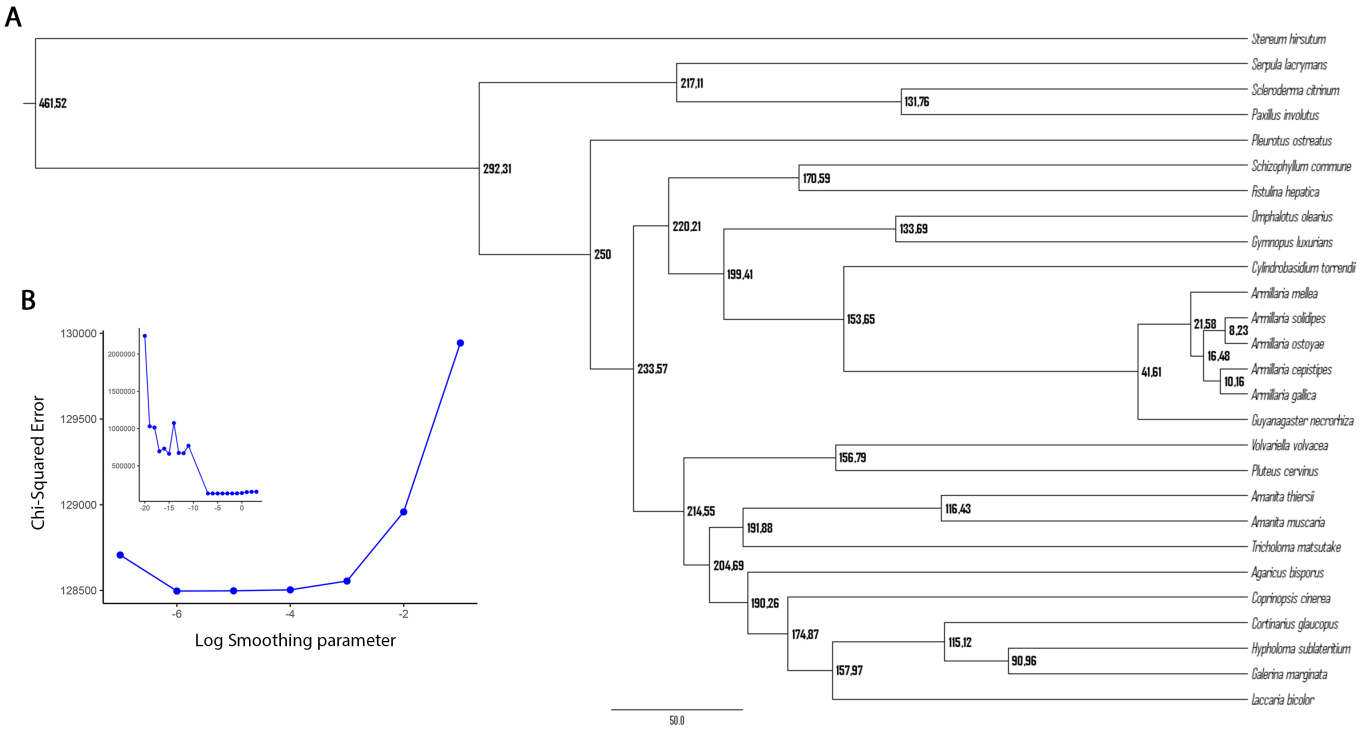
A. gallica 50kb windows



A. cepistipes 50kb windows

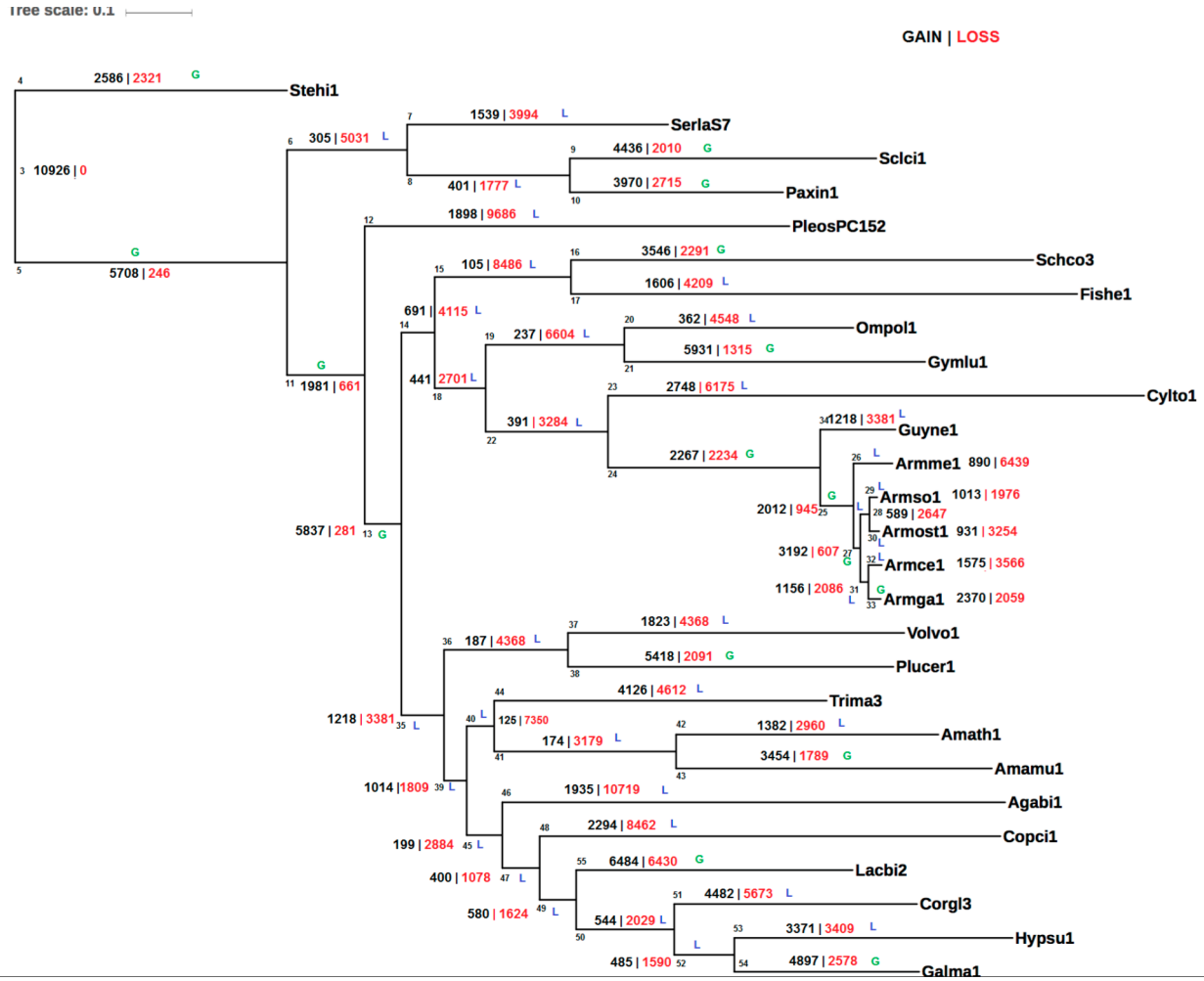


Supplementary Figure 3. Windowed analysis of clustering of TE content with respect to gene content. Each genome was partitioned into 50kb windows and binned according to proportions of TE and gene content. Shading of the cells indicates the number of windows in the bin. We did not find considerable structure in the outgroup *C. torrendii* likely due to the very low overall TE content. In *G. necrorhiza*, we found a moderate degree of partitioning of TE content and genic regions as indicated by the higher intensity of bins along the axes. Of the 13270 genes (after filtering) in *G. necrorhiza*, 3418 were found in chromosomal neighbourhoods devoid of TEs. This is considerably different from the *A.* species where TE content was well distributed all over the genome. We found only 700, 908, 816 and 1010 genes in 50kb windows without TEs in *A. cepistipes, A. gallica, A. ostoyae* and *A. solidipes*, respectively, indicating that vast majority of genes in these genomes is found in relatively close proximity to TEs. While we cannot exclude the existence of unassembled TE-only compartments, our data do not provide evidence for a "two-speed genome" found in some other plant pathogens[17], nor is there evidence for rampant TE expansion in the pathogenic *A.* species. Instead, the genomic clustering patterns observed suggest either continuing small scaled diversification of TEs over large parts of the genome, or a mixing and marbling of TE sequences throughout the genome mediated by structural changes such as segmental duplications, inversions or deletions. The latter mechanism would also provide a scenario for expansion of gene space and it is tempting to speculate that this may have facilitated the emergence of genome architectures found in extant *Armillaria.*

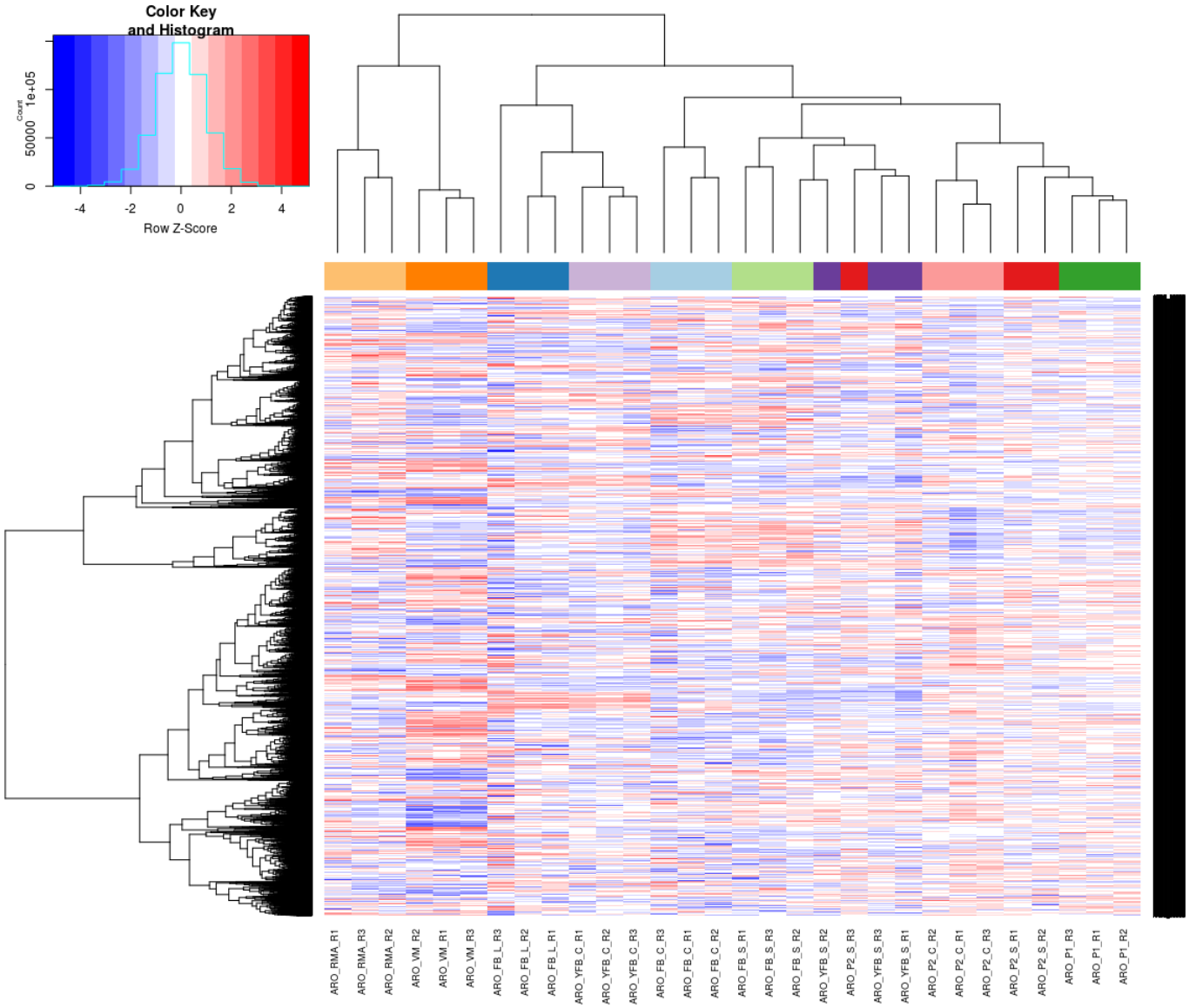


Supplementary Figure 4. Time calibrated phylogenomic tree (A) was inferred by penalized likelihood as implemented in r8s using the maximum likelihood tree and two fossils and one secondary calibration point. Smoothing parameters across 30 orders of magnitude, starting with 10^{-20} were tested with cross-validation analysis (B). Note that not all examined smoothing parameters are included in figure B, because in some cases and above 10^{3} smoothing parameter values were invalidated by errors during taxon prunings. The

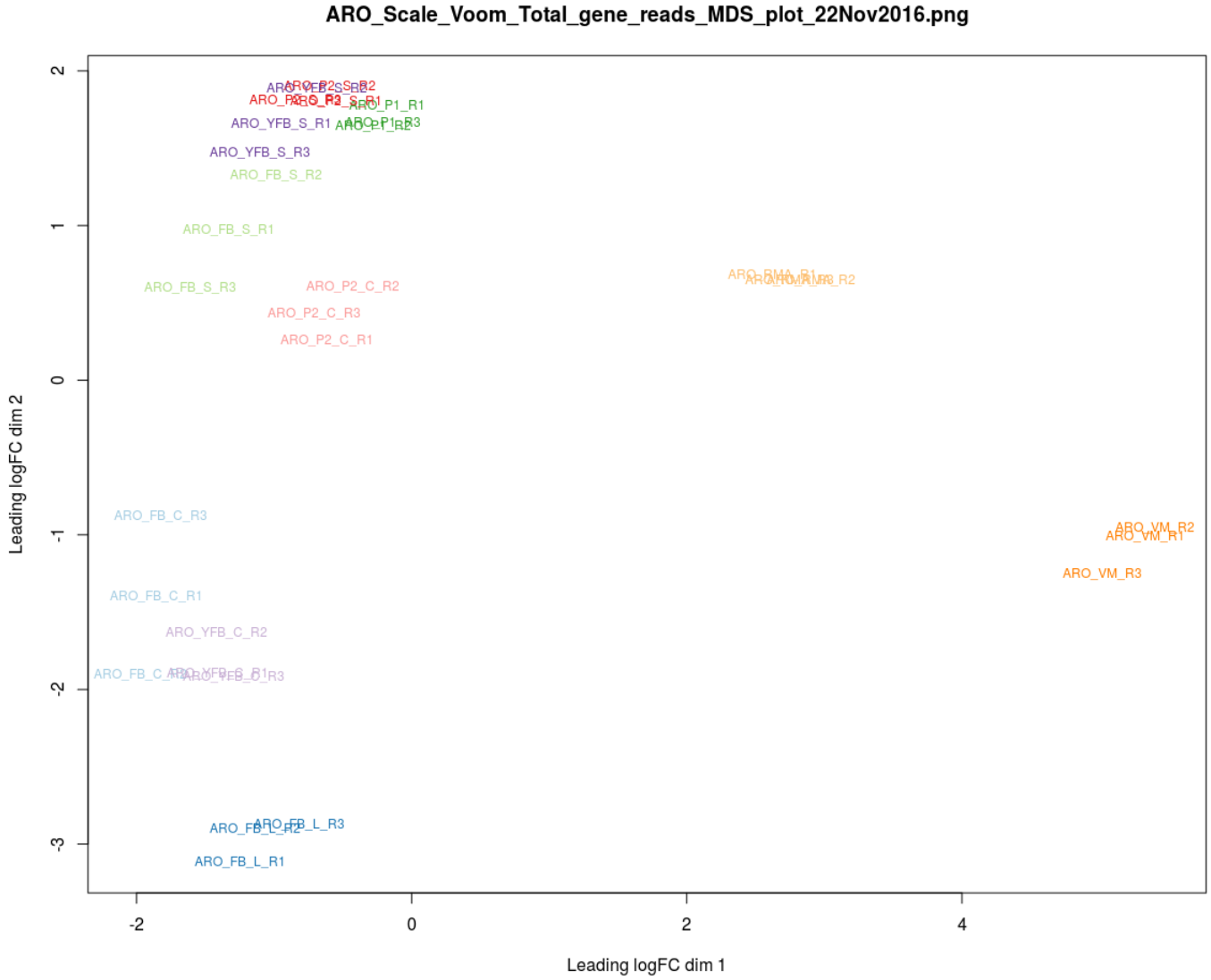
subplot on the upper left corner of figure B represents a broader range of smoothing parameter values than the main plot. In the final analysis a smoothing parameter of 10^{-6} was used, but calibration with smoothing parameters between 10^{-7}-10^{-3} resulted negligible differences in estimated ages. The values on the tree nodes are the estimated ages in million years. All nodes received maximal (100%) bootstrap support in ML bootstrapping.



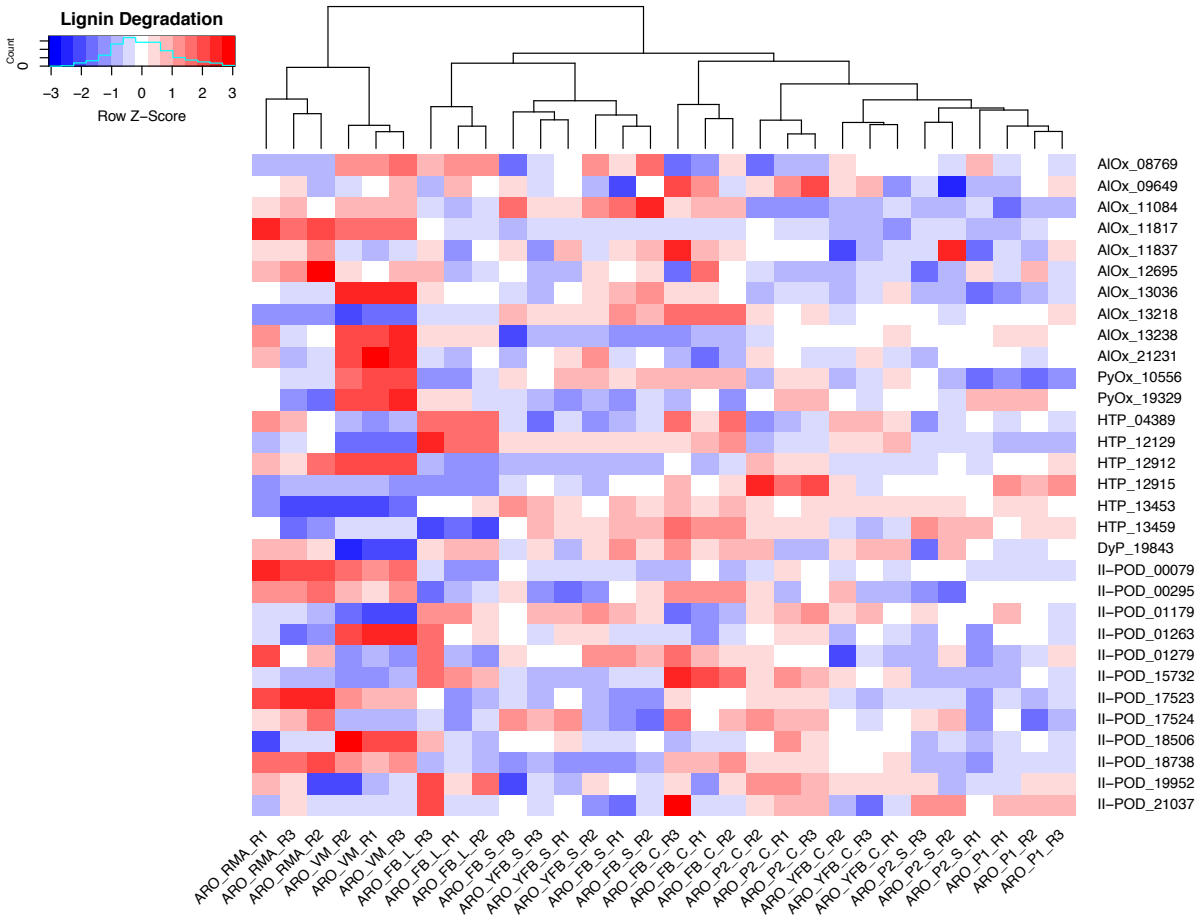
Supplementary Figure 5. Reconstructed gene duplication and loss histories summarized on the phylogenomic tree of 27 Agaricomycetes. Nodes are numbers adjacent to the branches and the inferred number of duplications (black) and losses (red) are shown for each branch.



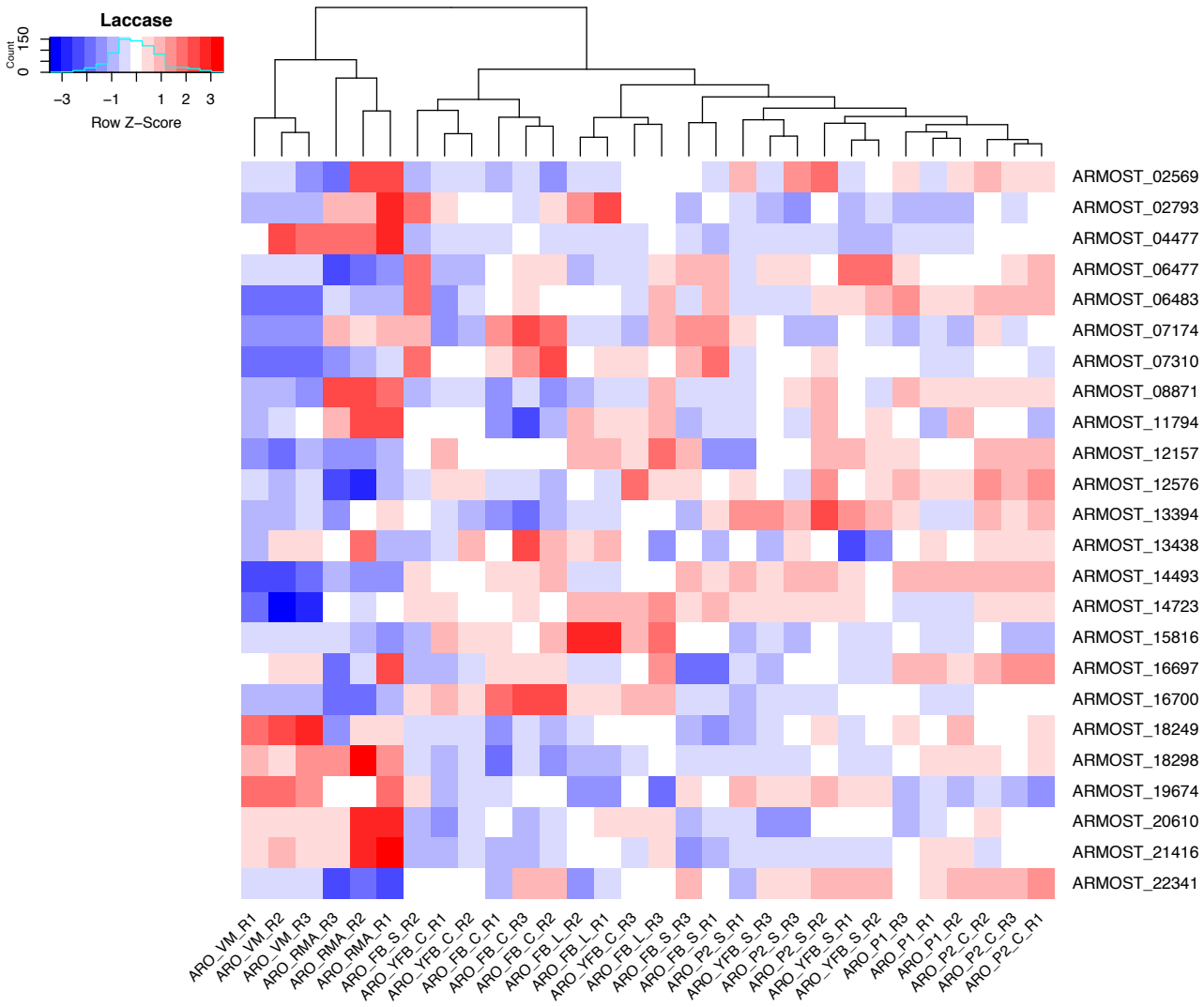
Supplementary Figure 6. Global gene expression patterns across three replicates of ten developmental stages of *A. ostoyae* C18 inferred using RNA-Seq. Abbreviations are as follows: ARO_RMA – rhizomorph, ARO_VM – vegetative mycelium, ARO_FB_L – mature fruiting body gills, ARO_FB_S – mature fruiting body stipes, ARO_FB_C – mature fruiting body caps, ARO_YFB_S – young fruiting body stipe, ARO_YFB_C – young fruiting body cap (including lamellae), ARO_P2_S – stage II primordium stipe, ARO_P2_C – stage II primordium cap, ARO_P1 – stage I primordium.



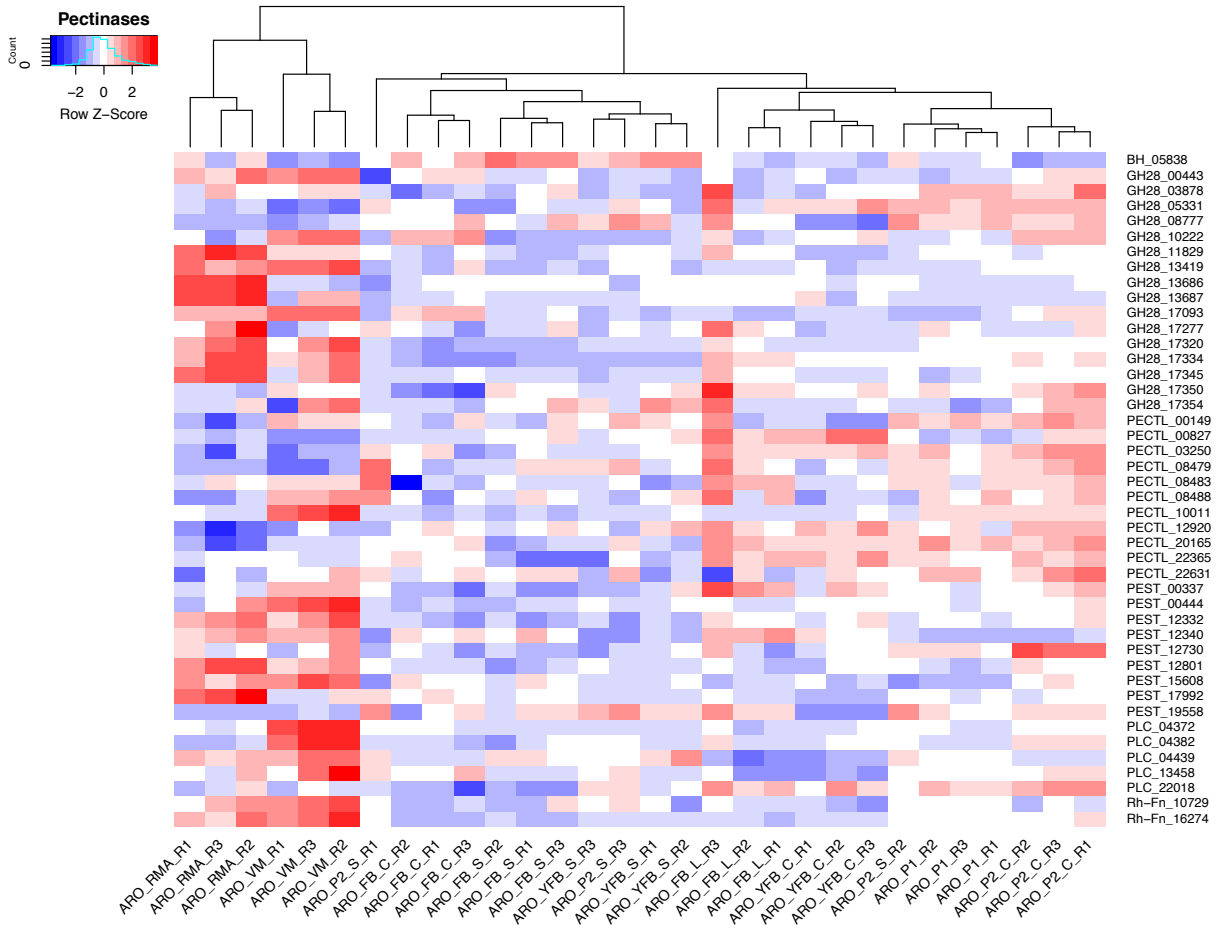
Supplementary Figure 7. MDS plot demonstrating the grouping of triplicate samples for 10 developmental stages and tissue types of *A. ostoyae* based on global gene expression patterns. Abbreviations are as follows: ARO_RMA – rhizomorph, ARO_VM – vegetative mycelium, ARO_FB_L – mature fruiting body gills, ARO_FB_S – mature fruiting body stipes, ARO_FB_C – mature fruiting body caps, ARO_YFB_S – young fruiting body stipe, ARO_YFB_C – young fruiting body cap (including lamellae), ARO_P2_S – stage II primordium stipe, ARO_P2_C – stage II primordium cap, ARO_P1 – stage I primordium.



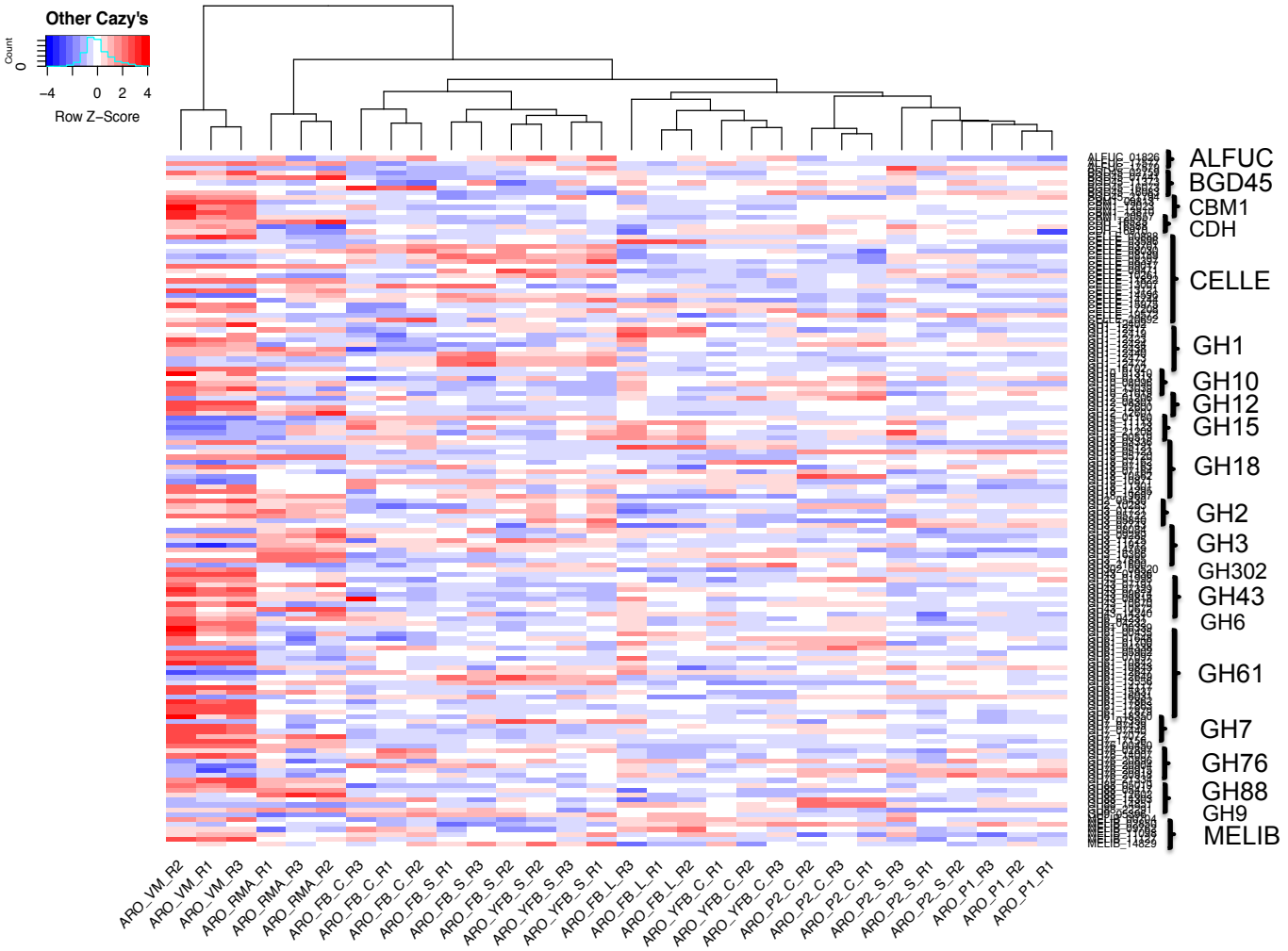
Supplementary Figure 8. Heatmap of ligninolytic gene expression patterns in *A. ostoyae* C18. Abbreviations are as follows: ARO_RMA – rhizomorph, ARO_VM – vegetative mycelium, ARO_FB_L – mature fruiting body gills, ARO_FB_S – mature fruiting body stipes, ARO_FB_C – mature fruiting body caps, ARO_YFB_S – young fruiting body stipe, ARO_YFB_C – young fruiting body cap (including lamellae), ARO_P2_S – stage II primordium stipe, ARO_P2_C – stage II primordium cap, ARO_P1 – stage I primordium. On the right, gene abbreviations AlOx, PyOx, HTP, Dyp and POD stand for aryl alcohol oxidase, pyranose oxidase, heme-thiolate peroxidase, dye decolorizing peroxidase and Class-II peroxidase, respectively. (ARMOST_XXXXX is replaced with CLASS_XXXXX)



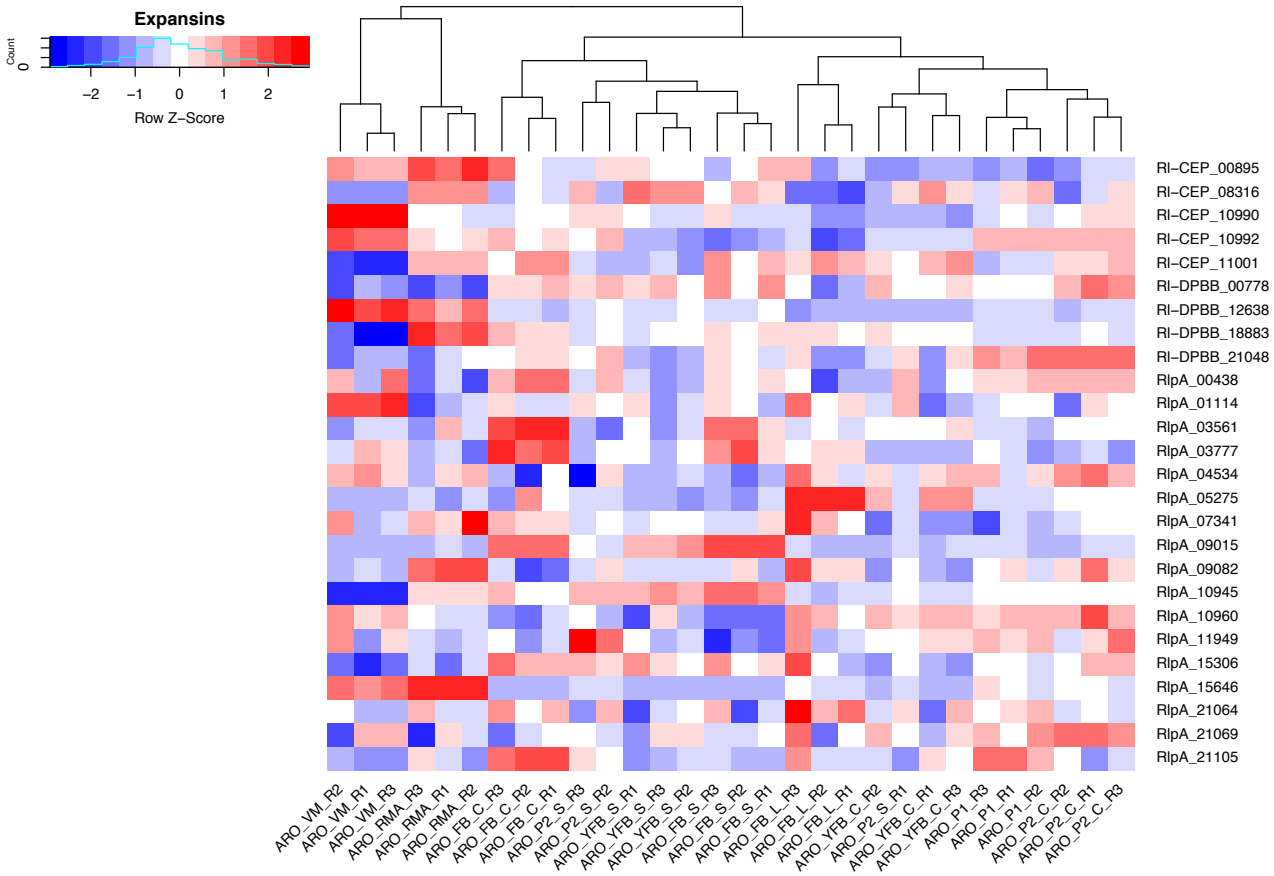
Supplementary Figure 9. Heatmap of laccase gene expression patterns in *A. ostoyae* C18. Abbreviations are as follows: ARO_RMA – rhizomorph, ARO_VM – vegetative mycelium, ARO_FB_L – mature fruiting body gills, ARO_FB_S – mature fruiting body stipes, ARO_FB_C – mature fruiting body caps, ARO_YFB_S – young fruiting body stipe, ARO_YFB_C – young fruiting body cap (including lamellae), ARO_P2_S – stage II primordium stipe, ARO_P2_C – stage II primordium cap, ARO_P1 – stage I primordium.



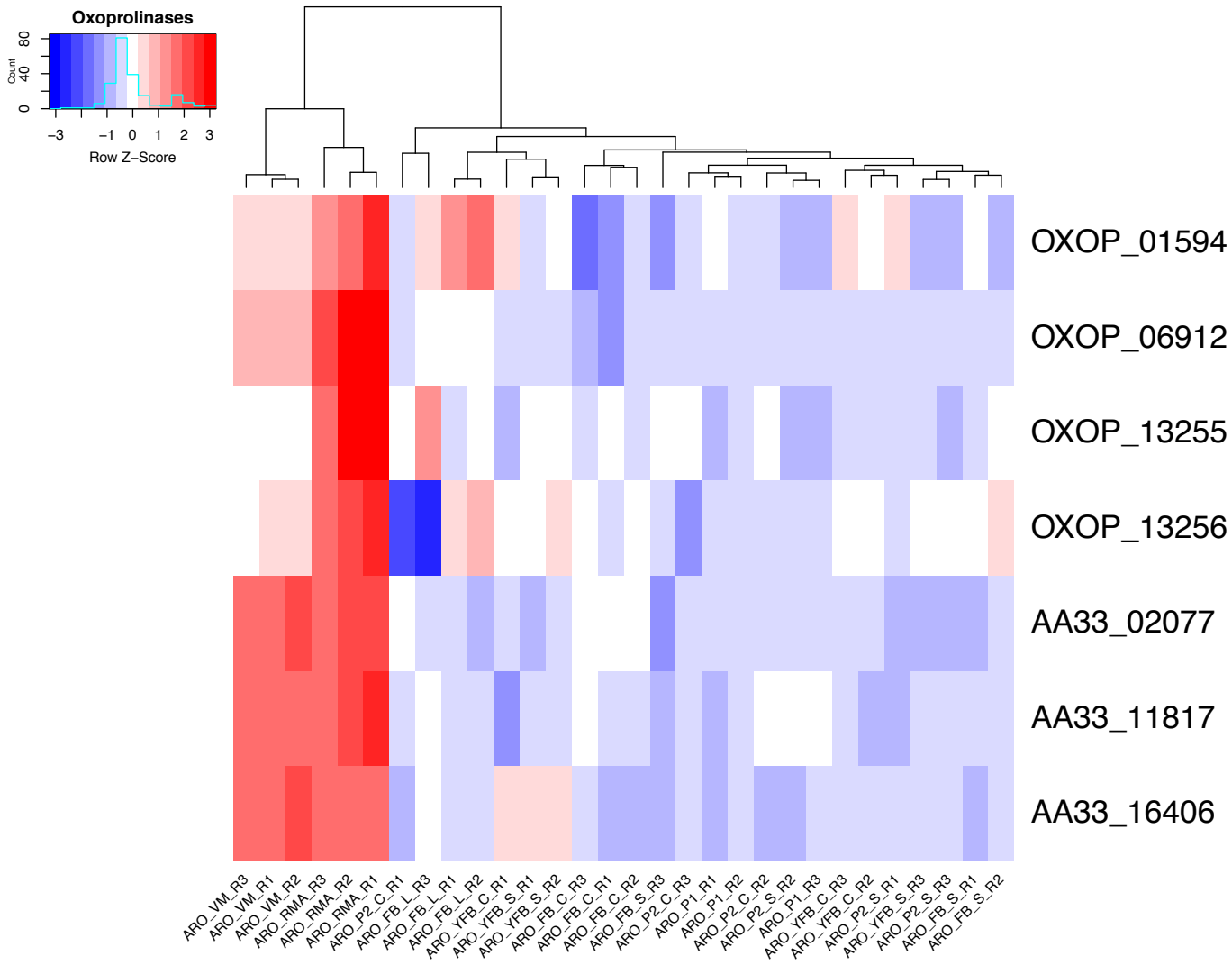
Supplementary Figure 10. Heatmap of pectinolytic CAZy gene expression patterns in *A. ostoyae* C18. Abbreviations are as follows: ARO_RMA – rhizomorph, ARO_VM – vegetative mycelium, ARO_FB_L – mature fruiting body gills, ARO_FB_S – mature fruiting body stipes, ARO_FB_C – mature fruiting body caps, ARO_YFB_S – young fruiting body stipe, ARO_YFB_C – young fruiting body cap (including lamellae), ARO_P2_S – stage II primordium stipe, ARO_P2_C – stage II primordium cap, ARO_P1 – stage I primordium. On the right, gene abbreviations BH, GH28, PECTL, PEST, PLC and Rh-Fn stand for beta-helix, pectate lyase, pectin esterase, Pec_Lyase_C and RhgB_N/fn3_3 respectively. (ARMOST_XXXXX is replaced with CLASS_XXXXX)



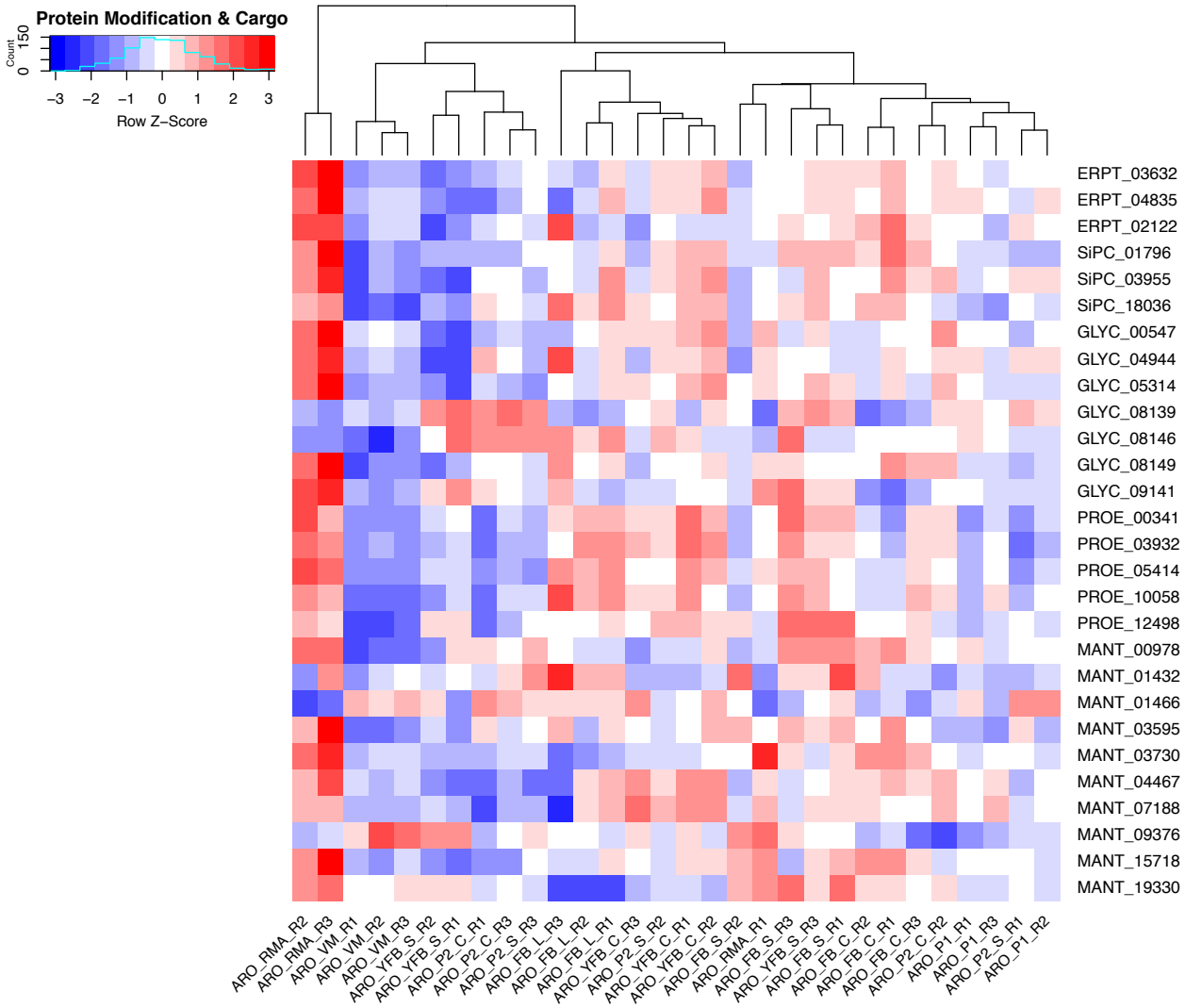
Supplementary Figure 11. Heatmap of (hemi-)cellulolytic CAZy expression patterns in *A. ostoyae* C18. Abbreviations are as follows: ARO_RMA – rhizomorph, ARO_VM – vegetative mycelium, ARO_FB_L – mature fruiting body gills, ARO_FB_S – mature fruiting body stipes, ARO_FB_C – mature fruiting body caps, ARO_YFB_S – young fruiting body stipe, ARO_YFB_C – young fruiting body cap (including lamellae), ARO_P2_S – stage II primordium stipe, ARO_P2_C – stage II primordium cap, ARO_P1 – stage I primordium. On the right, following are the gene abbreviations: ALFUC-alpha-L-fucosidase; BGD45-beta-galactosidase jelly roll domain; CBM1-carbohydrate binding module1; CDH-cellobiose dehydrogenase, CELLE – cellulase; GH1-Glyco_hydro_1; GH10-Glyco_hydro_10; GH12-Glyco_hydro_12; GH15-Glyco_hydro_15; GH18-Glyco_hydro_18; GH2-Glyco_hydro_2; GH3-Glyco_hydro_3; GH302-Glyco_hydr_30_2; GH43-Glyco_hydro_43; GH6-Glyco_hydro_6; GH61-Glyco_hydro_61; GH7-Glyco_hydro_7; GH76-Glyco_hydro_76; GH88-Glyco_hydro_88; GH9-Glyco_hydro_9; MELIB-melibiase_2. (ARMOST_XXXXX is replaced with CLASS_XXXXX)



Supplementary Figure 12. Heatmap of cerato-platanin (CEP) and expansin-like (DPBB, RLPA) gene expression patterns in *A. ostoyae* C18. Abbreviations are as follows: ARO_RMA – rhizomorph, ARO_VM – vegetative mycelium, ARO_FB_L – mature fruiting body gills, ARO_FB_S – mature fruiting body stipes, ARO_FB_C – mature fruiting body caps, ARO_YFB_S – young fruiting body stipe, ARO_YFB_C – young fruiting body cap (including lamellae), ARO_P2_S – stage II primordium stipe, ARO_P2_C – stage II primordium cap, ARO_P1 – stage I primordium. On the right gene abbreviation RI-CEP, RI-DPBB and RlpA stand for RlpA like ceratoplatanins, RlpA like-DPBB domain and rare lipoprotein A. (ARMOST_XXXXX is replaced with CLASS_XXXXX)

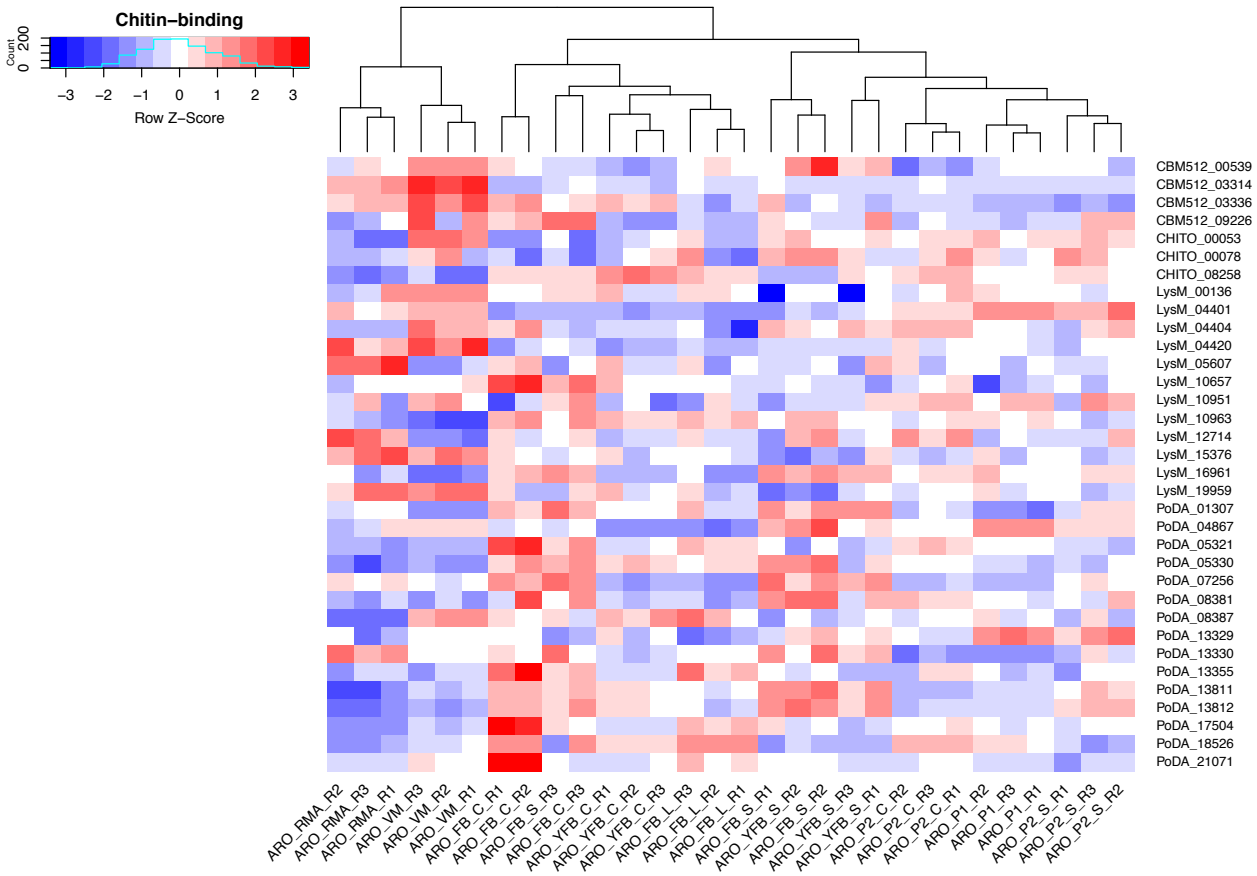


Supplementary Figure 13. Upregulation of four *A.*-specific 5-oxoprolinase (ARMOST_01594, ARMOST_06912, ARMOST_13255, ARMOST_13256) genes, representing a hydrolase family acting on carbon-nitrogen bonds of cyclic amides. Note that oxoprolinases are enriched in *A.* genomes (see pfam enrichment analysis), and form *A.*-specific clusters. 5-oxoprolinases eliminate 5-oxoproline (or pyroglutamic acid), which is toxic to the cell and accumulates as a byproduct of the pyroglutamyl cycle when cytoplasmic glutathione is exposed to elevated levels of H_2O_2. We hypothesize that H_2O_2 originate from the activity of intracellular alcohol oxidase genes (AA3_3, ARMOST_02077, ARMOST_11817, ARMOST_16406), of which three were found up-regulated in vegetative mycelia and rhizomorphs. Thus, 5-oxoprolinase upregulation probably marks a defense against intracellular H_2O_2 generated by alcohol oxidases, which can be tolerated by vegetative mycelia, but not rhizomorphs. It is tempting to speculate that the 3-dimensional organization of rhizomorphs hampers the passive diffusion of H_2O_2 out of the cell, requiring an additional mechanism to mitigate its effects. On the right, gene abbreviation for OXOP and AA33 stand for 5-oxoprolinase and alcohol oxidase (AA3_3), respectively. (ARMOST_XXXXX is replaced with CLASS_XXXXX)

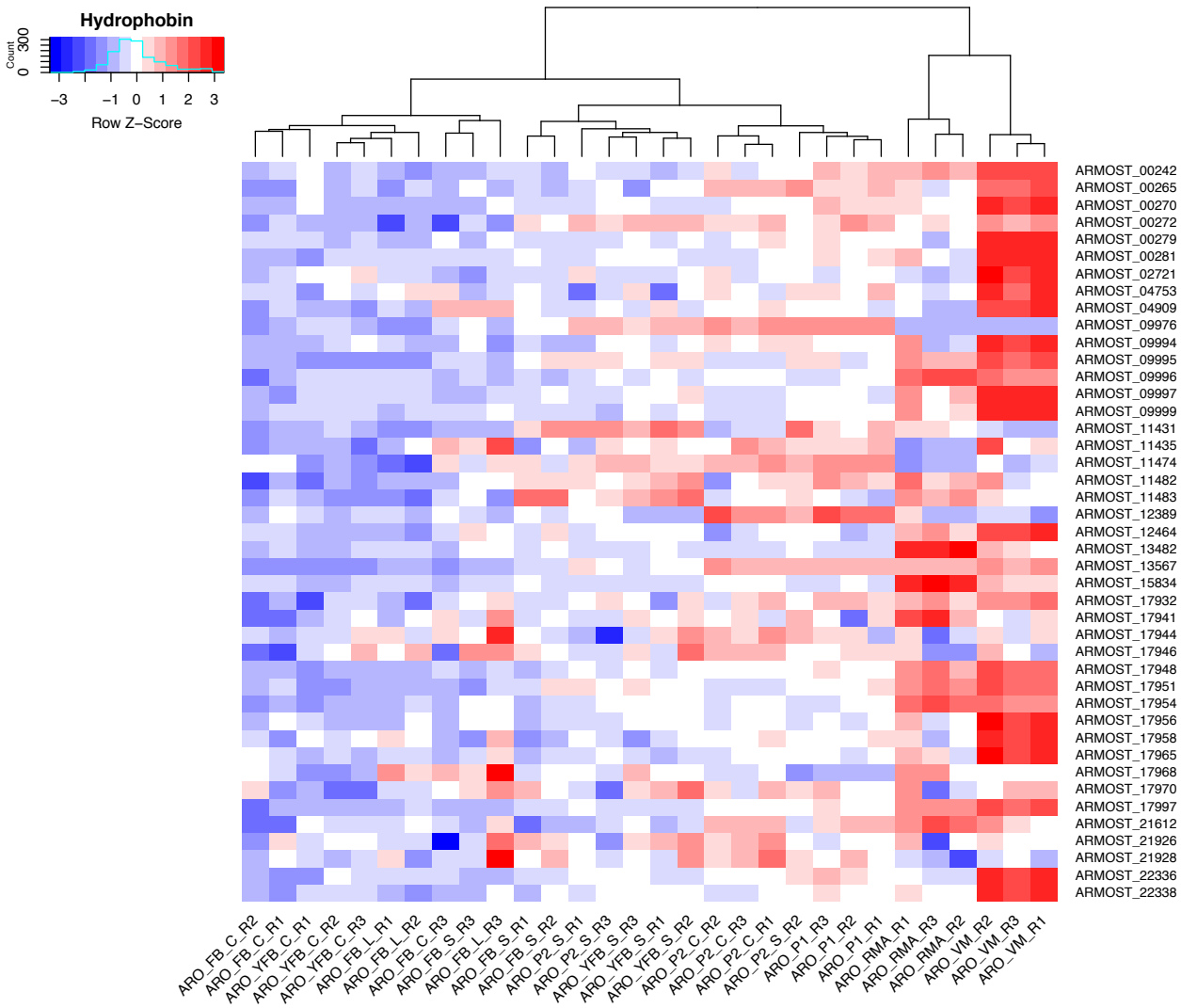


Supplementary Figure 14. The up-regulation of the genes encoding several components of the ER protein posttranslational import system (ARMOST_03632, ARMOST_04835, ARMOST_02122), including the signal peptidase complex (ARMOST_01796, ARMOST_03955, ARMOST_18036) and 10 mannosyl-transferases (ARMOST_00978, ARMOST_01432, ARMOST_01466, ARMOST_03595, ARMOST_03730, ARMOST_04467, ARMOST_07188, ARMOST_09376, ARMOST_15718, ARMOST_19330) of the N- and O-glycosylation machineries (ARMOST_00547, ARMOST_04944, ARMOST_05314, ARMOST_08139, ARMOST_08146, ARMOST_08149, ARMOST_09141), and subunits of the proteasome (ARMOST_00341, ARMOST_03932, ARMOST_05414, ARMOST_10058, ARMOST_12498) involved in the degradation of the unfolded, exported proteins on the ER surface, may indicate an intensified biogenesis and cargo of extracellular proteins along the secretory pathway. Although an upregulation of these genes was low (FC<4), it was statistically significant, suggesting that these functionalities are revved up in rhizomorphs.

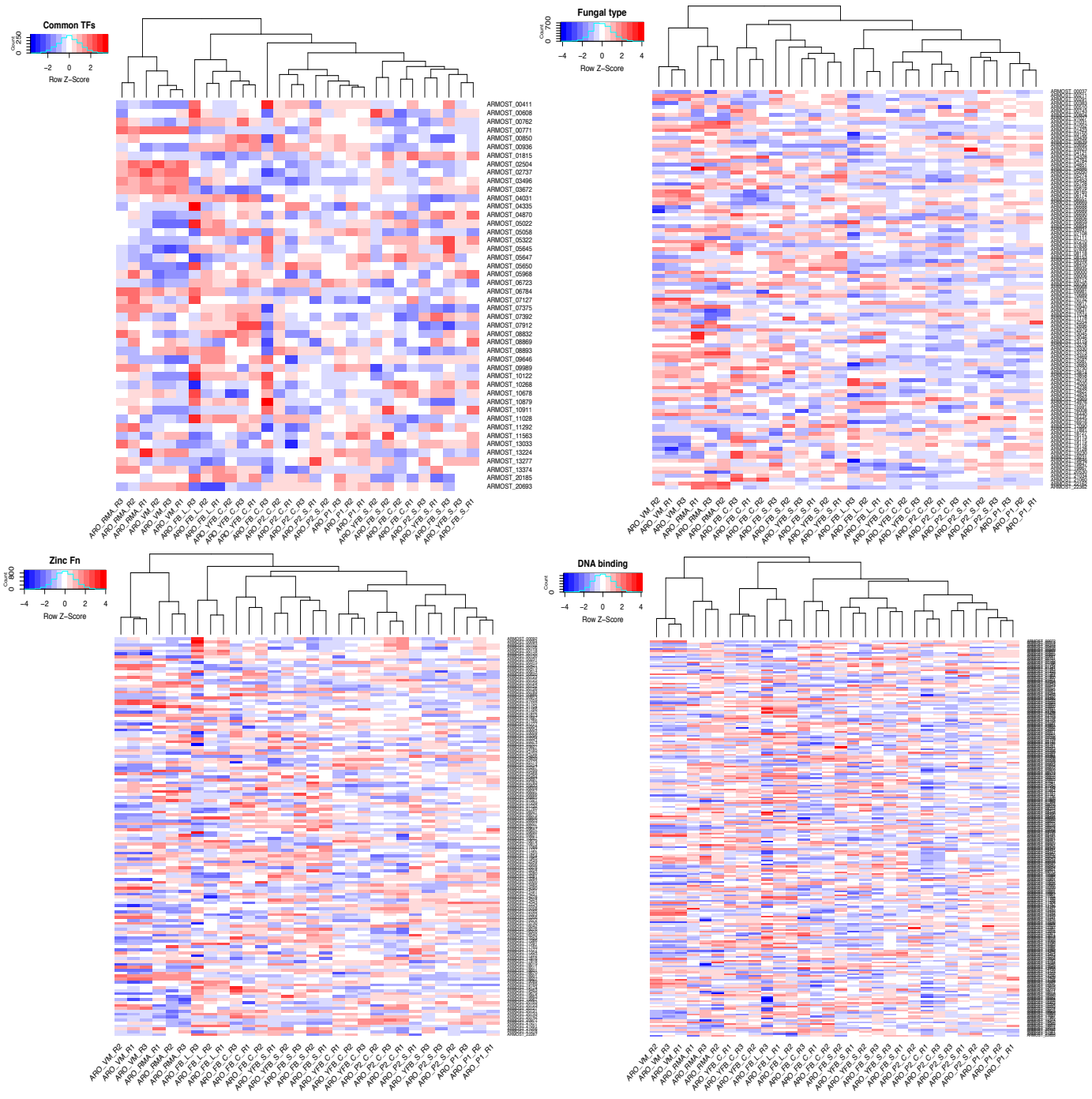
On the right, gene abbreviations ERPT, SiPC, GLYC, PROE and MANT stand for ER protein translocation system, Signal peptidase complex, N and O glycosylation machineries, subunits of Proteosome and Mannosyl transferases. (ARMOST_XXXXX is replaced with CLASS_XXXXX)



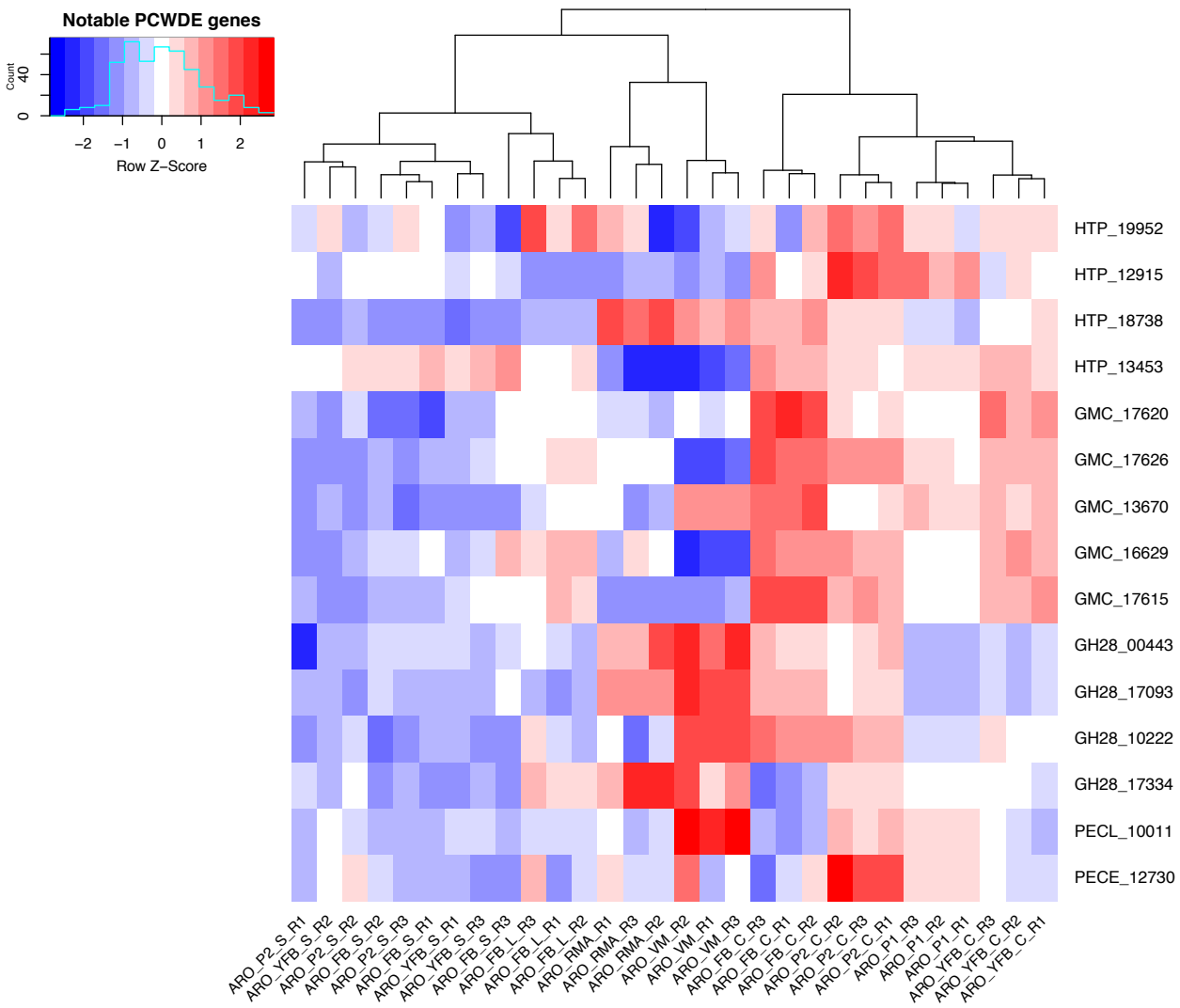
Supplementary Figure 15. Heatmap of chitin binding and metabolism-related gene expression patterns in *A. ostoyae* C18. Abbreviations are as follows: ARO_RMA – rhizomorph, ARO_VM – vegetative mycelium, ARO_FB_L – mature fruiting body gills, ARO_FB_S – mature fruiting body stipes, ARO_FB_C – mature fruiting body caps, ARO_YFB_S – young fruiting body stipe, ARO_YFB_C – young fruiting body cap (including lamellae), ARO_P2_S – stage II primordium stipe, ARO_P2_C – stage II primordium cap, ARO_P1 – stage I primordium. On the right, gene family abbreviations CBM512, CHITO, LysM and PoDA stand for carbohydrate binding module5/12, fungal chitosanase (GH75), LysM motif and polysaccharide deacethylase. (ARMOST_XXXXX is replaced with CLASS_XXXXX)



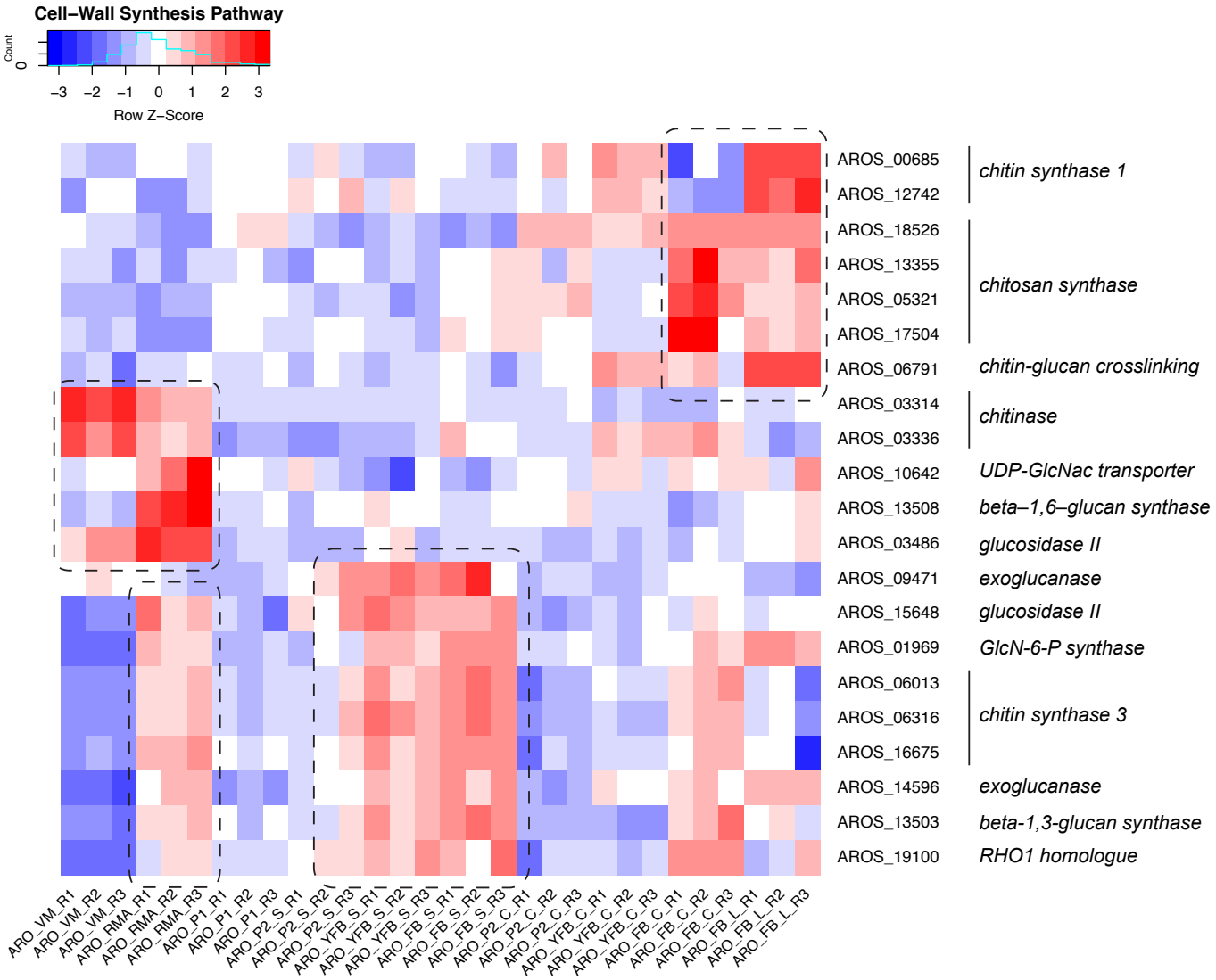
Supplementary Figure 16. Heatmap of hydrophobin gene expression patterns in *A. ostoyae* C18. Abbreviations are as follows: ARO_RMA – rhizomorph, ARO_VM – vegetative mycelium, ARO_FB_L – mature fruiting body gills, ARO_FB_S – mature fruiting body stipes, ARO_FB_C – mature fruiting body caps, ARO_YFB_S – young fruiting body stipe, ARO_YFB_C – young fruiting body cap (including lamellae), ARO_P2_S – stage II primordium stipe, ARO_P2_C – stage II primordium cap, ARO_P1 – stage I primordium.



Supplementary Figure 17. Heatmap of transcription factor gene expression patterns in *A. ostoyae* C18 broken down by major TF category: General transcription factors (TFs like TFIID, TATA box binding factors etc.,, upper left), Fungal specific transcription factors (Fungal type like Fungal transcription factor) , Zinc finger type transcription factors and DNA binding factors (Fork head domain, ARID DNA binding etc.,). Abbreviations are as follows: ARO_RMA – rhizomorph, ARO_VM – vegetative mycelium, ARO_FB_L – mature fruiting body gills, ARO_FB_S – mature fruiting body stipes, ARO_FB_C – mature fruiting body caps, ARO_YFB_S – young fruiting body stipe, ARO_YFB_C – young fruiting body cap (including lamellae), ARO_P2_S – stage II primordium stipe, ARO_P2_C – stage II primordium cap, ARO_P1 – stage I primordium.



Supplementary Figure 18. We detected several HTPs and GMC oxidoreductases upregulated in fruiting bodies. ARMOST_13453 (HTP) and ARMOST_07644 (GMC oxidoreductase) show uniformly high transcript levels in fruiting body samples, whereas 3 further HTPs and 5 GMC oxidoreductases show high expression in caps, but significantly lower in stipes. Certain GMC oxidoreductases were upregulated in fruiting body tissues of *Flammulina velutipes* too[73], which supports a potential role of HTPs and GMC oxidoreductases in the multicellular organization of rhizomorphs and fruiting bodies. Similarly, several pectin degradation-related genes, including multiple GH28 exo- and endopolygalacturonases, a pectate lyase and pectinesterase showed an expression maximum in cap tissues of stage II primordia and to a lesser extent in young fruiting body caps. Their upregulation in fruiting body tissues (and rhizomorphs) is remarkable as these genes are generally linked to PCW degradation, whereas to our best knowledge, they have not been implicated in fruiting body development so far. On the right, gene abbreviations HTP, GMC, GH28, PECL and PECE stand for heme-thiolate peroxidase, GMC oxidoreductase, GH28 exopolygalacturonases, pectate lyase and pectinesterase, respectively. (ARMOST_XXXXX is replaced with CLASS_XXXXX)



Supplementary Figure 19. Genes of the cell wall biosynthesis and remodeling pathways (including its regulator) exhibit tissue-specific expression patterns. Two homologs of the *Saccharomyces cerevisiae* chitin synthase 1 (CHS1), four chitin deacetylases or chitosan synthases (ARMOST_ 05321, ARMOST_13355, ARMOST_17504, ARMOST_18526, homologues of CDA1) and one gene involved in crosslinking glucan and chitin (ARMOST_06791, homolog of CRH1) are predominantly upregulated in cap and gill tissues. Two chitinase genes (ARMOST_03314 and ARMOST_03336, homologs of CTS1 and CTS2 respectively) showed highest expression in vegetative mycelia. On the other hand, genes showing peak expression in rhizomorphs included a homologue (ARMOST_03486) of the yeast glucosidase II (ROT2), critical in the elongation of N-glycans, a beta–1,6–glucan synthase (ARMOST_13508, homolog of KRE6) and a gene encoding transporter of UDP-GlcNac (ARMOST_10642, homolog of YEA4), which is the donor of GlcNac residues for glycosylated proteins and cell surface structures. In contrast to CHS1 homologs, three out of four yeast CHS3 homologs (ARMOST_06013, ARMOST_06316, ARMOST_16675), a second homolog of glucosidase II (ARMOST_15648, ROT2), a beta-1,3-glucan synthase (ARMOST_13503, homolog of FKS1), its RHO1 regulatory factor (ARMOST_19100), an exoglucanase (ARMOST_14596, homolog of EXG1) and a gene which catalyzes the formation of glucosamine-6-P (the first step of the chitin biosynthetic pathway,

ARMOST_01969, homolog of GFA1) show concomitant upregulation in stipes and rhizomorphs. These latter genes show modest upregulation in cap tissues of mature fruiting bodies too. Taken together, these data suggest that genes involved in chitin synthesis and remodeling are differentially used by *A. ostoyae* in a development- and tissue-specific manner.