Ecological generalism drives hyperdiversity of secondary metabolite gene clusters in xylarialean endophytes

SUPPORTING INFORMATION

Isolate selection and verification. Endophytic isolates are maintained as an axenic voucher in sterile water at the Robert L. Gilbertson Mycological Herbarium at the University of Arizona (ARIZ). Cultures of named taxa were obtained from the Westerdijk Fungal Biodiversity Institute (Netherlands) or from Dr. Yu-Ming Ju. Prior to genome and transcriptome sequencing, fungi were grown on 2% malt extract agar (MEA) to verify morphology and obtain tissue for a preliminary DNA extraction to verify isolate identity. Briefly, DNA was extracted using Extract n Amp (Sigma) following (U'Ren, 2016). For each isolate the ITS-LSU nrDNA region was PCR amplified using the primer pair ITS1F/LR3 and Sanger sequenced for each isolate as described by (U'Ren *et al.*, 2012). Sequences were edited in Sequencher v5.4.6 (Gene Codes Corporation, Ann Arbor, MI) and aligned with the original ITS nrDNA sequences for each isolate. For isolates without a prior ITS nrDNA sequence, we used Tree-Based Alignment Selector Toolkit (T-BAS) v2.2 (Miller *et al.*, 2015; Carbone *et al.*, 2017, 2019) to query sequences against the multilocus tree of the Xylariaceae from (U'Ren *et al.*, 2016). In some cases, names of reference taxa (previously named based only on morphological characters) were updated to reflect their phylogenetic placement (see Table S1). Ecological modes were assigned based on the substrate of isolation: fungi isolated from living plants and lichens with no signs of disease were classified as endophytic; fungi isolated from or collected as fruiting bodies from decomposing plant tissues (e.g., litter, wood, dung) were classified as saprotrophs; and fungi isolated or collected as fruiting bodies from living, diseased host tissues were classified as pathogens.

Fungal growth for DNA and RNA purification. For PacBio sequencing, isolates were first grown on multiple 2% MEA plates overlaid with sterile, cellophane membrane to allow mycelial harvesting without media carry-over. After ca. 5-10 days of growth, mycelium was removed using sterile forceps and scalpels, placed in 150 mL of 1% malt extract (ME) media in a sterile, stainless steel Eberbach blender cup (Fisher Scientific) and homogenized with 3-5 short pulses using a Waring blender. After homogenization, two 75 mL aliquots were placed in Erlenmeyer flasks and incubated on a shaker at room temperature for 3-7 days. Once sufficient growth was obtained samples were then filtered through sterilized Miracloth (Millipore, 475855-1R) in a Buchner funnel (Fisher Scientific), placed in a 50 mL centrifuge tube, flash-frozen in liquid nitrogen, and stored at -80°C. If isolates grew slowly, the contents of the inoculated flask were re-blended with an equal volume of fresh 1% ME media after 7 days, aliquoted into new flasks, and incubated on the shaker at room temperature for an additional 5-7 days prior to filtering. After filtration, mycelium was washed with sterile molecular grade water to remove media and excess polysaccharides.

We used a modified phenol:chloroform extraction method to achieve high molecular weight DNA for PacBio. Briefly, ca. 4 g (wet weight) of tissue was ground in liquid nitrogen with a sterile mortar and pestle. Ground tissue was transferred to a 50 mL Falcon tube containing 14 mL of SDS buffer and incubated at 65°C for 30 minutes, during which the tube was gently inverted 5X every 10 minutes. After incubation, 0.5X volume of 5M KOAc (pH 7.5) was added to each tube, mixed by inversion, and placed at 4°C for 30 minutes. Samples were then centrifuged at 4500 RPM for 10 minutes at 4°C. After centrifugation, the supernatant was removed, placed into a new tube, 0.7X volume of molecular grade isopropanol was added, and the tube was gently inverted to mix. The sample was then centrifuged at 4500 RPM for 20 minutes at 4°C to precipitate the DNA. After centrifugation the supernatant was removed, and the DNA pellet was washed with 5 mL of 70% EtOH and centrifuged for an additional 5 minutes at 4500 RPM. Residual EtOH was removed with a pipette, and the pellet was air dried. The DNA pellet was resuspended in 2 mL of TE buffer, 10 uL of RNase (20mg/mL; Invitrogen, Waltham, MA) and the sample was placed in a 37°C water bath for 1 hour. After incubation, DNA was purified with phenol:chloroform:IAA, washed with 0.3X volume of absolute molecular grade ethanol to remove polysaccharides, and precipitated by adding 1.7X volume of absolute molecular grade ethanol. The resulting DNA pellet was washed with 70% EtOH, air dried, and resuspended in low salt TE. After extraction, the purity of DNA for PacBio sequencing was verified with a EcoRI (New England BioLabs, Ipswich, MA) restriction digest and sized via electrophoresis on a 1% agarose gel with a clamped homogeneous electric field (CHEF) apparatus (Chu *et al.*, 1986) as described by (Luo & Wing, 2003).

For Illumina sequencing, isolates were first grown on multiple 2% MEA plates overlaid with sterile cellophane as described above, but harvested mycelium was placed in RNase free stainless-steel bead tubes (Next Advance, NAVYR5-RNA), flash frozen in liquid nitrogen, and stored at -80°C until extraction. DNA for Illumina sequencing was extracted using similar methods as above for PacBio, with the exception that only a small amount of tissue was used, samples were homogenized in 2 mL tubes with stainless steel beads rather than grinding in liquid N, and the initial purification with 5M KOAc was not performed (see (U'Ren & Moore)). DNA obtained from both methods was quantified with a Qubit fluorometer (Invitrogen, Carlsbad, CA) and sample purity was assessed with a NanoDrop 1000 (BioNordika, Herlev, Denmark).

To extract RNA, fungal isolates were grown on 2% MEA with sterile cellophane overlay. Mycelium was harvested after ca. one week of growth, placed in 2 mL tubes containing stainless steel beads, flash frozen in liquid N, and stored at -80°C until extraction. Frozen mycelium was homogenized for 5 seconds at 1400 RPM on a BioSpec, Mini-BeadBeater 96 115V (MP Biomedicals) with stainless steel beads. Following homogenization, 1 mL of TRIzol was added to each tube and the sample was incubated for 5 minutes at room temperature, followed by centrifugation at 4°C for 15 minutes at 12,000 RPM. Following centrifugation, the supernatant was transferred to a new tube and 0.2 mL of chloroform was added, mixed gently by inversion, and transferred to a column following the manufacturer's instructions.

Transcriptome sequencing. Plate-based RNA sample prep was performed on the PerkinElmer Sciclone NGS robotic liquid handling system using Illumina's TruSeq Stranded mRNA HT sample prep kit utilizing poly-A selection of mRNA following the Illumina protocol with the following conditions: 1 ug of total RNA per sample and eight cycles of PCR for library amplification. Libraries were quantified using KAPA Biosystems' NGS library qPCR kit and run on a Roche LightCycler 480 real-time PCR instrument. Sequencing of the flowcell was performed on the Illumina NovaSeq sequencer using NovaSeq XP V1 reagent kits, S4 flowcell, following a 2x150 indexed run recipe. Raw reads were evaluated with BBDuk (https://sourceforge.net/projects/bbmap/) for artifact sequences by kmer matching (kmer=25), allowing 1 mismatch and detected artifacts 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 1/3 of the original read length - whichever is longer). Filtered reads were assembled into consensus sequences using Trinity v2.3.2 (Grabherr *et al.*, 2011) with the --normalize_reads (In-silico normalization routine) and --jaccard_clip options. On average, ca. 90% of RNAseq reads mapped to each genome (Table S1).

Gene prediction pipeline and validation. Fungal genomes sequenced at JGI were annotated using the JGI Annotation Pipeline (Kuo *et al.*, 2014; Grigoriev *et al.*, 2014) Briefly, gene models are predicted using multiple approaches: (i) *ab initio*: FGENESH (Salamov & Solovyev, 2000) and GeneMark-ES (Ter-Hovhannisyan *et al.*, 2008); (ii) homology-based: GeneWise (Birney *et al.*, 2004) and FGENESH+ (Salamov & Solovyev, 2000); or (iii) transcriptome-based: EST_MAP (http://www.softberry.com/) and Combest (Zhou *et al.*, 2015). The best representative model at each locus was selected from the collection of all models through an automated filtering procedure based on homology and transcriptome support (Grigoriev *et al.*, 2014). To facilitate more efficient annotation for the large sequencing effort described here, we implemented a modified annotation pipeline using a subset of gene modelers: GeneMark-ES, FGENESH+, EST_MAP, and Combest. In addition to improving computational speed and storage improvements, this modified pipeline showed no significant impact on gene model quality with internal benchmarks using organisms with a large number of previously annotated genomes that can be used for homology-based gene predictions. The 34 and 87 genomes annotated using either the "standard" or "modified" annotation pipeline, respectively, are indicated in Table S1. We found no association between the sequencing method and genome size or gene content, even when analyses were restricted to closely related sister taxa with the same ecological mode (All genomes: t_{90} = -0.79, P = 0.4341; Sister taxa: t_{12} = 0.24, P = 0.8114). Genomes that were sequenced with PacBio did not contain a significantly greater number of total repetitive elements (All genomes t_{90} = 0.28, P = 0.7775; Sister taxa: t_{12} = 1.63, P = 0.1290) even though repetitive regions typically assemble better with long-read sequencing.

Functional annotation and ancestral state reconstruction of orthologous gene families. All 1,451,488 genes from the 121 genomes (ingroup and outgroup) were clustered into 104,604 orthologous groups (i.e., orthogroups). Approximately 25% (26,825) of orthogroups were assigned functional annotations with KinFin v1.0 (Laetsch & Blaxter, 2017) (Table S7), which performs a representative functional annotation of the orthogroups based on both the proportion of proteins in the group carrying a specific annotation as well as the proportion of taxa in the cluster with such annotation. Specifically, orthogroup annotation criteria were: (i) a minimum of 75% of the proteins in the orthogroup share the annotation and (ii) 30% of the taxa in the cluster with at least one protein annotated with that domain. Gene ontology (GO) terms were designated for 6,458 orthogroups (6.2%), while 10,820 (10.3%) and 11,144 (10.7%) of orthogroups were assigned to InterPro and Pfam domains, respectively. A small fraction of orthogroups were assigned IDs as carbohydrate-active enzymes (CAZyme; 720, 0.7%), peptidases and peptidase inhibitors (MEROPS DB; 443 orthogroups, 0.4%), and transporters (TCDB; 1154, 1.10%). The total number of gene families with signal peptides was 15,076 (14.4%), among which 2,869 (2.7%) were annotated as effectors (Table S7). We then used ancestral gene content reconstruction in Count v10.04 (Csurös, 2010) with the unweighted Wagner parsimony method (gain and loss penalties both set to 1) to assess changes in the size of orthologous gene families over evolutionary time. Functional annotation of orthogroups was imported into Count v10.04 GUI to reconstruct the ancestral gene content for subsets of orthologous gene families corresponding to CAZymes and PCWDEs.

Evolutionary relationships of endophytic, saprotrophic, and pathogenic Xylariaceae s.l. and Hypoxylaceae. Maximum likelihood phylogenomic analyses were performed with IQ-Tree using a concatenated matrix of 1,526 universal, single-copy orthogroups (Fig. 1a; Fig. S2). Phylogenomic results support the monophyly of the newly proposed families of Graphostromataceae and Hypoxylaceae (Wendt *et al.*, 2018), as well as previously observed relationships among genera (U'Ren *et al.*, 2016) (Fig. S2). Dense gene sampling resulted in improved resolution and statistical support for deeper internal branches compared to a previous five-gene analysis (U'Ren *et al.*, 2016). Inclusion of previously unstudied endophytic taxa markedly increased the known phylogenetic diversity of the family (U'Ren *et al.*, 2016) (Fig. S2), highlighting the importance of including unnamed endophytes (which are typically sterile mycelium in culture which precludes morphological characterization and formal naming; but see (Harrington *et al.*, 2019)) in phylogenetic studies.

Our analyses revealed seven endophytic isolates in five distinct clades (i.e., clades E2, E4, E5, E6, and E6) nested between the Graphostromataceae and Xylariaceae *sensu stricto* (Fig. S2). To better ascertain their taxonomic relationships, we performed additional phylogenetic analysis that included recently published xylarialean taxa closely related to Xylariaceae and Graphostromataceae (i.e., *Barrmaelia*, Barrmaeliaceae (Voglmayr *et al.*, 2018); *Linosporopsis* and *Clypeosphaeria*, Xylariaceae (Voglmayr & Beenken, 2020)). Briefly, we queried sequences of RPB2, alpha-actin, beta-tubulin, and ITS nrDNA for 35 taxa not included in previous multilocus analyses that contained xylarialean endophytes (U'Ren *et al.*, 2016) (e.g., *Barrmaelia*, *Linosporopsis, Clypeosphaeria Entosordaria, Graphostroma; Cryptostroma* (Li et al., 2021)) against the reference multilocus Xylariaceae tree (U'Ren *et al.*, 2016) in T-BAS v2.2 (Miller *et al.*, 2015; Carbone *et al.*, 2017, 2019) with the evolutionary placement algorithm in RAxML (Berger *et al.*, 2011). The settings that we used to place taxa within the reference tree were as follows: UNITE filter off, no clustering, likelihood weights (fast), with the outgroup selected, and data were retained for all isolates. This analysis revealed that endophytes in clade E4 are sister to *Barrmaelia*, endophytes in clade E5 are sister to *Linosporopsis*, and endophytes in clade E6 are sister to *Clypeosphaeria* (Fig. S2). Thus, our use of Xylariaceae *sensu stricto* and Xylariaceae *sensu lato* corresponds to (Voglmayr *et al.*, 2018) (see Fig. S2).

Phylogenomic results are robust to outgroup taxa, gene selection, and phylogenetic methodology. To assess the robustness of our phylogenetic results we reconstructed the phylogeny of Xylariaceae s.l. and Hypoxylaceae using four different approaches that differed in either outgroup taxon selection, model of inference, or orthologous gene set. First, we performed a maximum likelihood (ML) analysis of 1,526 single-copy orthogroups (found across all 121 ingroup and outgroup taxa) with the LG model of evolution (Fig. 1, Fig. S2). Second, we performed an ML analysis of the same orthologous genes and 121 taxa, but with the JTT+F+I+G4 model of evolution, which was the best evolutionary model selected by ModelFinder in IQ-TREE (Fig. S3). Third, we performed an ML analysis with the JTT+F+I+G4 model of evolution and the same orthologous genes, but after removing non-Xylariales taxa from the outgroup (data not shown). We performed a fourth ML analysis with all taxa (i.e., 121 ingroup and outgroup), but with 1,086 protein sequences identified as universal fungal orthologs with fungal genomes from JGI Mycocosm (Grigoriev *et al.*, 2014). JGI orthologs were identified in genomes using the PHYling pipeline

(DOI: 10.5281/zenodo.1257002; https://github.com/stajichlab/PHYling_unified). All phylogenetic analyses were performed with IQ-TREE multicore v1.6.1178 with 1,000 ultrafast bootstrap replicates (data not shown). All phylogenetic analyses resulted in similar topologies, although relationships among taxa in the Xylaria HY and E9 clades differed slightly with the LG model (analysis 1) and the JTT+F+I+G4 (analysis 2) (see Fig. S13).

We reconstructed the species tree using a coalescent-based approach. For each of the 1,526 universal single-copy orthogroups defined by OrthoFinder (Emms & Kelly, 2019), we perform a multiple sequence alignment using MAFFT v7.427 (Katoh & Standley, 2013), selected the best-fitting model of amino acid evolution with ModelFinder, and reconstructed its phylogeny in IQ-TREE multicore v1.6.1178 (Nguyen *et al.*, 2015) with 1,000 ultrafast bootstrap replicates. The gene trees were then used to reconstruct the species tree using ASTRAL version 5.15.4 (Mirarab *et al.*, 2014). The resultant species tree had a similar topology to trees reconstructed with concatenated supermatrix (including the placement of Xylariaceae sp. FL2044), although ASTRAL recovered different relationships among taxa in the Xylaria HY and E9 clades (see Fig. S13).

In previous multi-locus analyses, the endophytic isolate Xylariaceae sp. FL2044 was placed as a sister to the Xylariaceae s.l. and Hypoxylaceae (U'Ren *et al.*, 2016). However, both concatenated and coalescent phylogenomic analyses consistently placed FL2044 as basal within the monophyletic clade containing Xylariaceae s.l. (Fig. 1a; Fig. S2). Network analysis of shared orthogroups also supports the placement of FL2044 in the Xylariaceae s.l. clade (Fig. S14). To further investigate the placement of FL2044, we computed single-gene trees with IQ-TREE and used ETE Toolkit (http://etetoolkit.org/) to quantify the number of genes that supported the placement of FL2044 as recovered in our concatenated phylogenomic analyses. Overall, of the 882 single-gene trees where the placement of FL2044 was highly supported (i.e., >75% bootstrap), 297 (33.7%) agree with the placement of FL2044 in our concatenated analyses (Fig. 1; Figs. S2 and S3).

Determination of core, family-specific, clade-specific, and isolate-specific orthogroups and SMGCs. To visualize the distribution of orthogroups and SMGCs across taxa, we categorized orthogroups/SMGCs into 10 categories (see Tables S3 and S7). To visualize the relative abundance of these categories across the phylogeny, we combined categories in the following manner for Fig. S3e. Core: orthogroups/SMGCs present in all 121 taxa (cat a), as well as orthogroups/SMGCs present in all Xylariaceae s.l. and Hypoxylaceae taxa and in some outgroup taxa (cat c). Family-specific (i.e., Xylariaceae s.l. and Hypoxylaceae specific): orthogroups/SMGCs present in all or some Xylariaceae s.l. and Hypoxylaceae taxa but absent in outgroup taxa (cat b and cat d). Hypoxylaceae-specific: orthogroups/SMGCs present in all or some Hypoxylaceae taxa but absent in Xylariaceae s.l. taxa and outgroup taxa (cat e and cat f).

Xylariaceae s.l.-specific: orthogroups/SMGCs present in all or some Xylariaceae s.l. taxa but absent in Hypoxylaceae taxa and outgroup taxa (cat g and cat h). Isolate-specific: orthogroups/SMGCs found only in a single genome (cat j). Orthogroups/SMGCs that did not fall into any of these categories were defined as "other" (cat i). Examples of the "other" category include orthogroups/SMGCs that are present in some outgroup taxa, as well as some Hypoxylaceae and/or Xylariaceae s.l. taxa. Orthogroups/SMGCs distributions falling in the "other" category may have arisen through HGT, ancestral gene duplication and gene loss, or interspecific hybridization (Keeling & Palmer, 2008). We found that no orthogroups were both unique to- and universally present in all Xylariaceae s.l. and Hypoxylaceae taxa (Table S7d). A single orthogroup (annotated as a putative signaling peptide; OG0009755) was specific to and universally distributed in the Hypoxylaceae clade, but no orthogroups met these criteria for the Xylariaceae s.l. clade. Overall, ca. 21-37% of the orthogroups per genome (mean = 27.4%) represented orthogroups shared by all 121 taxa (i.e., core genes; n =2,656 total) (Fig. S3e, Table S7). An additional 1,831 orthogroups were present in all Xylariaceae s.l. and Hypoxylaceae and one or more outgroup taxa (Table S7d), representing an average of 14-23% orthologous gene families per genome (mean = 18.5%; Fig. S3e). Gene families unique to Xylariaceae s.l. and Hypoxylaceae (i.e., absent in the outgroups and present in at least one genome in both Hypoxylaceae and Xylariaceae s.l. clades) represented, on average, ca. 1.6% of orthogroups per genome (Fig. S3e, orange bars). An average of 3.0% and 3.8% of orthogroups were unique to Hypoxylaceae or Xylariaceae s.l. taxa, respectively (Fig. S3e, Table S7d).

Orthogroups unique to a single genome (i.e., dispensable orthogroups) represent ca. 1.4 to 15.6% of the orthogroups per genome for Xylariaceae s.l. and Hypoxylaceae (Fig. S3e). Functional annotation using euKaryotic Orthologous Groups (KOGs) revealed a greater fraction of dispensable orthogroups were predicted to be involved in cellular processes and signaling (i.e., 42.6%) compared to core orthogroups (27.7%), including a higher fraction of orthogroups annotated as defense mechanisms and extracellular structures (Fig. S15, Table S7f). Dispensable orthogroups also were more likely than core orthogroups to encode proteins secreted through the general secretory pathway (15.0% vs 2.7%), supporting the hypothesis that strain-specific genes may provide ecological adaptations (Haridas *et al.*, 2020). However, the functions of the majority of dispensable orthogroups remain unknown (i.e., only 20% had functional annotation vs. 90% of core orthogroups), similar to results from Dothideomycetes genomes (Haridas *et al.*, 2020).

Comparison of Hypoxylaceae and Xylariaceae s.l. SMGCs to MIBiG. Although there has been increasing biochemical characterization of metabolites from species of Xylariaceae s.l. and Hypoxylaceae (e.g., terpenes and polyketide compounds (Becker & Stadler, 2021)), fewer studies have linked metabolites to gene clusters. Here, we compared predicted SMGCs to a reference database of known metabolites clusters (MIBiG) (Medema *et al.*, 2015). Only 25% of predicted SMGCs (n = 1,711, belonging to 816 cluster families) had BLAST hits to 168 unique MIBiG (Medema *et al.*, 2015) accession numbers (Table S3b). The majority of MIBiG hits were classified as PKSI (808 hits), terpene synthases (268 hits), and PKS-NRPS hybrids (253 hits). The remaining 382 hits were classified as NRPS, PKS-Other, RiPPS, and Other SMGCs. The average similarity of SMGCs to a MIBiG accession was 54% (range 13-100%) (Table S3), but 587 xylarialean SMGCs were 100% similar to 38 MIBiG accessions (Table S3).

Similarity to MIBiG is currently defined as the percentage of genes in an SMGC with significant BLAST hits to a known SMGC (Medema *et al.*, 2011), yet similarity can be difficult to assess given the dynamic nature of SMGCs (i.e., frequent gene duplications, gene losses, and HGT (Wisecaver *et al.*, 2014; Lind *et al.*, 2017)) and the potential for *in silico* methods to misidentify cluster boundaries. For example, the griseofulvin cluster of *Penicillium aethiopicum* is predicted to contain 21 genes, but only core genes Gsf A, I, and G have been experimentally validated (Chooi *et al.*, 2010). *Xylaria* taxa, despite lacking 13 genes (GsfR2, GsfK, GsfR1, GsfJ, GsfH and all eight genes of unknown function; Fig. S5a), produce detectable levels of griseofulvin in culture (Fig. S5b, see also (Mead *et al.*, 2019)). However, lower similarity may also reflect true differences in cluster composition and the production of similar, but distinct metabolites. Variation may also represent null alleles unable to synthesize the metabolite (e.g., aflatoxin in *A. flavus* (Chang et al., 2005)). Currently, databases such as MIBiG primarily contain metabolites from bioactive fungi with important roles as human or plant pathogens, and increased effort is needed to link metabolites from xylarialean fungi to specific gene clusters.

Intergenic distances, repetitive elements, effectors, and SMGCs. The software BEDTools version 2.29.2 (Quinlan, 2014) was used to calculate the distance between adjacent genes (intergenic distance) and the distance between each gene and the closest repetitive element on the 5' and the 3' end following (Frantzeskakis *et al.*, 2018). Results were visualized using the package 'ggplot2' version 3.3.2 in R and previously published code (Frantzeskakis *et al.*, 2018) (https://github.com/lambros-f/blumeria_2017). The mean intergenic distance for all Xylariaceae s.l. and Hypoxylaceae genomes was 1,776 ± 415 bp. For all genomes, the distribution of intergenic distances followed a normal distribution, except for the genome of *Sodiomyces alkalinus*, which displayed an increase in the frequency of genes with an intergenic distance towards 10,000 bp. Repetitive elements occurred more frequently in gene-sparse regions and at the end of contigs (Table S2). Since *de novo* genome assemblers can collapse when reaching a repetitive region larger than the read length itself (Thomma *et al.*, 2016), we surmise that our genome assemblies may be fragmented because of complex regions rich in repetitive elements.

To identify whether SMGCs and effectors were in regions of the genome with high repeats and sparse gene content, we performed the same calculation of intergenic distances and visualized the locations as a function of gene density and TE location. Xylariaceae s.l. genomes had a higher density of repetitive elements surrounding genes (including genes annotated as effectors and genes identified as high-confidence HGT) compared to Hypoxylaceae (Fig. S6). However, we saw no significant difference in repeat density for effector genes vs. non-effector genes within a clade (Fig. S6). In the majority of Xylariaceae s.l. and Hypoxylaceae genomes, numerous SMGCs, and genes annotated as effectors are located at the edge of contigs in gene sparse/high repeat regions including the griseofulvin cluster in *Xylaria*. sp. However, there was no relationship between SMGC number (residuals after accounting for genome size) and the number of scaffolds obtained from genome assembly (Fig. S16) suggesting that fragmentation of genome assemblies did not artificially increase the predicted number of SMGCs (Navarro-Muñoz *et al.*, 2020). Repetitive-rich regions, often near telomeres and centromeres, can represent hotspots of gene gain/loss events as transposable elements facilitate gene dispersal both within and among genomes (Wisecaver & Rokas, 2015; Slot, 2017; Rokas *et al.*, 2018). The presence of SMGCs in these regions may drive the hyperdiversity of SMGCs within the Xylariales, as well as the discontinuous phylogenetic distribution of SMGCs across the studied genomes (see Figs. 1 and 3).

Confirmation of griseofulvin HGT. We examined regions flanking in *Xylaria* sp. with and without the griseofulvin cluster to further confirm HGT. Briefly, 30 kbp sequences located up- and downstream of the griseofulvin cluster of *Xylaria flabelliformis* CBS 123580 were queried with BLASTn against closely related genomes without the griseofulvin cluster to identify homologous regions (*X. longipes* CBS 148.73 scaffold 57 and *X. acuta* CBS 122032 scaffold 139). Scaffolds containing these homologous regions, along with the scaffolds containing the griseofulvin cluster in *X. flabelliformis* NC1011 (scaffold 71), *X. flabelliformis* CBS 124033 (scaffold 75), *X. flabelliformis* CBS 123580 (scaffold 16), *X. flabelliformis* CBS 114988 (scaffold_56), *X. flabelliformis* CBS 116.85 (scaffold 29), were then aligned using Mauve (Darling *et al.*, 2004). In *X. flabelliformis* isolates, the scaffold alignment contains the up- and downstream homology blocks with the intervening griseofulvin cluster. Up- and downstream homology blocks were also found in *X. longipes* CBS 148.73; however, the griseofulvin cluster was not present, thus supporting the HGT of griseofulvin cluster in some taxa.

Comparison of leaf litter decomposition among clades. To assess the ability of Xylariaceae s.l. and Hypoxylaceae fungi to degrade lignocellulose, we collected fresh, healthy, green leaf material from two individuals of *Quercus virginiana* and *Pinus halepensis* at the University of Arizona campus arboretum. Trees are cultivated in a park-like setting with supplemental water and appear healthy. For both species, leaves were washed in tap water to remove any surface debris. Washed leaves were autoclaved for 20 min to inactivate endogenous microbes and then dried overnight at 60°C. Autoclaved leaves (0.5 g) were placed into individual, sterile 100 mm Petri plates (three replicate plates per leaf substrate type for each fungal isolate). For each fungal isolate, a 6 mm plug of mycelium (actively growing on 2% MEA) was briefly homogenized with a sterile minipestle in 1 mL of sterile water until mycelia had visually separated from the agar chunks. From this 1 mL mixture, 75 µL was diluted with 3 mL of sterile water and mixed via pipetting to create the fungal inoculum. One mL of the diluted ground mycelium was placed directly on the sterile leaf surface in each Petri dish. Negative control samples were inoculated in parallel with sterile water. In total, we inoculated three replicate plates per fungal isolate per plant species (total of 120 plates). Petri plates were sealed with Parafilm and weighed on an analytical balance ($\text{mass}_{\text{original}}$). Plates were stored in the dark at 26°C for the duration of the experiment (12 weeks). Each plate was weighed weekly, and the percent of leaf tissue covered with mycelium was visually scored (0 = no visible growth; 1 = 1-25% leaf coverage; 2 = 26-50% leaf coverage; 3 = 51-75% leaf coverage; 4 = 76-100% leaf coverage). Negative controls did not display fungal growth. We calculated mass loss for each replicate and control as $\text{mass}_{\text{final}} = \text{mass}_{\text{week12}} – \text{mass}_{\text{original}}$. To account for water loss due to evaporation, we then subtracted the average value of the negative control plates ($\text{mass}_{\text{norm}} = \text{mass}_{\text{final}} – \text{mass}_{\text{control}}$). We compared the normalized mass loss among clades with ANOVA (Fig. S12).

Comparison of Xylariaceae s.l. and Hypoxylaceae geographic ranges. Increased metabolic diversity and host breadth of Xylariaceae species is likely to impact species geographical distributions (Barberán *et al.*, 2014). Xylariaceae genera such as *Xylaria* and *Nemania* occur worldwide as fruiting bodies in temperate, subtropical, and tropical forests, whereas Hypoxylaceae genera such as *Daldinia* and *Hypoxylon* are more common in boreal and temperate forests (U'Ren *et al.*, 2016), but taxonomic uncertainty for many specimens and sequences (Persoh *et al.*, 2009) combined with a lack of biome metadata for the majority of reference taxa (U'Ren *et al.*, 2016) precludes robust statistical comparisons of biogeographic ranges for named taxa. However, a recent global survey of boreal endophytes demonstrated that host generalist species occupy larger geographic ranges (U'Ren *et al.*, 2019). As a preliminary assessment to address this question, we performed an analysis of previously published data from ecological surveys in boreal, temperate montane, and subtropical forests in Alaska, Arizona, North Carolina, and Florida. This analysis reveals a higher fraction of our Xylariaceae endophyte species were cultured from hosts in more than one site (i.e., 28% Xylariaceae vs. 20% for Hypoxylaceae), including six Xylariaceae endophyte species that were found in >3 sites (U'Ren *et al.*, 2016). In contrast, no Hypoxylaceae endophyte species were found in more than two sites (U'Ren *et al.*, 2016). Future studies will robustly address the relationship between metabolic diversity and host breadth in endophytic fungi.

Metabolite extraction and identification. To induce the production of SMs and potentially verify SMGCs, we performed co-culture experiments with three isolates: *X. flabelliformis* NC1011, *Xylaria arbuscula* FL1030, and *Daldinia* sp. FL1419. Isolates were grown on *Aspergillus* defined media with glucose. After one week, we removed 6 mm diameter plugs of actively growing mycelium from each isolate for three pairwise combinations of co-culture plates (i.e., NC1011 vs. FL1419; FL1419 vs. NC1030; NC1030 vs. NC1011). Briefly, agar plugs of two isolates were placed ~4.5 cm apart across the horizontal diameter of a 100 mm Petri dish (see Fig. S5b). We inoculated four replicate co-culture plates for each combination (total 12 interaction plates) and four plates containing each isolate alone (total 12 positive control plates). Plates were incubated at room temperature for 8-10 days or until the mycelium from the two isolates was ~1cm apart. Using a sterile transfer tube, we harvested five 6 mm plugs of agar either (i) next to a single culture (i.e, positive control plates); (ii) in the space between isolates (i.e., interaction plates) to ensure the capture of exogenous SMs; or (iii) in the middle of media control plate. After harvesting agar plugs were placed into sterile, 2.0 mL microcentrifuge tubes, flash frozen in liquid Nitrogen, and stored at -80°C. Frozen samples were shipped to JGI for extraction and stored and -80℃ until processed.

To extract metabolites for LC-MS/MS, samples were lyophilized dry (FreeZone 2.5 Plus, Labconco), then bead-beaten to a fine powder with a 3.2 mm stainless steel bead for 5 seconds (2x) in a bead-beater (Mini-Beadbeater-96, BioSpec Products). For extraction, 500 µL of MeOH was added to each sample, briefly vortexed, sonicated in a water bath for 5 minutes, and centrifuged for 5 min at 5000 rpm to pellet agar and cellular debris. The supernatant was transferred to a 2 mL Eppendorf, dried in a SpeedVac (SPD111V, Thermo Scientific), and stored at -80 °C. Extraction controls were prepared similarly but using empty tubes exposed to the same extraction procedures. In preparation for LC-MS/MS analysis, dried extracts were resuspended by adding 300 µL methanol containing 10 µg/mL of 2-Amino-3-bromo-5methylbenzoic acid (#R435902, Sigma) as internal standard, vortexed briefly, sonicated in a water bath for 10 min, and centrifuged (5 min at 5000 rpm). After centrifugation, 150 µL of the resuspended extract was filtered via centrifugation (2.5 min at 2500 rpm) through a 0.22 µm filter (UFC40GV0S, Millipore) and transferred to a glass autosampler vial.

Samples were analyzed on a system consisting of an Agilent 1290 UHPLC coupled to a Thermo QExactive Orbitrap HF (Thermo Scientific, San Jose, CA) mass spectrometer. Reverse phase chromatography was performed by injecting 2 µL extract into a C18 column (Agilent ZORBAX Eclipse Plus C18, 2.1x50 mm, 1.8 µm) warmed to 60°C with a flow rate of 0.4 mL/min equilibrated with 100% buffer A (100% LC-MS water with 0.1% formic acid) for 1 minute, followed by a linear gradient to 100% buffer B (100% acetonitrile w/ 0.1% formic acid) for 7 minutes, and then held at 100% B for 1.5

minutes. MS and MS/MS data were collected in both positive and negative ion mode, with full MS spectra acquired ranging from 90-1350 *m/z* at 60,000 resolution, and fragmentation data acquired using an average of stepped collision energies of 10, 20 and 40 eV at 17,500 resolution. Orbitrap instrument parameters included a sheath gas flow rate of 50 (au), an auxiliary gas flow rate of 20 (au), sweep gas flow rate of 2 (au), 3 kV spray voltage, and 400 °C capillary temperature. Sample injection order was randomized, and an injection blank of methanol only was run between each sample. Metabolites were identified based on comparing exact mass (ppm difference between detected m/z to a compound's theoretical m/z) and comparing experimental MS/MS fragmentation spectra to that of standards. These data confirmed the production of griseofulvin by NC1011 when grown in co-culture with FL1419 (Fig. S5b).

REFERENCES

Barberán A, Ramirez KS, Leff JW, Bradford MA, Wall DH, Fierer N. **2014**. Why are some microbes more ubiquitous than others? Predicting the habitat breadth of soil bacteria. *Ecology letters* **17**: 794–802.

Bastian M, Heymann S, Jacomy M. **2009**. Gephi: an open source software for exploring and manipulating networks. In: Third international AAAI conference on weblogs and social media. aaai.org.

Becker K, Stadler M. **2021**. Recent progress in biodiversity research on the Xylariales and their secondary metabolism. *The Journal of antibiotics* **74**: 1–23.

Berger SA, Krompass D, Stamatakis A. **2011**. Performance, accuracy, and Web server for evolutionary placement of short sequence reads under maximum likelihood. *Systematic biology* **60**: 291–302.

Birney E, Clamp M, Durbin R. **2004**. GeneWise and Genomewise. *Genome research* **14**: 988–995.

Carbone I, White JB, Miadlikowska J, Arnold AE, Miller MA, Kauff F, U'Ren JM, May G, Lutzoni F. **2017**. T-BAS: Tree-Based Alignment Selector toolkit for phylogenetic-based placement, alignment downloads and metadata visualization: an example with the Pezizomycotina tree of life. *Bioinformatics* **33**: 1160–1168.

Carbone I, White JB, Miadlikowska J, Arnold AE, Miller MA, Magain N, U'Ren JM, Lutzoni F. **2019**. T-BAS Version 2.1: Tree-Based Alignment Selector Toolkit for Evolutionary Placement of DNA Sequences and Viewing Alignments and Specimen Metadata on Curated and Custom Trees. *Microbiology resource announcements* **8**.

Chang P-K, Horn BW, Dorner JW. **2005**. Sequence breakpoints in the aflatoxin biosynthesis gene cluster and flanking regions in nonaflatoxigenic Aspergillus flavus isolates. *Fungal genetics and biology: FG & B* **42**: 914–923.

Chooi Y-H, Cacho R, Tang Y. **2010**. Identification of the viridicatumtoxin and griseofulvin gene clusters from Penicillium aethiopicum. *Chemistry & biology* **17**: 483–494.

Chu G, Vollrath D, Davis RW. **1986**. Separation of large DNA molecules by contour-clamped homogeneous electric fields. *Science* **234**: 1582–1585.

Csurös M. **2010**. Count: evolutionary analysis of phylogenetic profiles with parsimony and likelihood. *Bioinformatics* **26**: 1910–1912.

Darling ACE, Mau B, Blattner FR, Perna NT. **2004**. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome research* **14**: 1394–1403.

Emms DM, Kelly S. **2019**. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome biology* **20**: 238.

Frantzeskakis L, Kracher B, Kusch S, Yoshikawa-Maekawa M, Bauer S, Pedersen C, Spanu PD, Maekawa T, Schulze-Lefert P, Panstruga R. **2018**. Signatures of host specialization and a recent transposable element burst in the dynamic one-speed genome of the fungal barley powdery mildew pathogen. *BMC genomics* **19**: 381.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, *et al.* **2011**. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature biotechnology* **29**: 644–652.

Grigoriev IV, Nikitin R, Haridas S, Kuo A, Ohm R, Otillar R, Riley R, Salamov A, Zhao X, Korzeniewski F, *et al.* **2014**. MycoCosm portal: gearing up for 1000 fungal genomes. *Nucleic acids research* **42**: D699–704.

Haridas S, Albert R, Binder M, Bloem J, LaButti K, Salamov A, Andreopoulos B, Baker SE, Barry K, Bills G, *et al.* **2020**. 101 Dothideomycetes genomes: A test case for predicting lifestyles and emergence of pathogens. *Studies in mycology* **96**: 141–153.

Harrington AH, Olmo-Ruiz M del, U'Ren JM, Garcia K, Pignatta D, Wespe N, Sandberg DC, Huang Y-L, Hoffman MT, Arnold AE. **2019**. Coniochaeta endophytica sp. nov., a foliar endophyte associated with healthy photosynthetic tissue of Platycladus orientalis (Cupressaceae). *Plant and Fungal Systematics* **64**: 65–79.

Hsieh H-M, Ju Y-M, Rogers JD. **2005**. Molecular phylogeny of Hypoxylon and closely related genera. *Mycologia* **97**: 844–865.

Hsieh H-M, Lin C-R, Fang M-J, Rogers JD, Fournier J, Lechat C, Ju Y-M. **2010**. Phylogenetic status of Xylaria subgenus Pseudoxylaria among taxa of the subfamily Xylarioideae (Xylariaceae) and phylogeny of the taxa involved in the subfamily. *Molecular Phylogenetics and Evolution* **54**: 957–969.

Katoh K, Standley DM. **2013**. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution* **30**: 772–780.

Keeling PJ, Palmer JD. **2008**. Horizontal gene transfer in eukaryotic evolution. *Nature reviews. Genetics* **9**: 605–618.

Kuo A, Bushnell B, Grigoriev IV. **2014**. Fungal genomics: sequencing and annotation. *Advances in botanical research* **70**: 1–52.

Laetsch DR, Blaxter ML. **2017**. KinFin: Software for Taxon-Aware Analysis of Clustered Protein Sequences. *G3* **7**: 3349–3357.

Li Q, Gong X, Zhang X, Pi Y, Long S, Wu Y, Shen X, Kang Y, Kang J. **2021**. Phylogeny of graphostromatacea with Three species isolated in China. *Research Square*.

Lind AL, Wisecaver JH, Lameiras C, Wiemann P, Palmer JM, Keller NP, Rodrigues F, Goldman GH, Rokas A. **2017**. Drivers of genetic diversity in secondary metabolic gene clusters within a fungal species. *PLoS biology* **15**: e2003583.

Luo M, Wing RA. **2003**. An improved method for plant BAC library construction. *Methods in molecular biology* **236**: 3–20.

Mead ME, Raja HA, Steenwyk JL, Knowles SL, Oberlies NH, Rokas A. **2019**. Draft Genome Sequence of the Griseofulvin-Producing Fungus Xylaria flabelliformis Strain G536. *Microbiology resource announcements* **8**.

Medema MH, Blin K, Cimermancic P, de Jager V, Zakrzewski P, Fischbach MA, Weber T, Takano E, Breitling R. **2011**. antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic acids research* **39**: W339–46.

Medema MH, Kottmann R, Yilmaz P, Cummings M, Biggins JB, Blin K, de Bruijn I, Chooi YH, Claesen J, Coates RC, *et al.* 2015. Minimum Information about a Biosynthetic Gene cluster. *Nature chemical biology* **11**: 625–631.

Miller MA, Schwartz T, Pickett BE, He S, Klem EB, Scheuermann RH, Passarotti M, Kaufman S, O'Leary MA. **2015**. A RESTful API for Access to Phylogenetic Tools via the CIPRES Science Gateway. *Evolutionary bioinformatics online* **11**: 43–48.

Mirarab S, Reaz R, Bayzid MS, Zimmermann T, Swenson MS, Warnow T. **2014**. ASTRAL: genome-scale coalescent-based species tree estimation. *Bioinformatics* **30**: i541–8.

Navarro-Muñoz JC, Selem-Mojica N, Mullowney MW, Kautsar SA, Tryon JH, Parkinson EI, De Los Santos ELC, Yeong M, Cruz-Morales P, Abubucker S, *et al.* 2020. A computational framework to explore large-scale biosynthetic diversity. *Nature chemical biology* **16**: 60–68.

Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. **2015**. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular biology and evolution* **32**: 268–274.

Persoh D, Melcher M, Graf K, Fournier J, Stadler M, Rambold G. **2009**. Molecular and morphological evidence for the delimitation of Xylaria hypoxylon. *Mycologia* **101**: 256–268.

Quinlan AR. **2014**. BEDTools: the Swiss-army tool for genome feature analysis. *Current protocols in bioinformatics / editoral board, Andreas D. Baxevanis ... [et al.]* **47**: 11–12.

Rokas A, Wisecaver JH, Lind AL. **2018**. The birth, evolution and death of metabolic gene clusters in fungi. *Nature reviews. Microbiology* **16**: 731–744.

Salamov AA, Solovyev VV. **2000**. Ab initio gene finding in Drosophila genomic DNA. *Genome research* **10**: 516–522.

Slot JC. **2017**. Fungal Gene Cluster Diversity and Evolution. *Advances in genetics* **100**: 141–178.

Ter-Hovhannisyan V, Lomsadze A, Chernoff YO, Borodovsky M. **2008**. Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training. *Genome research* **18**: 1979–1990.

Thomma BPHJ, Seidl MF, Shi-Kunne X, Cook DE, Bolton MD, van Kan JAL, Faino L. **2016**. Mind the gap; seven reasons to close fragmented genome assemblies. *Fungal genetics and biology: FG & B* **90**: 24–30.

U'Ren JM. **2016**. DNA extraction from fungal mycelium using Extract-n-Amp.

U'Ren JM, Lutzoni F, Miadlikowska J, Laetsch AD, Arnold AE. **2012**. Host and geographic structure of endophytic and endolichenic fungi at a continental scale. *American journal of botany* **99**: 898–914.

U'Ren JM, Lutzoni F, Miadlikowska J, Zimmerman NB, Carbone I, May G, Arnold AE. **2019**. Host availability drives distributions of fungal endophytes in the imperiled boreal realm. *Nature Ecology and Evolution* **3**: 1430–1437.

U'Ren JM, Miadlikowska J, Zimmerman NB, Lutzoni F, Stajich JE, Arnold AE. **2016**. Contributions of North American endophytes to the phylogeny, ecology, and taxonomy of Xylariaceae (Sordariomycetes, Ascomycota). *Molecular phylogenetics and evolution* **98**: 210–232.

U'Ren JM, Moore L. *Small volume fungal genomic DNA extraction protocol for Illumina genome*.

Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, Burovski E, Peterson P, Weckesser W, Bright J, *et al.* 2020. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature methods* **17**: 261–272.

Voglmayr H, Beenken L. **2020**. Linosporopsis, a new leaf-inhabiting scolecosporous genus in Xylariaceae. *Mycological progress* **19**: 205–222.

Voglmayr H, Friebes G, Gardiennet A, Jaklitsch WM. **2018**. Barrmaelia and Entosordaria in Barrmaeliaceae (fam. nov., Xylariales) and critical notes on Anthostomella -like genera based on multigene phylogenies. *Mycological progress* **17**: 155–177.

Wendt L, Sir EB, Kuhnert E, Heitkämper S, Lambert C, Hladki AI, Romero AI, Luangsa-ard JJ, Srikitikulchai P, Peršoh D, *et al.* 2018. Resurrection and emendation of the Hypoxylaceae, recognised from a multigene phylogeny of the Xylariales. *Mycological progress* **17**: 115–154.

Wisecaver JH, Rokas A. **2015**. Fungal metabolic gene clusters—caravans traveling across genomes and environments. *Frontiers in microbiology*.

Wisecaver JH, Slot JC, Rokas A. **2014**. The evolution of fungal metabolic pathways. *PLoS genetics* **10**: e1004816.

Zhou K, Salamov A, Kuo A, Aerts AL, Kong X, Grigoriev IV. **2015**. Alternative splicing acting as a bridge in evolution. *Stem cell investigation* **2**: 19.

SUPPORTING TABLES

Table S1. (a) Information for the 121 genomes included in this study; (b) Genome and assembly information for 96 Xylariaceae s.l. and Hypoxylaceae genomes included in this study.

Table S2. RepeatMasker, RepeatScout, and RepBase Update classification of repetitive elements for 96 genomes of Xylariaceae s.l. and Hypoxylaceae.

Table S3. (**a**) Secondary metabolite gene cluster (SMGC) annotations for the 121 genomes included in this study (according to antiSMASH) and grouped into families with BiG-SCAPE; (**b**) Distribution and percent similarity of Xylariaceae s.l. and Hypoxylacae SMGCs to 168 MIBiG accessions; (**c**) Count and percentage of all SMGCs and SMGC families per category (A-J); (**d-j**) Count and percentage of SMGCs per type (e.g., NRPS, Terpene, Other PKS, PKS-NRP Hybrids, Other, RiPP) per category (A-J).

Table S4. (**a**) Count of catabolic gene clusters (CGCs) by anchor gene; (**b**) Presence/Absence of CGC families per genome; (**c**) Composition of the CGC families; (**d**) Genomic position and annotation of CGCs.

Table S5. (**a**) Taxonomic and phylogenetic information for 4,262 putative HGT candidate genes identified by Alien Index (AI); (**b**) Manual curation of phylogenetic trees reveals 168 HGT candidates (each row is a unique transfer event; orthogroups may appear more than once); (**c**) Distribution of HGT counts per genome (HGT001-HGT-129 are high confidence transfers and HGT130-HGT290 are ambiguous transfers); (**d**) Functional annotation of 1,148 SMGC genes identified by the second Alien Index as candidate HGTs.

Table S6. (a) Number of genes annotated as MEROPS, CAZymes, PCWDCs, SMGCs, CGCs, and putative HGTs for genomes of 96 Xylariaceae s.l. and Hypoxylaceae; (**b**) Statistical comparison between Xylariaceae s.l. and Hypoxylaceae genomes; (**c**) Statistical comparison between endophytic and non-endophytic genomes with phylogenetic independent contrasts (PICS); (**d**) Statistical analysis of genomic features for paired endophyte/non-endophyte sister taxa using least-squares means contrasts; (**e**) Pearson correlation of genomic features as a function of ecological mode and clade.

Table S7. (a) Orthogroup summary statistics; (**b**) Orthogroup annotations; (**c**) Count and percentage of orthogroups and proteins per orthogroup category (A-J). (**d**) Orthogroups that comprise each category (A-J).

SUPPORTING APPENDICES (Available on FigShare Repository; DOI 10.6084/m9.figshare.c.5314025)

Appendix S1. InterProScan annotations for 96 Xylariaceae s.l. and Hypoxylaceae genomes.

Appendix S2. AntiSMASH output for the 96 Hypoxylaceae and Xylariaceae s.l genomes.

Appendix S3. Table summarizing the ancestral gene reconstruction by Count v10.04. The ancestral gene content was reconstructed for the entire data set, as well as for subsets of orthologous gene families from KinFin corresponding to different functional groups including (i) CAZymes; (ii) plant cell wall degrading CAZymes (PCWDCs); (iii) PCWDCs involved in the degradation of cellulose, hemicellulose, lignin, pectin, starch and inulin; (iv) peptidases; (v) peptidase inhibitors; (vi) transporters; (vii) transporters involved in the exchange of carbohydrates; (viii) transporters involved in the exchange of amino acids; (ix) transporters involved in the exchange of lipids; (x) transporters involved in the exchange of nitrogen; and (xi) effectors.

Appendix S4. Graphs of intergenic distances for each genome of Xylariaceae s.l. and Hypoxylaceae, overlaid with the location of secondary metabolite gene clusters, repeat elements, and effector genes.

Appendix S5. Graphs depicting the frequency of repetitive elements surrounding genes for each genome of Xylariaceae s.l. and Hypoxylaceae.

Appendix S6. Phylogenomic trees inferred by maximum-likelihood under the JTT+F+I+G4 model for (a) the whole dataset of 121 taxa and 1,526 protein sequences; (b) a subset of Xylariales taxa only and 1,526 protein sequences; and (c) the entire dataset of 121 taxa and 1,086 protein sequences.

Appendix S7. Table showing the sister clades to Xylariaceae sp. FL2044 recovered by the phylogenetic analysis of each of the 1,526 single-copy orthologous genes.

Appendix S8. Alignment of regions flanking the griseofulvin cluster in *Xylaria* sp. (a) Mauve (Darling *et al.*, 2004) alignment of the scaffolds containing the griseofulvin cluster in *X. flabelliformis* NC1011, *X. flabelliformis* CBS 124033, *X. flabelliformis* CBS 123580, *X. flabelliformis* CBS 114988, *X. flabelliformis* CBS 116.85, and scaffolds of the closely related *Xylaria longipes* CBS 148.73 and *Xylaria acuta* CBS 122032 showing similarity to the griseofulvin flanking regions of *X. flabelliformis* CBS 123580. (b) Same alignment after hiding the scaffolds of *Xylaria acuta* CBS 122032, *X. flabelliformis* CBS 124033, *X. flabelliformis* NC1011, *X. flabelliformis* CBS 114988. Locally collinear blocks are shown in the same colors. The plot inside the blocks indicates the level of sequence similarity. The ruler above each scaffold represents the nucleotide positions. The white boxes below represent coding sequences. The griseofulvin cluster is highlighted in light blue for *X. flabelliformis* CBS 123580; the purple block contains the griseofulvin protocluster.

SUPPORTING FIGURES AND LEGENDS

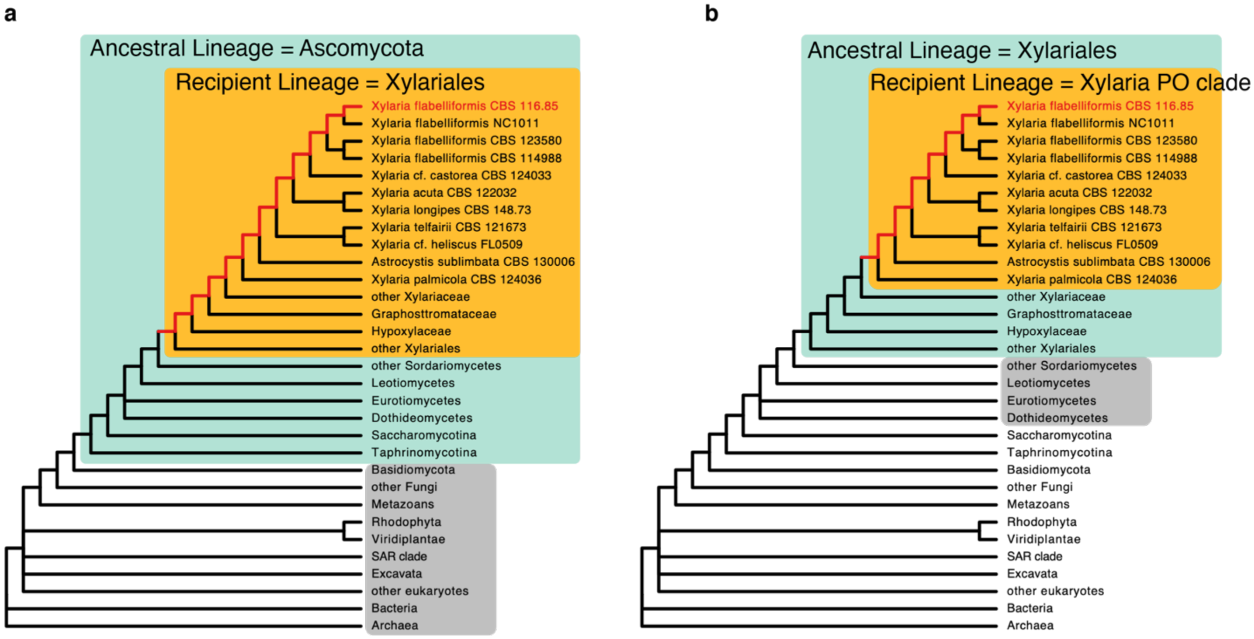
a
Ancestral Lineage = Ascomycota
Recipient Lineage = Xylariales
Xylaria flabelliformis CBS 116.85
Xylaria flabelliformis NC1011
Xylaria flabelliformis CBS 123580
Xylaria flabelliformis CBS 114988
Xylaria cf. castorea CBS 124033
Xylaria acuta CBS 122032
Xylaria longipes CBS 148.73
Xylaria telfairii CBS 121673
Xylaria cf. heliscus FL0509
Astrocystis sublimbata CBS 130006
Xylaria palmicola CBS 124036
other Xylariaceae
Graphosttromataceae
Hypoxylaceae
other Xylariales
other Sordariomycetes
Leotiomycetes
Eurotiomycetes
Dothideomycetes
Saccharomycotina
Taphrinomycotina
Basidiomycota
other Fungi
Metazoans
Rhodophyta
Viridiplantae
SAR clade
Excavata
other eukaryotes
Bacteria
Archaea
b
Ancestral Lineage = Xylariales
Recipient Lineage = Xylaria PO clade
Xylaria flabelliformis CBS 116.85
Xylaria flabelliformis NC1011
Xylaria flabelliformis CBS 123580
Xylaria flabelliformis CBS 114988
Xylaria cf. castorea CBS 124033
Xylaria acuta CBS 122032
Xylaria longipes CBS 148.73
Xylaria telfairii CBS 121673
Xylaria cf. heliscus FL0509
Astrocystis sublimbata CBS 130006
Xylaria palmicola CBS 124036
other Xylariaceae
Graphosttromataceae
Hypoxylaceae
other Xylariales
other Sordariomycetes
Leotiomycetes
Eurotiomycetes
Dothideomycetes
Saccharomycotina
Taphrinomycotina
Basidiomycota
other Fungi
Metazoans
Rhodophyta
Viridiplantae
SAR clade
Excavata
other eukaryotes
Bacteria
Archaea

Fig. S1. Overview of Alien Index (AI) calculations to identify HGT. In this example, *Xylaria flabelliformis* CBS 116.85 is the query genome. (**a**) AI screen to identify HGT candidates from more distant gene donors (grey box); candidates must have a better hit to sequences outside the ancestral lineage (Ascomycota; green box). By skipping all sequences to other Xylariales (orange box), HGT candidates could have been acquired at any point back to their last common ancestor (red branches) (**b**) AI screen to identify more recently acquired HGT candidates from other filamentous fungi (grey box). For this screen, candidates must have a better hit to sequences outside the Xylariales (green box). All sequences to other *Xylaria* “PO” clade were skipped (orange box) to identify shared HGT candidates acquired at any point back to the last common ancestor of the clade (red branches).

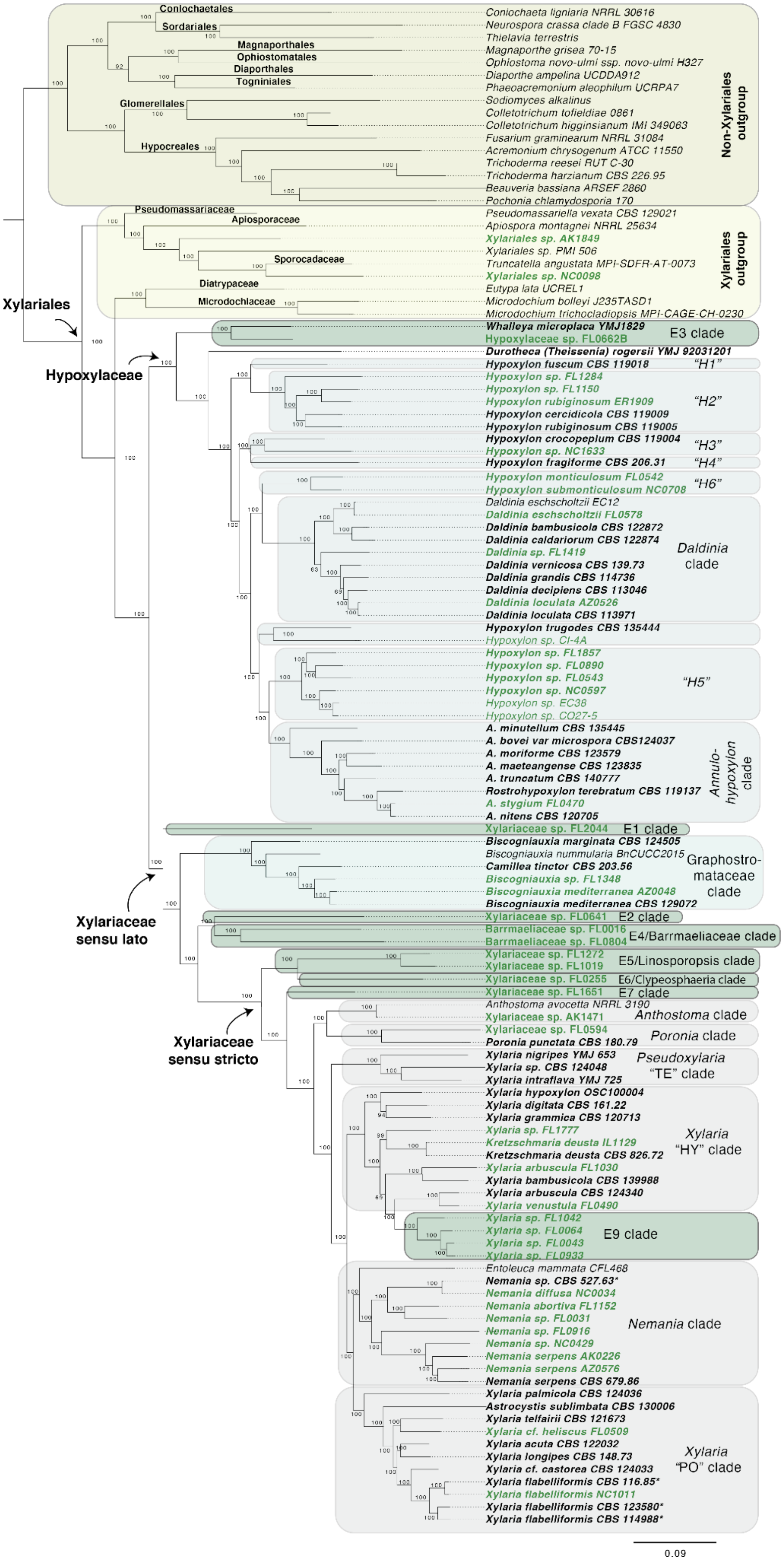
Coniochaetales
Sordariales
Magnaporthales
Ophiostomatales
Diaporthales
Togniniales
Glomerellales
Hypocreales
Coniochaeta ligniaria NRRL 30616
Neurospora crassa clade B FGSC 4830
Thielavia terrestris
Magnaporthe grisea 70-15
Ophiostoma novo-ulmi ssp. novo-ulmi H327
Diaporthe ampelina UCDDA912
Phaeoacremonium aleophilum UCRPA7
Sodiomyces alkalinus
Colletotrichum tofieldiae 0861
Colletotrichum higginsianum IMI 349063
Fusarium graminearum NRRL 31084
Acremonium chrysogenum ATCC 11550
Trichoderma reesei RUT C-30
Trichoderma harzianum CBS 226.95
Beauveria bassiana ARSEF 2860
Pochonia chlamydosporia 170
Non-Xylariales outgroup
Pseudomassariaceae
Apiosporaceae
Sporocadaceae
Diatrypaceae
Microdochiaceae
Pseudomassariella vexata CBS 129021
Apiospora montagnei NRRL 25634
Xylariales sp. AK1849
Xylariales sp. PMI 506
Truncatella angustata MPI-SDFR-AT-0073
Xylariales sp. NC0098
Eutypa lata UCREL1
Microdochium bolleyi J235TASD1
Microdochium trichocladiopsis MPI-CAGE-CH-0230
Xylariales outgroup
Xylariales
Hypoxylaceae
Whalleya microplaca YMJ1829
Hypoxylaceae sp. FL0662B
E3 clade
Durotheca (Theissenia) rogersii YMJ 92031201
Hypoxylon fuscum CBS 119018
"H1"
Hypoxylon sp. FL1284
Hypoxylon sp. FL1150
Hypoxylon rubiginosum ER1909
Hypoxylon cercidicola CBS 119009
Hypoxylon rubiginosum CBS 119005
"H2"
Hypoxylon crocopeplum CBS 119004
Hypoxylon sp. NC1633
"H3"
Hypoxylon fragiforme CBS 206.31
"H4"
Hypoxylon monticulosum FL0542
Hypoxylon submonticulosum NC0708
"H6"
Daldinia eschscholtzii EC12
Daldinia eschscholtzii FL0578
Daldinia bambusicola CBS 122872
Daldinia caldariorum CBS 122874
Daldinia sp. FL1419
Daldinia vernicosa CBS 139.73
Daldinia grandis CBS 114736
Daldinia decipiens CBS 113046
Daldinia loculata AZ0526
Daldinia loculata CBS 113971
Daldinia clade
Hypoxylon trugodes CBS 135444
Hypoxylon sp. CI-4A
Hypoxylon sp. FL1857
Hypoxylon sp. FL0890
Hypoxylon sp. FL0543
Hypoxylon sp. NC0597
Hypoxylon sp. EC38
Hypoxylon sp. CO27-5
"H5"
A. minutellum CBS 135445
A. bovei var microspora CBS124037
A. moriforme CBS 123579
A. maeteangense CBS 123835
A. truncatum CBS 140777
Rostrohypoxylon terebratum CBS 119137
A. stygium FL0470
A. nitens CBS 120705
Annulo-hypoxylon clade
Xylariaceae sp. FL2044
E1 clade
Biscogniauxia marginata CBS 124505
Biscogniauxia nummularia BnCUCC2015
Camillea tinctor CBS 203.56
Biscogniauxia sp. FL1348
Biscogniauxia mediterranea AZ0048
Biscogniauxia mediterranea CBS 129072
Graphostro-mataceae clade
Xylariaceae sensu lato
Xylariaceae sp. FL0641
E2 clade
Barrmaeliaceae sp. FL0016
Barrmaeliaceae sp. FL0804
E4/Barrmaeliaceae clade
Xylariaceae sp. FL1272
Xylariaceae sp. FL1019
E5/Linosporopsis clade
Xylariaceae sp. FL0255
E6/Clypeosphaeria clade
Xylariaceae sp. FL1651
E7 clade
Xylariaceae sensu stricto
Anthostoma avocetta NRRL 3190
Xylariaceae sp. AK1471
Anthostoma clade
Xylariaceae sp. FL0594
Poronia punctata CBS 180.79
Poronia clade
Xylaria nigripes YMJ 653
Xylaria sp. CBS 124048
Xylaria intraflava YMJ 725
Pseudoxylaria "TE" clade
Xylaria hypoxylon OSC100004
Xylaria digitata CBS 161.22
Xylaria grammica CBS 120713
Xylaria sp. FL1777
Kretzschmaria deusta IL1129
Kretzschmaria deusta CBS 826.72
Xylaria arbuscula FL1030
Xylaria bambusicola CBS 139988
Xylaria arbuscula CBS 124340
Xylaria venustula FL0490
Xylaria "HY" clade
Xylaria sp. FL1042
Xylaria sp. FL0064
Xylaria sp. FL0043
Xylaria sp. FL0933
E9 clade
Entoleuca mammata CFL468
Nemania sp. CBS 527.63*
Nemania diffusa NC0034
Nemania abortiva FL1152
Nemania sp. FL0031
Nemania sp. FL0916
Nemania sp. NC0429
Nemania serpens AK0226
Nemania serpens AZ0576
Nemania serpens CBS 679.86
Nemania clade
Xylaria palmicola CBS 124036
Astrocystis sublimbata CBS 130006
Xylaria telfairii CBS 121673
Xylaria cf. heliscus FL0509
Xylaria acuta CBS 122032
Xylaria longipes CBS 148.73
Xylaria cf. castorea CBS 124033
Xylaria flabelliformis CBS 116.85*
Xylaria flabelliformis NC1011
Xylaria flabelliformis CBS 123580*
Xylaria flabelliformis CBS 114988*
Xylaria "PO" clade
0.09

Fig. S2. Phylogenomic tree inferred by maximum likelihood based on a combination of 1,526 universal single-copy orthologous protein sequences. Twenty-five Sordariomycetes species outside Xylariales were used as the outgroup (Table S1a). Isolates sequenced in this study are highlighted in bold. Endophytes (i.e., fungi isolated from living, photosynthetic tissues of plants and lichens (U’Ren *et al.*, 2016)) are indicated in green. Clade information is based on previously published studies (see (Hsieh *et al.*, 2005, 2010; U’Ren *et al.*, 2016; Voglmayr *et al.*, 2018; Wendt *et al.*, 2018)). Numbers at nodes indicate ultrafast bootstrap support values from IQ-TREE (Nguyen *et al.*, 2015). The scale bar corresponds to the number of substitutions per site.

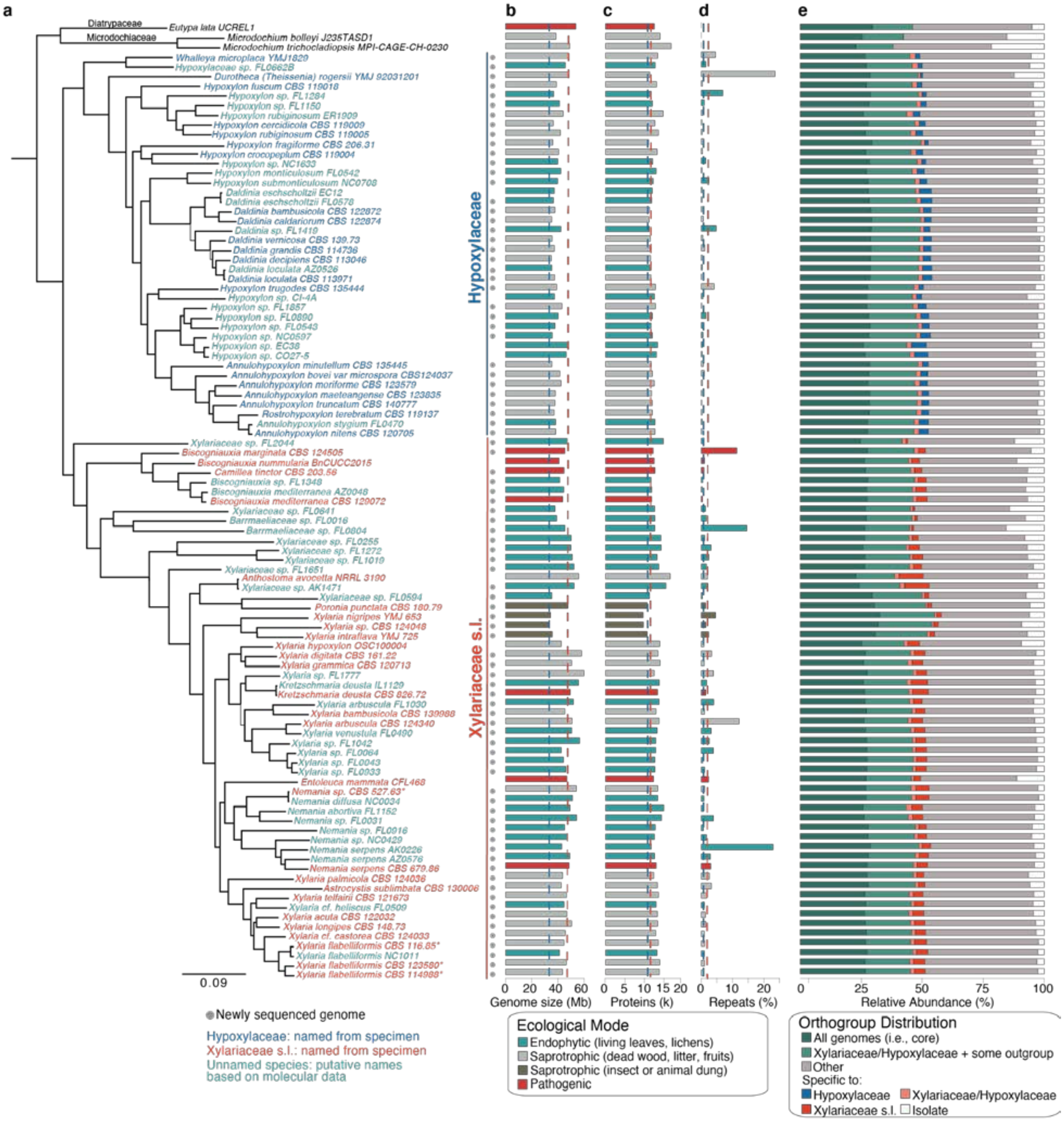

Fig. S3. **Phylogenomic reconstruction of Xylariaceae s.l. and Hypoxylaceae and genome statistics.** (**a**) The maximum likelihood phylogram is based on 1,526 single-copy orthologous genes present in all genomes. Bootstrap values are shown in Fig. S2. The scale bar indicates the number of substitutions per site. Names of reference taxa are colored according to their clade affiliation (dark blue: Hypoxylaceae; red: Xylariaceae s.l.). Undescribed endophyte species, putatively named based on phylogenetic analyses (U'Ren *et al.*, 2016), are shown in teal blue; (**b**) genome size; (**c**) predicted protein coding genes; and (**d**) percent transposable element (TE) content (bar colors correspond to ecological mode; see legend). Averages per major clade are shown with dotted lines in panels a-d; (**e**) relative abundance of core, family-specific, clade-specific, and isolate-specific orthogroups (see legend; Table S3d).

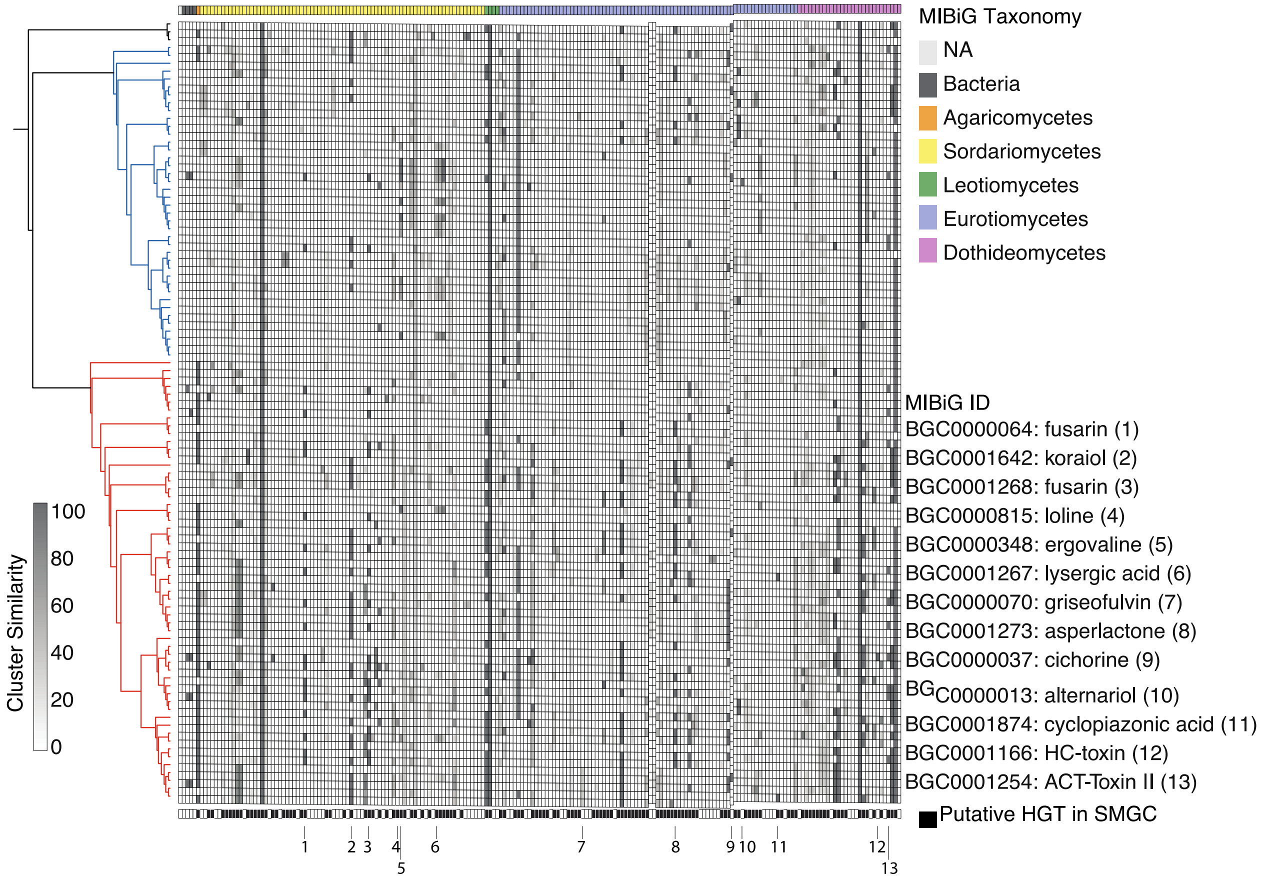

Fig S4. Dynamic distribution of 168 Xylariaceae and Hypoxylaceae SMGCs with hits to known metabolites in the MIBiG repository. Rows are sorted by the taxonomic identity (class and species) of the best MIBiG hit (top). Shading indicates the similarity of predicted SMGCs to reference metabolites, defined as the percentage of genes in an SMGC with significant BLAST hits to a known SMGC in the MIBiG database (Medema *et al.*, 2011). Black boxes (bottom) indicate SMGCs predicted by Alien Index (Wisecaver *et al.*, 2016; Verster *et al.*, 2019) to contain at least one gene putatively transferred via HGT (Table S5). For MIBiG clusters that occurred more than once per genome, only the hit with the highest similarity is shown (Table S3).

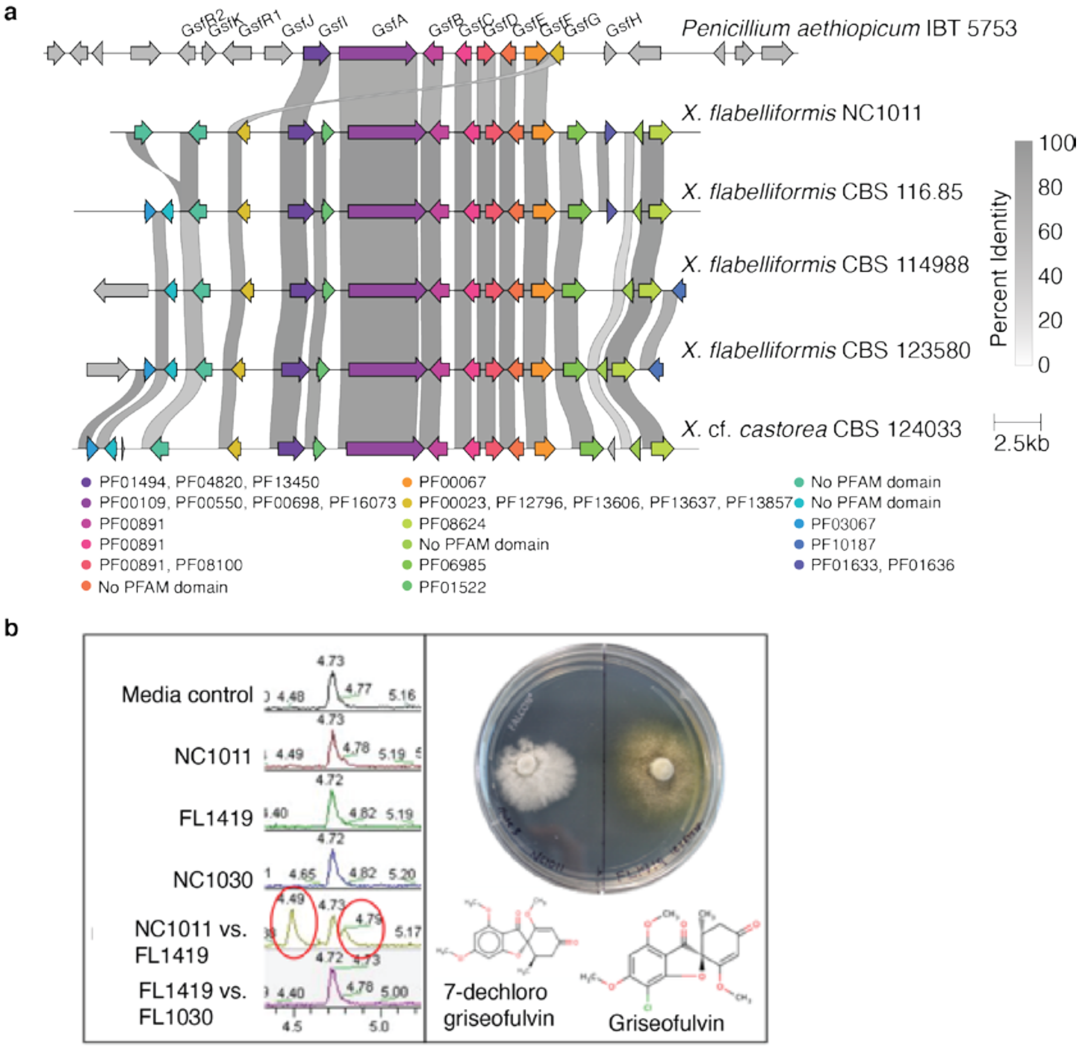

Fig. S5. Similarity of the griseofulvin SMGC in *Penicillium* and *Xylaria* supports HGT. (a) Comparison of the griseofulvin cluster from *Penicillium aethiopicum* IBT 5753 (top) to five newly sequenced *Xylaria* genomes. Homologous genes are colored by PFAM domain. Connecting ribbons indicate percent amino acid identity to genes in the *Penicillium* cluster; (**b**) Metabolomic analysis of pairwise comparisons of *X. flabelliformis* NC1011, *Xylaria arbuscula* FL1030, and *Daldinia* sp. FL1419 illustrates production of griseofulvin by NC1011 during the interaction with FL1419, but not when grown alone or with isolate FL1030.

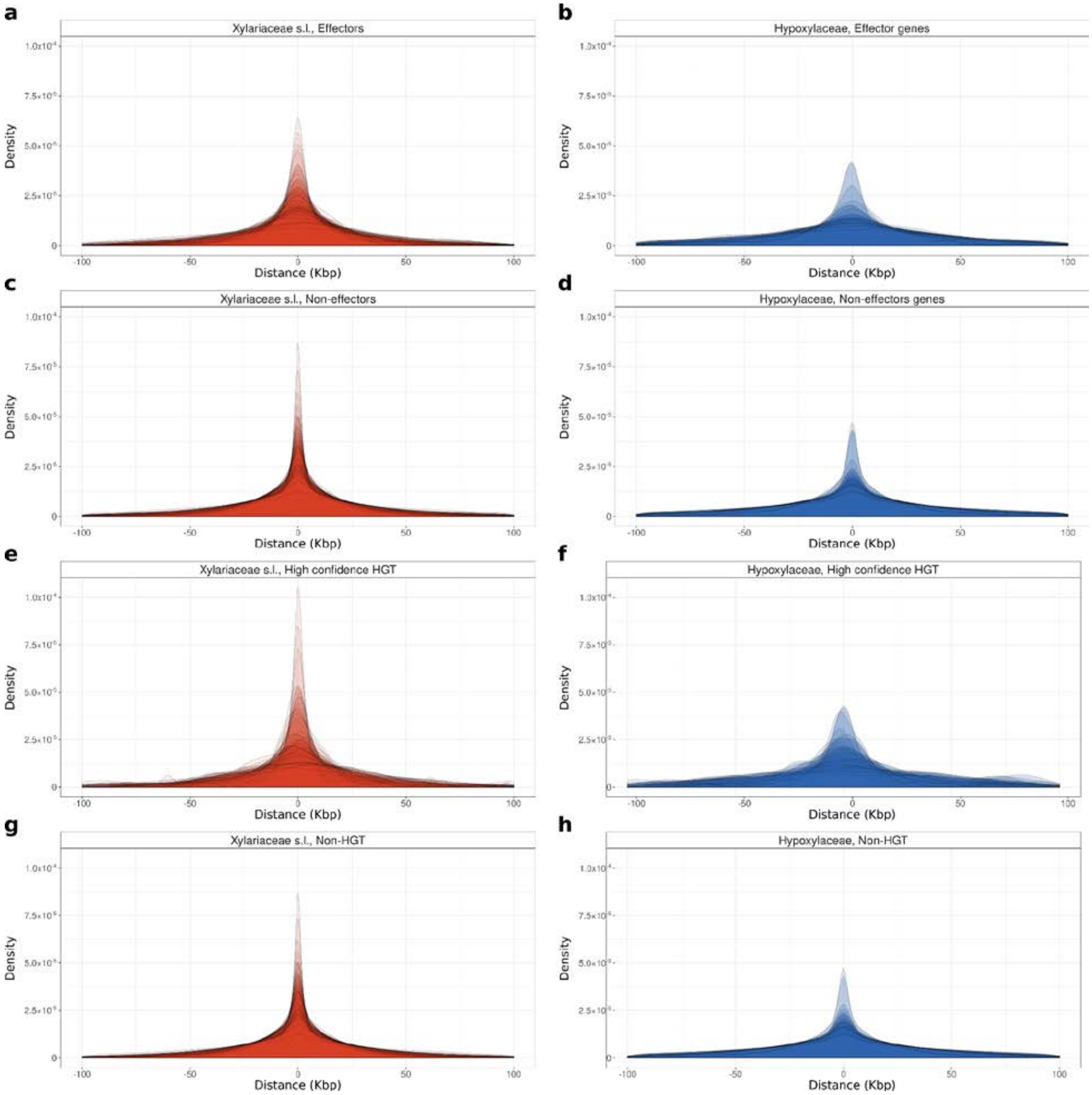

Fig. S6. The density of repetitive elements surrounding genes was higher for Xylariaceae s.l. than for Hypoxylaceae genomes. Overlapped density plot of all genomes in each clade (red: Xylariaceae s.l.; blue: Hypoxylaceae), illustrating the distance of the nearest repetitive elements from genes in the following categories: (**a, b**) effectors, (**c, d**) non-effector genes; (**e, f**) high confidence HGT candidate genes, and (**g, h**) non-HGT genes. Negative distances indicate that repetitive elements are located upstream of genes, while positive distances indicate repetitive elements downstream. Repetitive elements were identified by RepeatScout and RepeatMasker. Effector genes were predicted by EffectorP 2.0. High confidence HGT candidates were predicted using the first Alien Index analysis. Distances were computed using BEDTools v2.29.2.

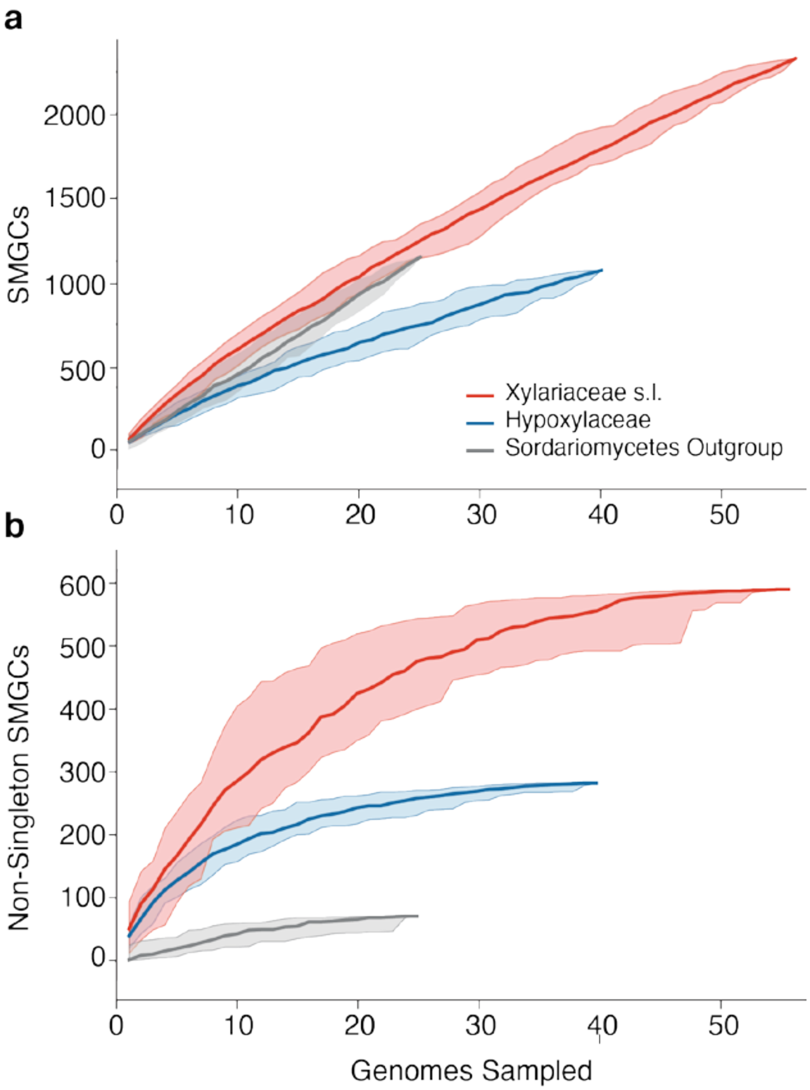

Fig. S7. Rarefaction analysis illustrates higher SMGC diversity in Xylariaceae compared to Hypoxylaceae. Rarefaction curves of **(a)** all SMGCs and **(b)** non-singleton SMGCs by clade (Xylariaceae s.l. Hypoxylaceae, and Sordariomycetes outgroup). Comparison of rarefaction curves for all SMGCs vs. non-singleton SMGCs illustrates the high number of singleton SMGCs present in the outgroup, which is consistent with the phylogenetic diversity of outgroup genomes that span 13 orders of Sordariomycetes. In contrast, richness of non-singleton SMGCs is ca. 4-7X greater for Xylariaceae and Hypoxylaceae genomes compared to outgroup genomes (n = 71 SMGCs).

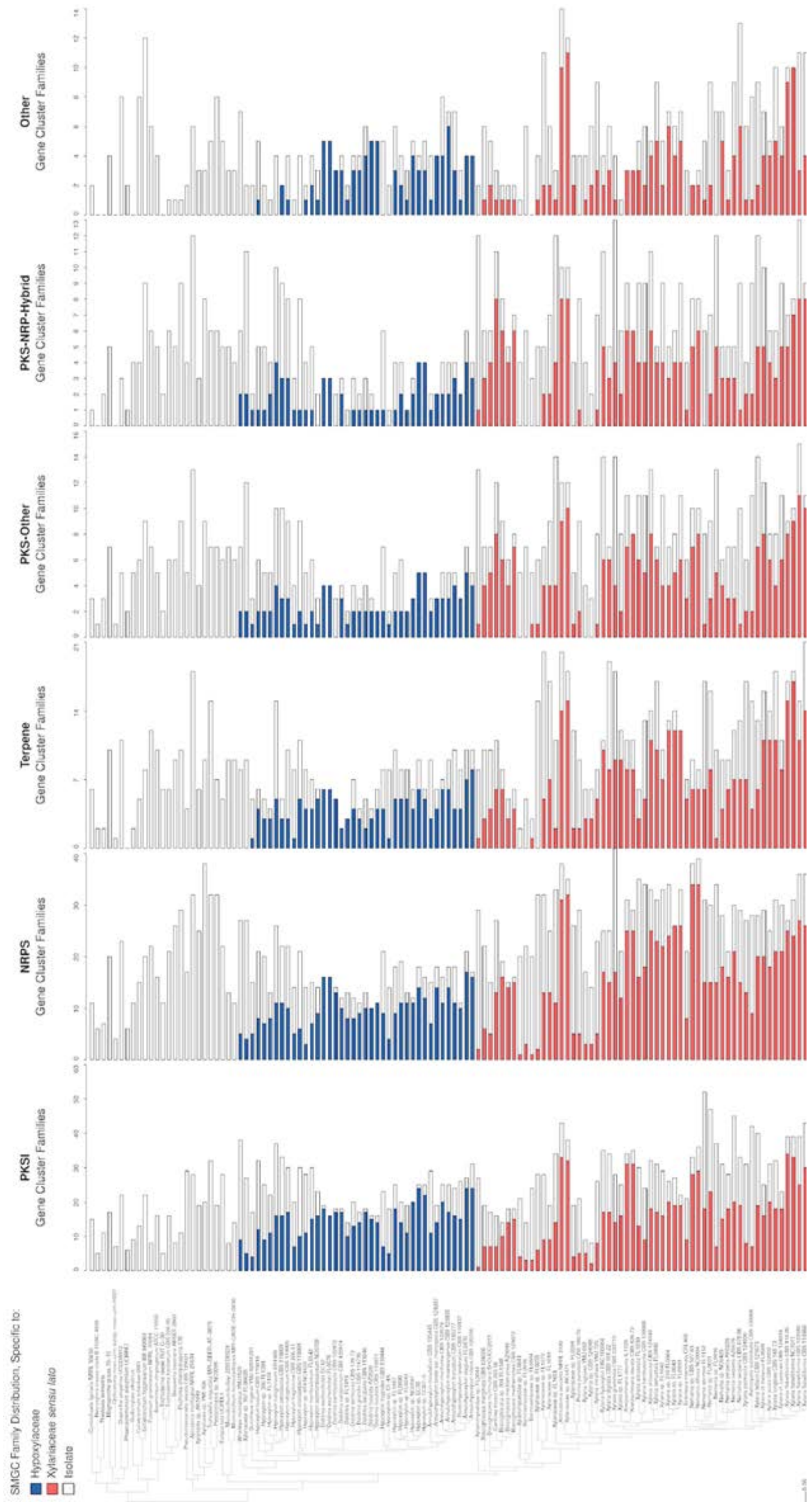

Fig. S8. The majority of SMGCs are specific to Hypoxylaceae or Xylariaceae s.l. clades or individual isolates regardless of SMGC type. Phylogenomic tree of Xylariaceae s.l. and Hypoxylaceae and outgroup taxa with bar plots illustrating the number of SMGC families per genome, as well as the percentage of clade-specific and isolate-specific SMGC families for (**a**) PKSI; (**b**) NRPS; (**c**) Terpene; (**d**) PKS-Other; (**e**) PKS-NRP Hybrid; and (**f**) Other.

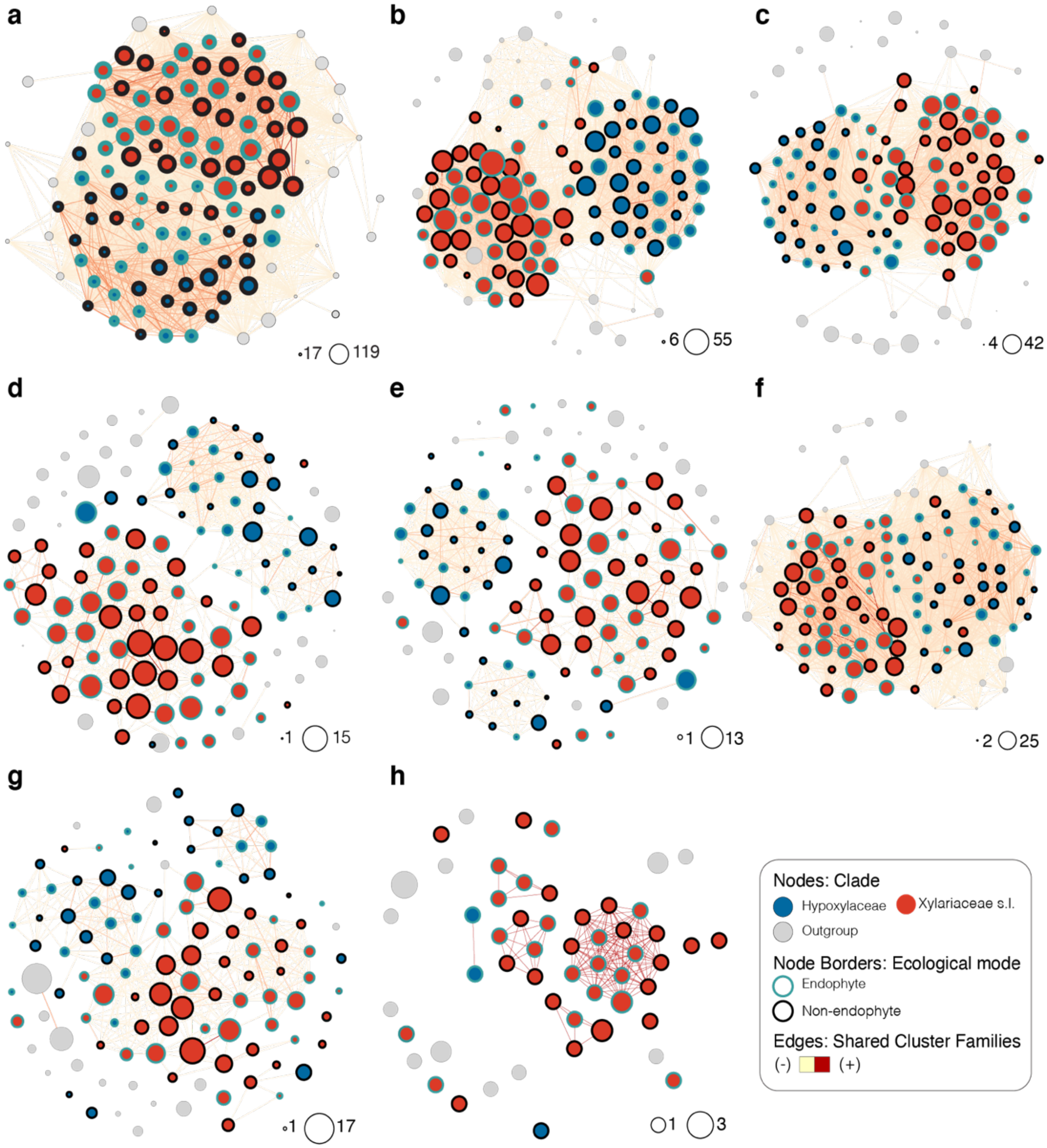

Fig. S9. Network analysis illustrates the importance of clade rather than ecological mode for SMGC content. Network representation of SMGCs clustering from BiG-SCAPE. Each node represents the SMGC content per genome for (**a**) all SMGCs and SMGC sub-types:(**b**) PKSI; (**c**) NRPS; (d) PKS other; (e) PKS-NRPS Hybrids; (f) terpenes; (g) other; and (h) RiPPs. Networks edited with Gephi v0.9.1 (Bastian *et al.*, 2009), wherenodes were scaled by the count of gene clusters and positioned by a force-directed layout algorithm (as described by ref(Laetsch & Blaxter, 2017)). Edges between two nodes are weighted by the number of shared clusters. Node color corresponds to clade. Nodes representing endophytic isolates are shown with blue borders. To compare the distribution of all SMGC families (**a**), BiG-SCAPE families representing different SMGC types were combined into a single dataset and SMGCs assigned to multiple families were arbitrarily assigned to the largest family.

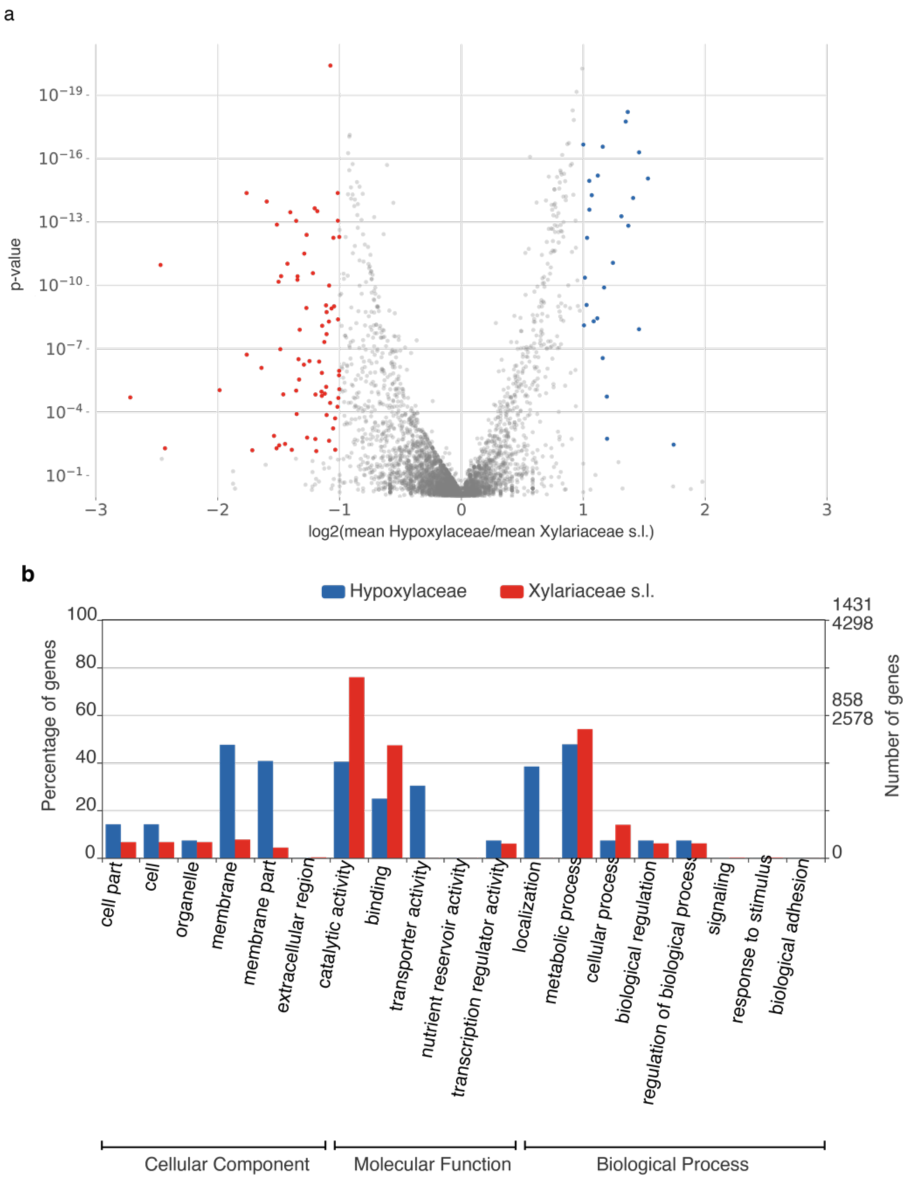

Fig. S10. Orthogroup enrichment suggests functional differences for Xylariaceae and Hypoxylaceae. Twenty-six orthogroups were significantly enriched in the Hypoxylaceae clade, while 74 orthogroups were significantly expanded in the Xylariaceae s.l. clade. (**a**) Volcano plot of the protein count representation tests for orthogroups shared between the Hypoxylaceae and Xylariaceae s.l. clades. Orthogroups significantly enriched in Xylariaceae s.l. taxa are colored in red, while orthogroups significantly enriched in Hypoxylaceae taxa are colored in blue. Two-sided Mann-Whitney U-tests, p-value ≤ 0.01 and |log2FC| ≥ 1. (**b**) Comparison of enriched GO terms (level 2) of orthogroups significantly enriched in Hypoxylaceae taxa (blue) vs. Xylariaceae s.l. taxa (red). GO terms were analyzed and visualized using Web Gene Ontology Annotation Plot 2.0 (WEGO). See also Table S3f for KOG annotation of enriched orthologs. The two-sided Mann-Whitney U-test was performed using SciPy (Virtanen *et al.*, 2020) through KinFin v1.0 (Laetsch & Blaxter, 2017).

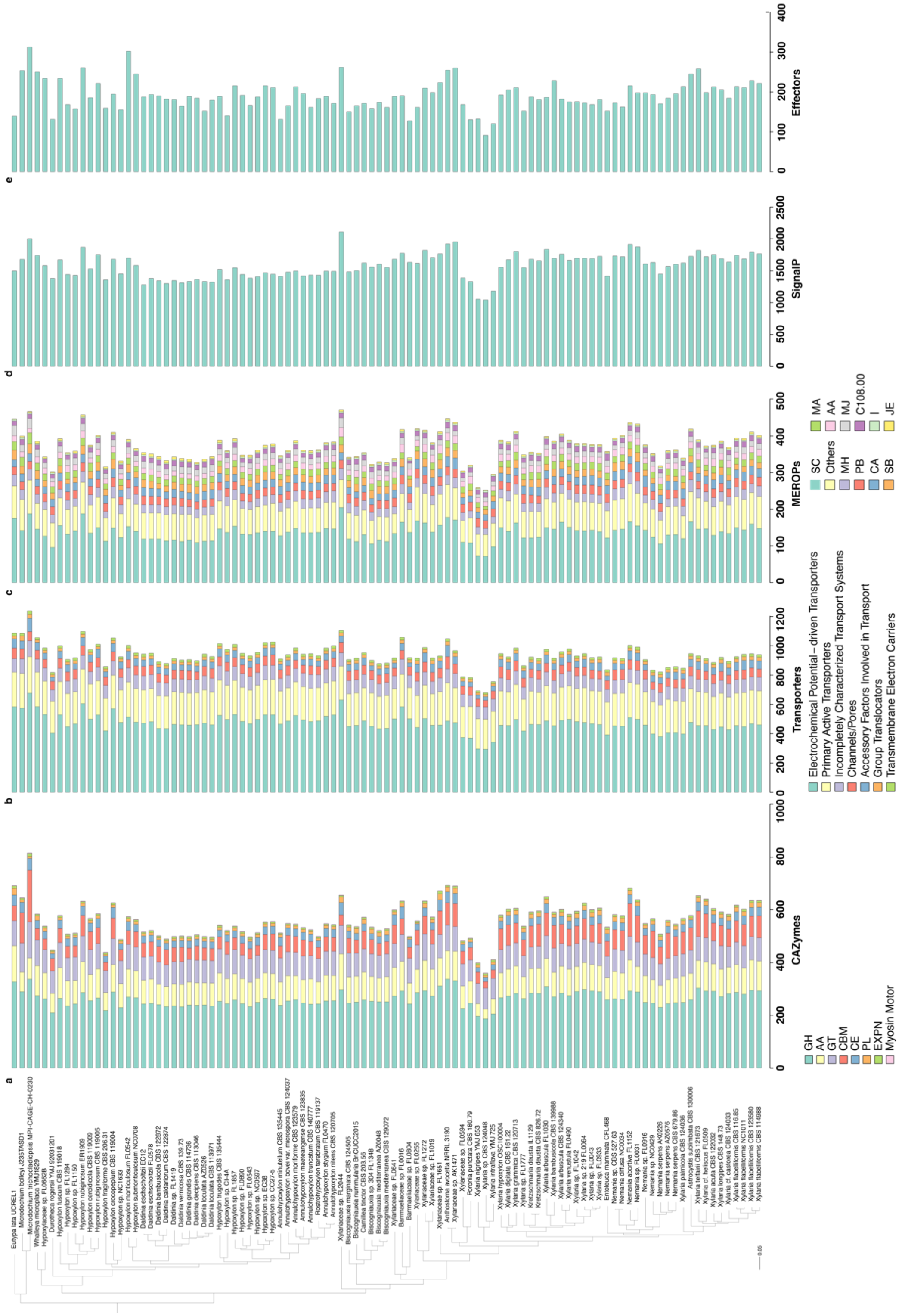

Fig. S11. Relative abundance of functional gene categories across Xylariaceae s.l. and Hypoxylaceae. Phylogenomic tree and bar plot showing the abundance and identity of (**a**) carbohydrate-active enzymes (CAZyme); (**b**) peptidases and their inhibitors (MEROPs); (**c**) transporters (TCDB); (**d**) secreted proteins (SignalP); and (**e**) effectors (EffectorP). Colors refer to different classifications within each database (see legends).

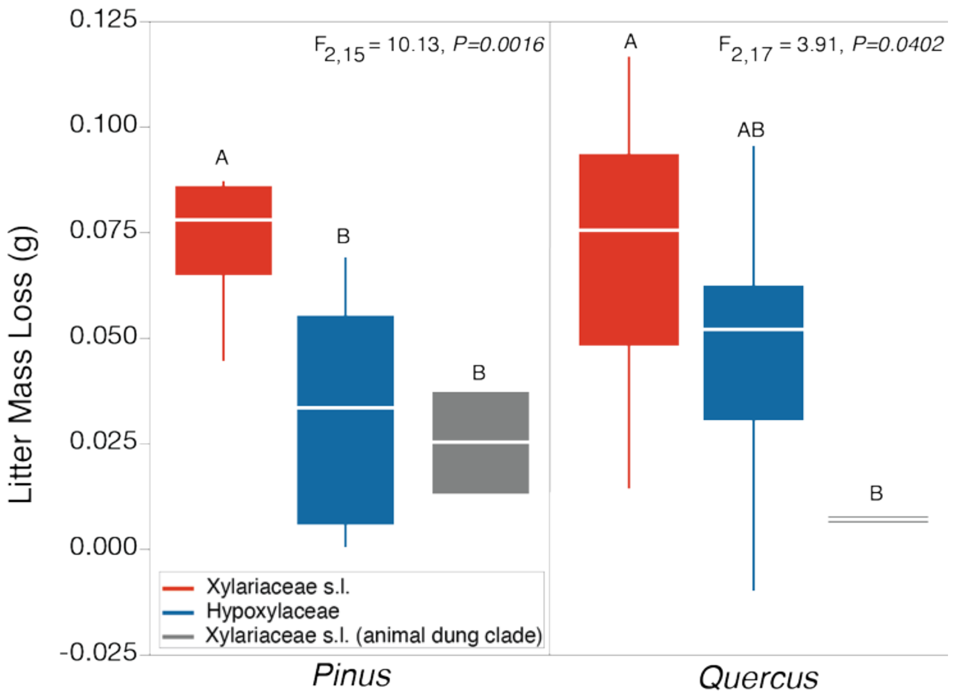

Fig. S12. Xylariaceae s.l. taxa demonstrate increased decomposition abilities (estimated via mass loss) on leaf litter compared to fungi with reduced genomes (i.e., Hypoxylaceae and animal dung Xylariaceae s.l. in the *Poronia* clade). Interquartile box plots showing median and interquartile range. We observed significant differences among means of each clade on both *Pinus* and *Quercus* leaves (ANOVA). Letters indicate significant differences after post-hoc Tukey's HSD. See Table S1 for a list of isolates included in the mass loss experiment.

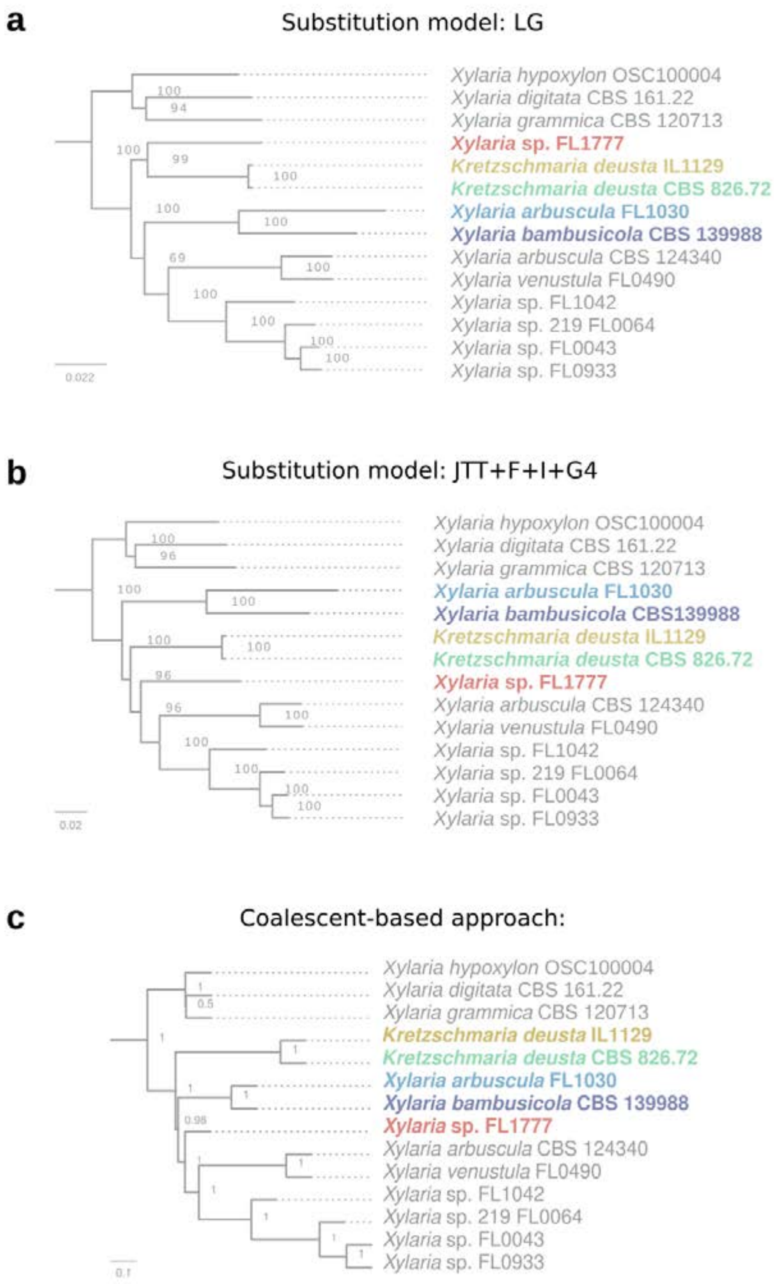
a
Substitution model: LG
Xylaria hypoxylon OSC100004
Xylaria digitata CBS 161.22
Xylaria grammica CBS 120713
Xylaria sp. FL1777
Kretzschmaria deusta IL1129
Kretzschmaria deusta CBS 826.72
Xylaria arbuscula FL1030
Xylaria bambusicola CBS 139988
Xylaria arbuscula CBS 124340
Xylaria venustula FL0490
Xylaria sp. FL1042
Xylaria sp. 219 FL0064
Xylaria sp. FL0043
Xylaria sp. FL0933
100
94
100
99
100
100
100
69
100
100
100
100
100
0.022
b
Substitution model: JTT+F+I+G4
Xylaria hypoxylon OSC100004
Xylaria digitata CBS 161.22
Xylaria grammica CBS 120713
Xylaria arbuscula FL1030
Xylaria bambusicola CBS139988
Kretzschmaria deusta IL1129
Kretzschmaria deusta CBS 826.72
Xylaria sp. FL1777
Xylaria arbuscula CBS 124340
Xylaria venustula FL0490
Xylaria sp. FL1042
Xylaria sp. 219 FL0064
Xylaria sp. FL0043
Xylaria sp. FL0933
100
96
100
100
100
100
96
96
100
100
100
100
100
0.02
c
Coalescent-based approach:
Xylaria hypoxylon OSC100004
Xylaria digitata CBS 161.22
Xylaria grammica CBS 120713
Kretzschmaria deusta IL1129
Kretzschmaria deusta CBS 826.72
Xylaria arbuscula FL1030
Xylaria bambusicola CBS 139988
Xylaria sp. FL1777
Xylaria arbuscula CBS 124340
Xylaria venustula FL0490
Xylaria sp. FL1042
Xylaria sp. 219 FL0064
Xylaria sp. FL0043
Xylaria sp. FL0933
1
0.5
1
1
1
1
0.98
1
1
1
1
1
0.1

Fig. S13. Phylogenetic tree topology was similar regardless of methodology, except for relationships among taxa in the Xylaria HY and E9 clades. Subclade topology for three phylogenetic analyses: concatenated analysis of 1,526 single-copy orthogroups with the (**a**) LG model of evolution (i.e., analysis 1; see also Fig. S2) or (**b**) JTT + F + I + G4 model of evolution (i.e., analysis 2); and (**c**) ASTRAL coalescent analysis of gene trees.

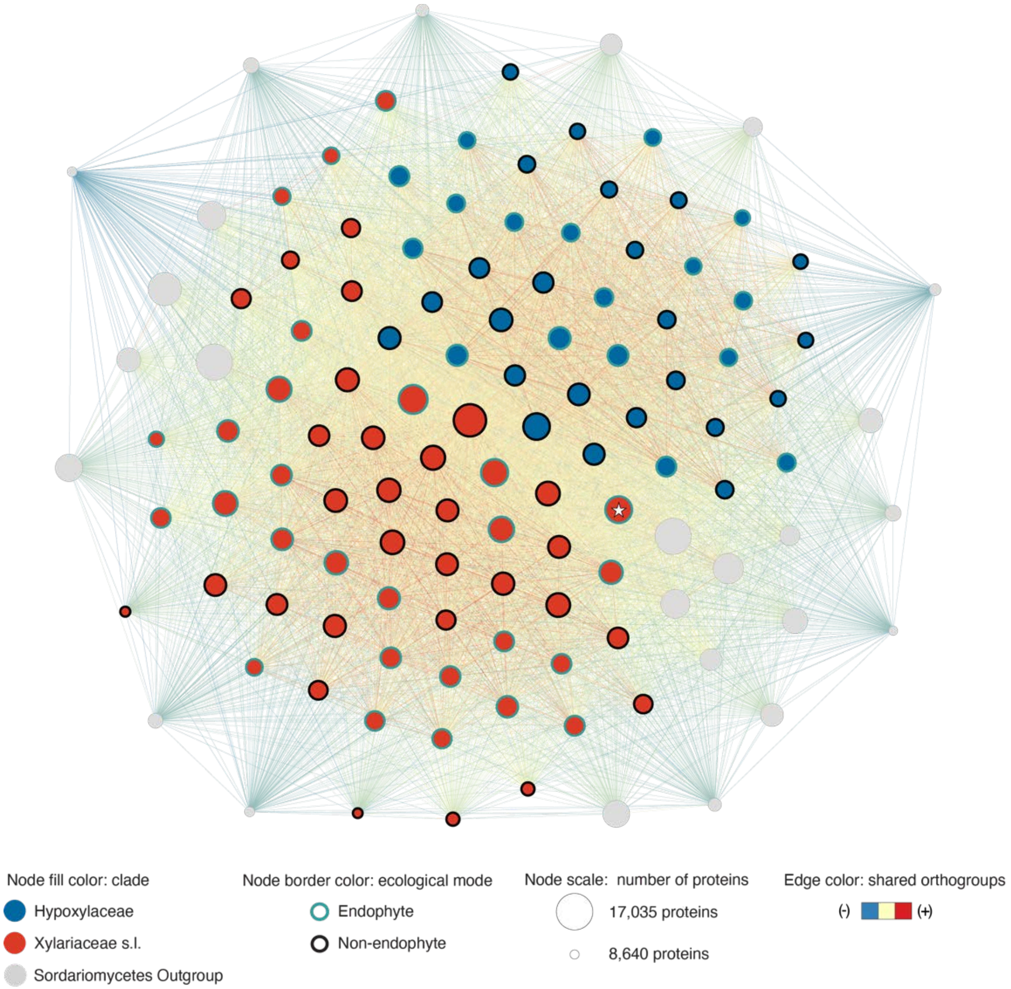

Fig. S14. Network analysis of individual proteomes illustrates the importance of major clade affiliation. Proteomes are represented by nodes, scaled by the count of proteins, colored by clade (fill) and ecological mode (border), and positioned by a force-directed layout algorithm. Edges between two nodes are weighted by the number of shared orthogroups. The node with a star represents Xylariaceae sp. FL2044.

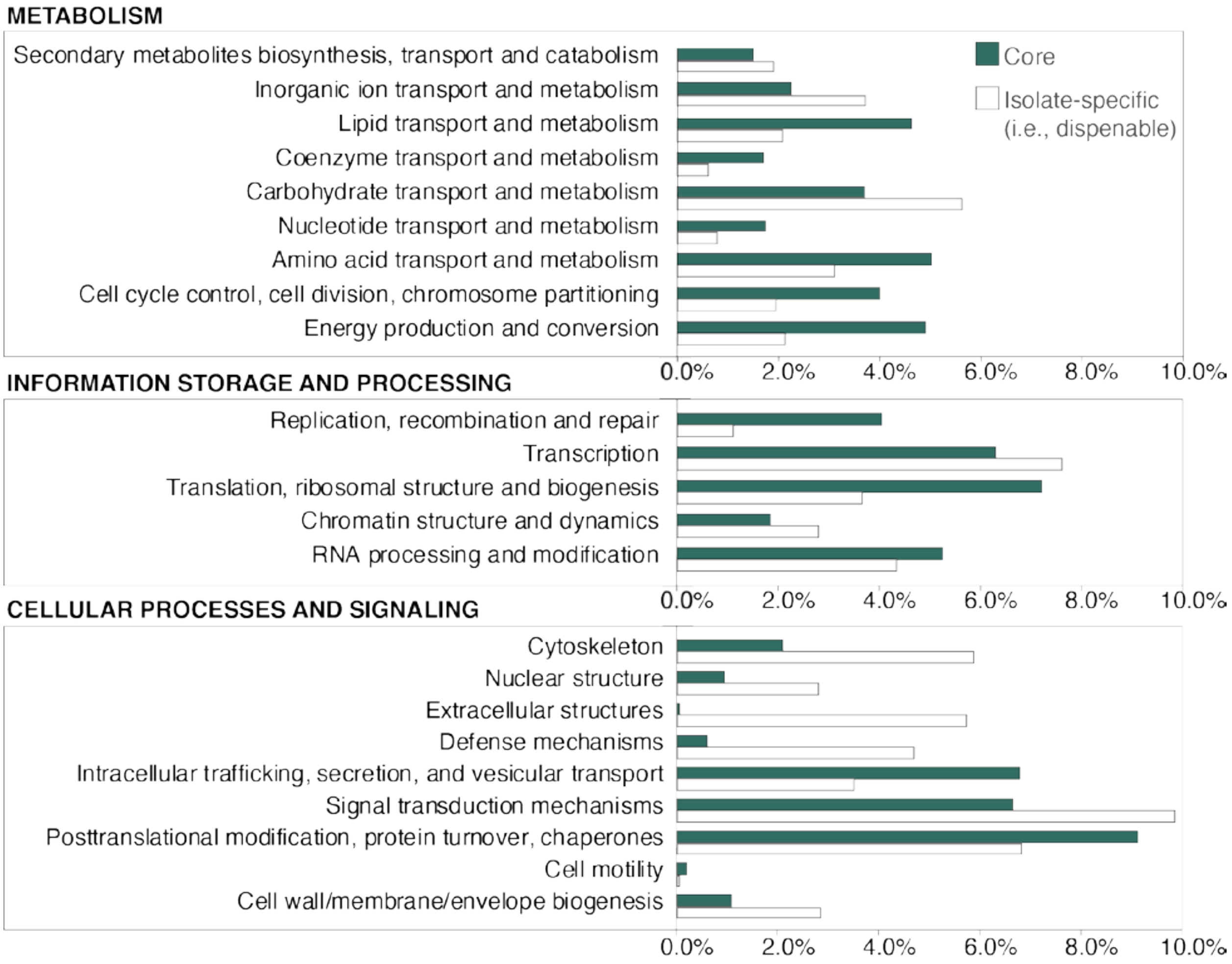

Fig S15. Comparison of functional annotations for core and dispensable orthogroups. Bar graphs showing the relative abundance of different functional categories represented by “core” vs. “dispensable” orthogroups. Orthogroups were annotated with euKaryotic Orthologous Groups (KOGs; see Table S7f).

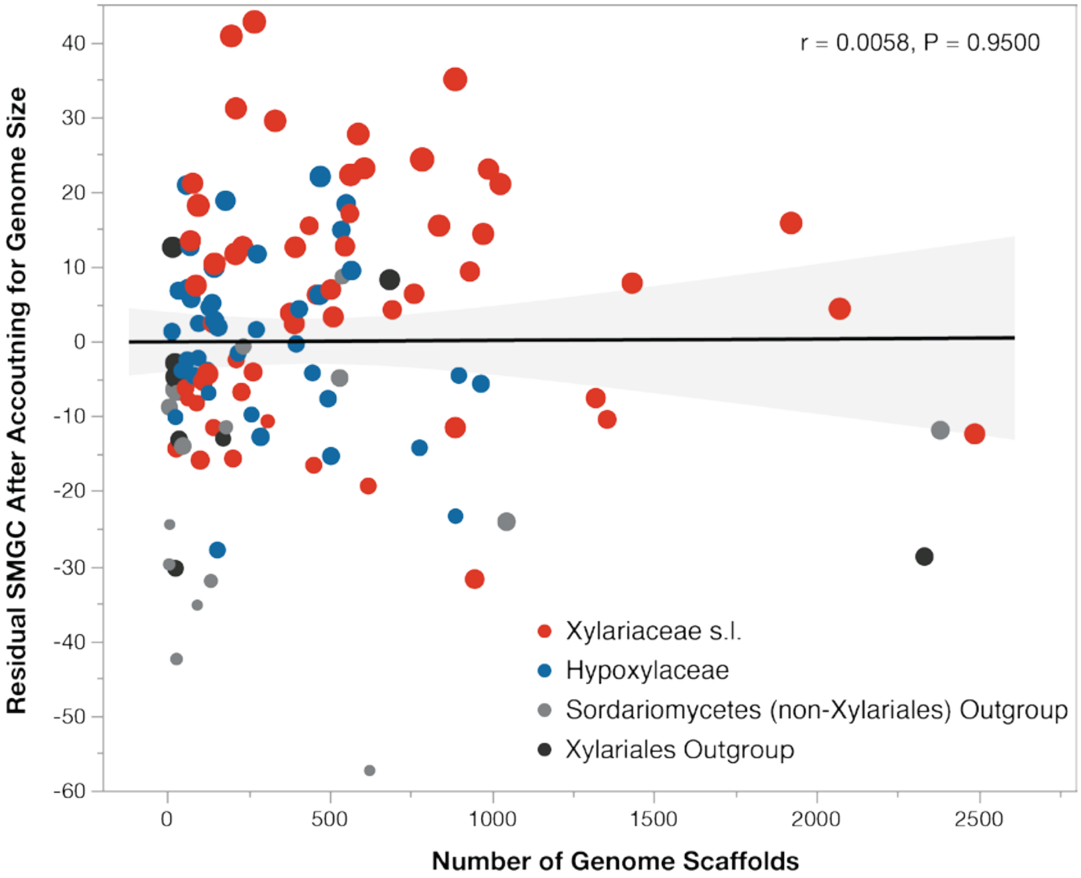

Fig S16. **The number of predicted SMGCs is not related to genome assembly.** Relationship of predicted SMGC content (residuals after accounting for genome size) and the number of scaffolds for 121 genomes. Points are colored by clade and their size is proportional to the raw number of SMGC per genome (range 16-119).