nature ecology & evolution

Article

https://doi.org/10.1038/s41559-025-02785-6

Cross-kingdom comparative genomics reveal the metabolic potential of fungi for lignin turnover in deadwood

Received: 26 February 2025

Accepted: 4 June 2025

Published online: 9 July 2025

Check for updates

Teeratas Kijpornyongpan[1], Eugene Kuatsjah[1,4], Evan Komp[1], James E. Evans[2], Francisco Javier Ruiz-Dueñas[3] & Davinia Salvachúa[1]

Deadwood is a major carbon source in forests, and yet the fate of this carbon remains a gap in our understanding of global carbon cycling. Lignin, the most recalcitrant biopolymer in wood, is mainly decayed through extracellular enzymatic and chemical processes initiated by white-rot fungi. However, the intracellular conversion of lignin decay products has been overlooked in the fungal kingdom. Here we integrate comparative genomic and phylogenetic analyses to understand the distribution and evolution of enzymes responsible for modifying lignin-related aromatic compounds—such as decarboxylases, hydroxylases, dioxygenases and other downstream ring-cleavage enzymes—that funnel carbon to central metabolism across the bacterial and the fungal kingdoms. We demonstrate that specific fungal lineages conserve these enzyme families, and that the abilities to enzymatically depolymerize lignin and catabolize lignin-related aromatic compounds are not necessarily coupled. Our analyses also reveal an expanded substrate specificity of aromatic ring-cleavage enzymes during fungal evolution, as well as a clade of extracellular enzymes among them, broadening the spatial range of these biochemical capabilities. Together, our results highlight a large diversity of fungal enzymes and hosts that warrant further investigation for inclusion into carbon cycling models and biotechnological applications for the conversion of aromatic compounds.

Wood decay plays a critical role in global carbon cycling. Approximately 28% of the carbon in terrestrial ecosystems is attributed to plant biomass[1,2], of which 8% (73 ± 6 Pg) resides in deadwood[3,4]. Lignin is the second most abundant plant biopolymer[5] and is a major source of long-lasting carbon in deadwood. In forests, lignin decay is largely initiated by ligninolytic enzymes (for example, lignin-oxidizing peroxidases and laccases) secreted by white-rot fungi (WRF)[6,7]. The emergence of this unique capability in WRF occurred ~295 million years ago and possibly contributed to the end of the Carboniferous period[6]. WRF peroxidases, the key lignin-degrading enzymes, continued to evolve since then, which has been correlated to increasing lignin complexity in plants[8]. In addition, the pool of enzymes secreted by wood decay fungi—including WRF and other non-lignin-degrading fungi[7,9]—has evolved over time to target specific plant species[10–12]. Together, these insights have advanced our understanding of the biological processes involved in wood turnover in forests.

The lignin polymer is heterogeneous and composed of dimethoxylated (syringyl (S)-type), monomethoxylated (guaiacyl (G)-type),

[1]Renewable Resources and Enabling Sciences Center, National Renewable Energy Laboratory, Golden, CO, USA. [2]Environmental and Molecular Sciences Division, Pacific Northwest National Laboratory, Richland, WA, USA. [3]Centro de Investigaciones Biológicas Margarita Salas (CIB), CSIC, Madrid, Spain. [4]Present address: Institute of Sustainability for Chemicals, Energy and Environment (ISCE2), Agency for Science, Technology and Research (A*STAR), Singapore, Republic of Singapore. e-mail: davinia.salvachua@nrel.gov

https://doi.org/10.1038/s41559-025-02785-6

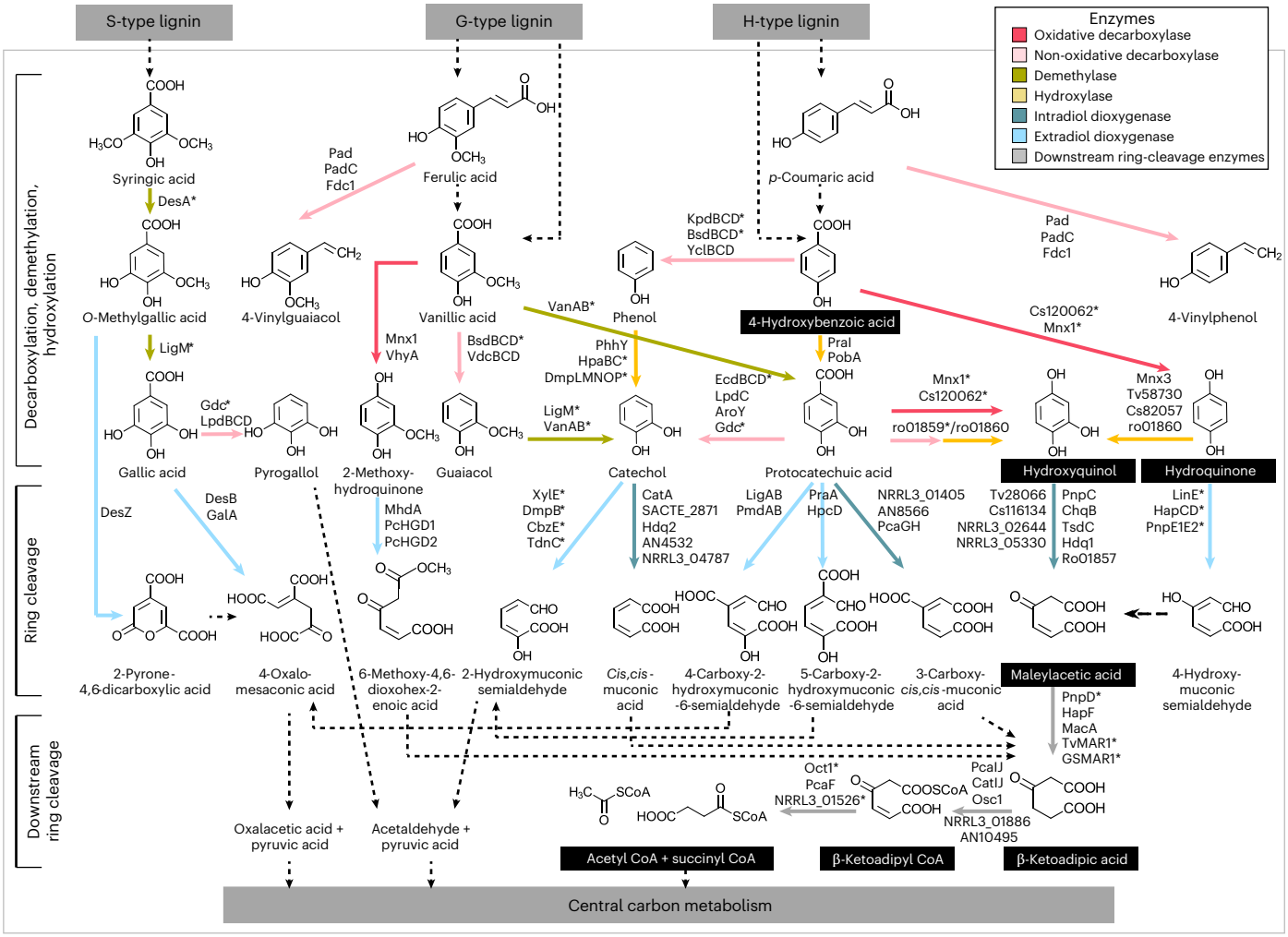

Fig. 1 | Simplified scheme of pathways and enzymes involved in the conversion of lignin-related aromatic compounds in fungi and/or bacteria towards central carbon metabolism. Catabolic pathways are shown for common lignin-related aromatic compounds: non-methoxylated *p*-coumaric acid and 4-hydroxybenzoic acid (derived from H-type lignin), monomethoxylated ferulate and vanillate (derived from G-type lignin) and dimethoxylated syringic acid (derived from S-type lignin). Reference enzymes included in this study are indicated at each biochemical step and the enzymes lacking proxy IPR domains are marked with an asterisk. Reference enzymes involved in the conversion of phenylpropanoids[74], salicylic acid[75] and benzoic acid[76], which are aromatic compounds derived from plants, were also included in this study (Supplementary Tables 6–9) but are not shown in this scheme to facilitate the visualization. IPR ID and IPR description for all the reference enzymes are included in Table 1. Additional details about these reference enzymes, including enzyme description, substrate specificity, associated microbial species, identified domains and the source publications, can be found in Supplementary Tables 6–10 and 12. Black boxes highlight the metabolites in the hydroxyquinol pathway, a recently proposed[18] and validated pathway (down to βKA)[21] for the catabolism of 4-hydroxybenzoic acid in WRF. Dashed lines indicate several reactions that are not considered in this study.

non-methoxylated (*p*-hydroxyphenyl (H)-type) phenylpropanoid units[5], along with other recently discovered monomers[13], and its decay can produce monomeric aromatic compounds[14,15] (Fig. 1). While intracellular conversion and use of aromatic compounds as carbon sources has been extensively documented for bacteria[14,16], research on this capability is limited in fungi[17]. Indeed, the ability of WRF to catabolize lignin-related aromatic compounds (that is, by *Trametes versicolor* and *Ceriporiopsis* (*Gelatoporia*) *subvermispora*)[18] and lignin (by *Agaricus bisporus*)[19] as carbon sources was only recently described. The intracellular catabolic pathways involved in the conversion of aromatic compounds to central carbon metabolism are diverse (Fig. 1). The β-ketoadipate (βKA) pathway is a well-established catabolic route for H- and G-type aromatic compounds[20], whereas the hydroxyquinol pathway (Fig. 1) has been only reported in few bacteria and in fungi[21]. We also note that the latter pathway was recently validated in one WRF through gene-function studies, specifically for the catabolism of H-type aromatic compounds[21]. Knowledge of intracellular catabolic pathways for S-type units is generally restricted to bacteria[17]; however, recent findings suggest that fungi may use extracellular polyphenol oxidases to demethoxylate S-type substrates[22]. Despite substantial pathway diversity, there are common enzyme reactions involved in processing monomeric aromatic compounds, including decarboxylation, hydroxylation, aromatic *O*-demethylation and ring cleavage[23] (Fig. 1). Our understanding of the distribution and evolution of these aromatic catabolic pathways and enzymes remains limited, especially in the fungal kingdom. This knowledge gap highlights the need to assess the metabolic potential of fungi for both lignin breakdown and downstream turnover, which plays an important role in carbon sequestration as fungal biomass and CO_2 release during wood decay in forests[4,24,25].

Here, we integrate comparative genomics and phylogenetics to analyse the distribution, prevalence and evolution of crucial enzymes involved in the catabolism of H-, G- and S-type lignin-related aromatic compounds in the fungal and bacterial kingdoms. Although emphasis is placed on results from the Agaricomycetes class in fungi, which

comprises most wood-degrading fungi (including WRF), parallel analyses within kingdoms provide more comprehensive insights into enzyme origins. Overall, this work establishes a foundation for a deeper understanding of carbon flow from lignin degradation products in forests and the discovery of diverse biocatalysts, either microbial or enzymatic, in the emerging field of lignin valorization[26].

Results

Genome sampling and enzyme selection for lignin catabolism

To study the distribution and evolution of enzymes that catabolize lignin-related aromatic compounds in the bacterial and fungal kingdoms, we used the workflow in Supplementary Fig. 1. We sampled 255 bacterial and 317 fungal genomes (Supplementary Tables 1–4), across different lineages and nutritional modes (phylogenetic trees in Supplementary Figs. 2 and 3 and Supplementary Data 1, which includes high-resolution figures from the main text and Supplementary Information). In the fungal kingdom, 73 genomes belong to the class Agaricomycetes. This class encompasses most wood decay fungi, including WRF (cellulose, hemicellulose and lignin degraders), brown-rot fungi (cellulose and hemicellulose degraders) and grey-rot fungi (uncertain classification between WRF and brown-rot fungi)[9] (Supplementary Table 5). Soft-rot fungi can decay wood polysaccharides[7] and are included in various groups, including the Pezizomycotina subphylum. Then, we selected 'reference enzymes' known experimentally to modify lignin-related aromatic compounds (Supplementary Tables 6–10). Overall, 32 decarboxylases (28 non-oxidative and 4 oxidative), 25 hydroxylases, 6 *O*-demethylases and 46 aromatic ring-cleavage dioxygenases (17 intradiol and 29 extradiol) were selected (Fig. 1 and Table 1). Given the limited information on the hydroxyquinol pathway and its existence in WRF[18,21] (Fig. 1 and Table 1), we chose 16 enzymes downstream hydroxyquinol ring cleavage towards central metabolism (hereafter called 'downstream ring-cleavage enzymes'), including maleylacetate reductases, βKA succinyl-CoA transferases and β-ketoadipyl CoA thiolases (Fig. 1 and Table 1). To explore associations between lignin decay (depolymerization) and the catabolism of lignin-related compounds, we also selected 17 lignin-oxidizing enzymes (Supplementary Table 11), including peroxidases and laccases[7].

The protein sequences of all the 'reference enzymes' were then examined for domain composition through InterProScan[27] (known as IPR (InterPro) domains) (Supplementary Tables 6–11). The IPR domains were mapped to show gene abundance in bacteria, fungi and Agaricomycetes (Supplementary Figs. 4–6 for dioxygenases, Supplementary Figs. 7–9 for decarboxylases, Supplementary Figs. 10–12 for hydroxylases, Supplementary Figs. 13–15 for *O*-demethylases and Supplementary Figs. 16–18 for downstream ring-cleavage enzymes). Some of these IPR domains lack description related to enzyme activities on lignin-related aromatic compounds and were highly conserved with high gene copy numbers in fungi and bacteria (that is, IPR domains corresponding to oxidative decarboxylases and *O*-demethylases; Supplementary Tables 7, 9 and 12 and Supplementary Figs. 7–9 and 13–15). This suggested the lack of specificity of these IPR domains in the catabolism of lignin-related compounds. Thus, we further down selected IPR domains (hereafter proxy IPR domains) that represent specific catabolic activities (Table 1 and Methods) to continue the analyses.

Proxy IPR domains are conserved in the fungal kingdom

Proxy IPR domains were used to assess the distribution (prevalence and abundance) of putative enzymes involved in the conversion of lignin-related aromatic compounds across the fungal and bacterial kingdoms (Table 1, Fig. 2 and Supplementary Table 12). Certain domains are conserved in over 50% of fungal genomes (Fig. 2a) and Agaricomycetes (Supplementary Fig. 19), but not in bacteria (Fig. 2b), including putative hydroxylases (with IPR012941), dioxygenases (with IPR000627, IPR007535 and IPR004183) and downstream ring-cleavage enzymes (with IPR012791). In bacteria, only UbiD–UbiX-type non-oxidative decarboxylases (with IPR002830 and IPR004507) were conserved in over 50% of the genomes. Conversely, some proxy IPR domains are prevalent in few bacterial lineages (Alphaproteobacteria, Betaproteobacteria, Gammaproteobacteria and Actinobacteria) and absent in fungi such as certain intradiol and extradiol dioxygenases (with IPR024756 and IPR011986), hydroxylases (with IPR000391 and IPR012733) and the last enzyme in the hydroxyquinol pathway (with IPR012793, a β-ketoadipyl CoA thiolase) (Fig. 1 and Table 1). Notably, the absence of certain domains in the sampled fungal genomes does not necessarily indicate the lack of the corresponding enzyme activity. For instance, protocatechuate 3,4-dioxygenase has been characterized in some Ascomycetes (for example, *Aspergillus niger*)[28] despite lacking the domain IPR024756 (Table 1)—instead, these fungal dioxygenases contain IPR domains not specific to protocatechuate, such as IPR000627 and IPR007535 (Table 1). Another example is β-ketoadipyl CoA thiolases from fungi, of which IPRs are too general to serve as proxy domains (Table 1). Considering this potential limitation, and to correlate IPR domains with substrate specificity, we further integrate genomic with phylogenetic analyses (including selected enzymes with validated activities from fungi and/or bacteria). In general, this dataset suggests that while bacteria appear to exhibit a broader variety of proxy IPR domains related to lignin-related aromatic compound catabolism compared to fungi, the fungal kingdom shows higher conservation of these domains (when present) than do bacteria.

Proxy IPR domain distribution associates with fungal lineage

We then investigated whether the distribution of proxy IPR domains is correlated with microbial lineages and aimed to identify genomes with high domain abundance (Fig. 2a,b). While gene abundance does not directly correlate with the capacity to catabolize lignin-related aromatic compounds, it provides a means to discover enzymes and microbes for their conversion. In bacteria, we observed that proxy IPR domains are predominantly found in Alphaproteobacteria, Betaproteobacteria, Gammaproteobacteria and Actinobacteria (Fig. 2b and Supplementary Table 13) and that the genomes with the highest abundance of intradiol dioxygenases (10 genes), extradiol dioxygenases (11 genes), non-oxidative decarboxylases (12 genes) and hydroxylases (38 genes) corresponded to *Rhodococcus* spp. (Actinobacteria), *Sphingobium* sp. SYK-6 (Alphaproteobacteria), *Aromatoleum aromaticum* EbN1 (Betaproteobacteria) and *Sphingobium wittichii* RW1 (Alphaproteobacteria), respectively (Supplementary Table 14). Multivariate analyses revealed that the association between bacterial lineages and domain distribution is weak (Supplementary Fig. 20), indicating that proxy IPR domains are not primarily affected by bacterial lineages. However, the phylogenetic tree revealed that the domains were often conserved at the genus level (Fig. 2b and Supplementary Data 2).

In the fungal kingdom, proxy IPR domains are conserved in most filamentous fungi (Pezizomycotina and Agaricomycotina) but less so in yeast-like lineages (Saccharomycotina, Taphrinomycotina and Ustilaginomycotina) and early-diverging fungi (Fig. 2a and Supplementary Table 15). Some of these enzymes (for example, dioxygenases and hydroxylases) require oxygen, which partly explains their absence in clades that include anaerobic fungi (early-diverging fungi)[29]. Ascomycota generally exhibit a higher abundance of enzymes with proxy IPR domains than do Basidiomycota. In addition, the proxy IPR domains for non-oxidative decarboxylases IPR002830 and IPR004507 (UbiD–UbiX type) are mostly found in Ascomycota (Fig. 2a and Supplementary Table 15). Genomes with the highest abundance of intradiol dioxygenases (14 genes), extradiol dioxygenases (14 genes), non-oxidative decarboxylases (9 genes) and hydroxylases (18 genes), were *Armillaria* spp. (Agaricomycotina), *Cadophora* sp. DSE1049 (Pezizomycotina), *Fusarium verticillioides* (Pezizomycotina) and *Nectria haematococca* (Pezizomycotina), respectively. Conversely to bacteria, multivariate analyses revealed strong association between lineages and the distribution of the proxy IPR domains (Supplementary Fig. 20), indicating that

Table 1 | Proxy IPR domains of enzymes related to the catabolism of lignin-related aromatic compounds

IPR ID	IPR description	Type of enzyme	Reference enzyme[a]	Number of genomes with IPR domain (average gene number with that IPR domain)	
				Bacteria	Fungi
IPR000391	Ring-hydroxylating dioxygenase beta subunit	Hydroxylase	B: HcaA2 (=HcaF)	43 (2.91)	0
IPR000627	Intradiol dioxygenase, C-terminal	Intradiol dioxygenase	B: CatA, ChqB, PnpC, PcaGH, TsdC, SACTE_2871 F: AN4532, AN8566, Hdq1, Hdq2, Tv28066, Cs116134, Ro01857, NRRL3-04787, NRRL3-01405, NRRL3-02644, NRRL3-05330	77 (2.74)	229 (4.05)
IPR001663	Aromatic ring-hydroxylating dioxygenase, alpha subunit	Hydroxylase	B: HcaA1 (=HcaE)	79 (3.29)	105 (2.02)
IPR002830	UbiD domain	Non-oxidative decarboxylase	B: AroY, BsdC, EcdC, KpdC, LpdC, VdcC, YclC F: Fdc1	138 (1.22)	54 (1.70)
IPR002938[b]	FAD-binding 3	Oxidative decarboxylase/ hydroxylase	B: MobA, PobA, PraI, ro01860 F: Mnx1, Mnx3, PhhY, VhyA, Tv58730[c], Cs82057[c], Cs120062[c]	199 (4.96)	304 (28.26)
IPR004183	Extradiol dioxygenase subunit B	Extradiol dioxygenase	B: LigB, PmdB, GalA DesB, PraA, DesZ, HpcD	109 (1.71)	267 (1.73)
IPR004507	Flavin prenyltransferase UbiX-like family	Non-oxidative decarboxylase	B: BsdB, EcdB, KpdB, LpdB, VdcB, YclB	139 (1.12)	52 (1.50)
IPR005708	Homogentisate-1,2-dioxygenase	Extradiol dioxygenase	F: MhdA, PcHGD1, PcHGD2	62 (1.06)	221 (1.96)
IPR007535	Catechol dioxygenase, N-terminal	Intradiol dioxygenase	B: CatA, ChqB, PnpC, TsdC F: AN4532, AN8566, Hdq1, Hdq2, Tv28066, Cs116134	33 (2.21)	219 (2.55)
IPR008729	Phenolic acid decarboxylase	Non-oxidative decarboxylase	B: PadC F: Pad	13 (1.00)	69 (1.09)
IPR011986	Extradiol ring-cleavage dioxygenase LigAB, LigA subunit	Extradiol dioxygenase	B: LigA, PmdA, GalA DesB	22 (1.45)	0
IPR012733	4-Hydroxybenzoate 3-monooxygenase	Hydroxylase	B: PraI, PobA	36 (1.03)	0
IPR012791	3-Oxoacid CoA transferase subunit B	Downstream ring-cleavage enzyme	B: CatJ, PcaJ F: Osc1, NRRL3_01886, AN10495	107 (1.50)	245 (1.68)
IPR012792	3-Oxoacid CoA transferase subunit A	Downstream ring-cleavage enzyme	B: PcaI, CatI	105 (1.56)	199 (1.46)
IPR012793	β-Ketoadipyl CoA thiolase family	Downstream ring-cleavage enzyme	B: PcaF, F[d]	53 (1.38)	0
IPR012941	Phenol hydroxylase domain	Hydroxylase	F: Mnx3, PhhY, Tv58730[c], Cs82057[c]	5 (1.20)	232 (3.96)
IPR024756	Protocatechuate 3,4-dioxygenase beta subunit, N-terminal	Intradiol dioxygenase	B: PcaH	42 (1.07)	0
IPR034786	Maleylacetate reductase family	Downstream ring-cleavage enzyme	B: HapF, MacA	19 (1.79)	111 (1.22)

This table includes IPR IDs, IPR descriptions and the source of reference enzymes (fungal (F) or bacterial (B)) organized by IPR ID number. The biochemical activities of these reference enzymes are illustrated in Fig. 1. Additional details about these enzymes, including enzyme description, substrate specificity, associated microbial species, identified domains and the source publication can be found in Supplementary Tables 6–10. This table also indicates the number of bacterial and fungal genomes (out of 255 and 317 genomes, respectively) that contain at least 1 gene with each IPR domain and, in parentheses, the average number of genes in these genomes that contain that IPR domain. [a]B, bacteria; F, fungi. [b]This IPR domain was not classified as 'proxy IPR domain' based on the IPR down selection pipeline (Methods). However, this domain belongs to three enzymes involved in the catabolism of lignin-related compounds recently validated oxidative decarboxylases and hydroxylases[21], which are also shown in Fig. 1. [c]Enzymes from WRF validated via biochemical analyses[21]. [d]Fungal enzymes with β-ketoadipyl CoA thiolase activity have been previously characterized. However, their IPR descriptions are very general and do not correspond to IPR012793. As a result, they are not included in Table 1 but are presented in Fig. 1 and Supplementary Table 10.

the aromatic catabolic enzyme suite is coupled with lineage divergence. The phylogenetic tree also confirms a tight linkage of phylogenetic relatedness with the profiles of proxy IPR domains throughout the fungal kingdom (Fig. 2a). We also examined whether the distribution of proxy IPR domains is correlated with nutritional modes in bacteria, fungi and Agaricomycotina (Supplementary Fig. 20, Supplementary Tables 17–19, Supplementary Text 1 and Supplementary Data 2). Multivariate analyses indicated a weak association between IPR domain distribution and nutritional modes. These results show that while there are notable differences in gene count among nutritional modes, the enzyme types remain independent of nutritional modes.

Correlation between lignin depolymerization and catabolism

WRF are the primary microbes responsible for lignin depolymerization in forests; however, bacteria also exhibit a limited capacity to degrade lignified plant cell walls[30]. Thus, we examined the distribution of lignin-oxidizing enzymes (lignin peroxidases (LiP), manganese peroxidases (MnP), versatile peroxidases (VP), dye-decolourizing peroxidases (DyP) and laccases) to determine whether enzymes involved in lignin depolymerization and catabolism of lignin-related aromatic compounds co-occur (Supplementary Table 11, Supplementary Figs. 21–23 and Methods). In bacteria, only laccases and DyPs were identified—the latter being the most abundant and prevalent in

https://doi.org/10.1038/s41559-025-02785-6

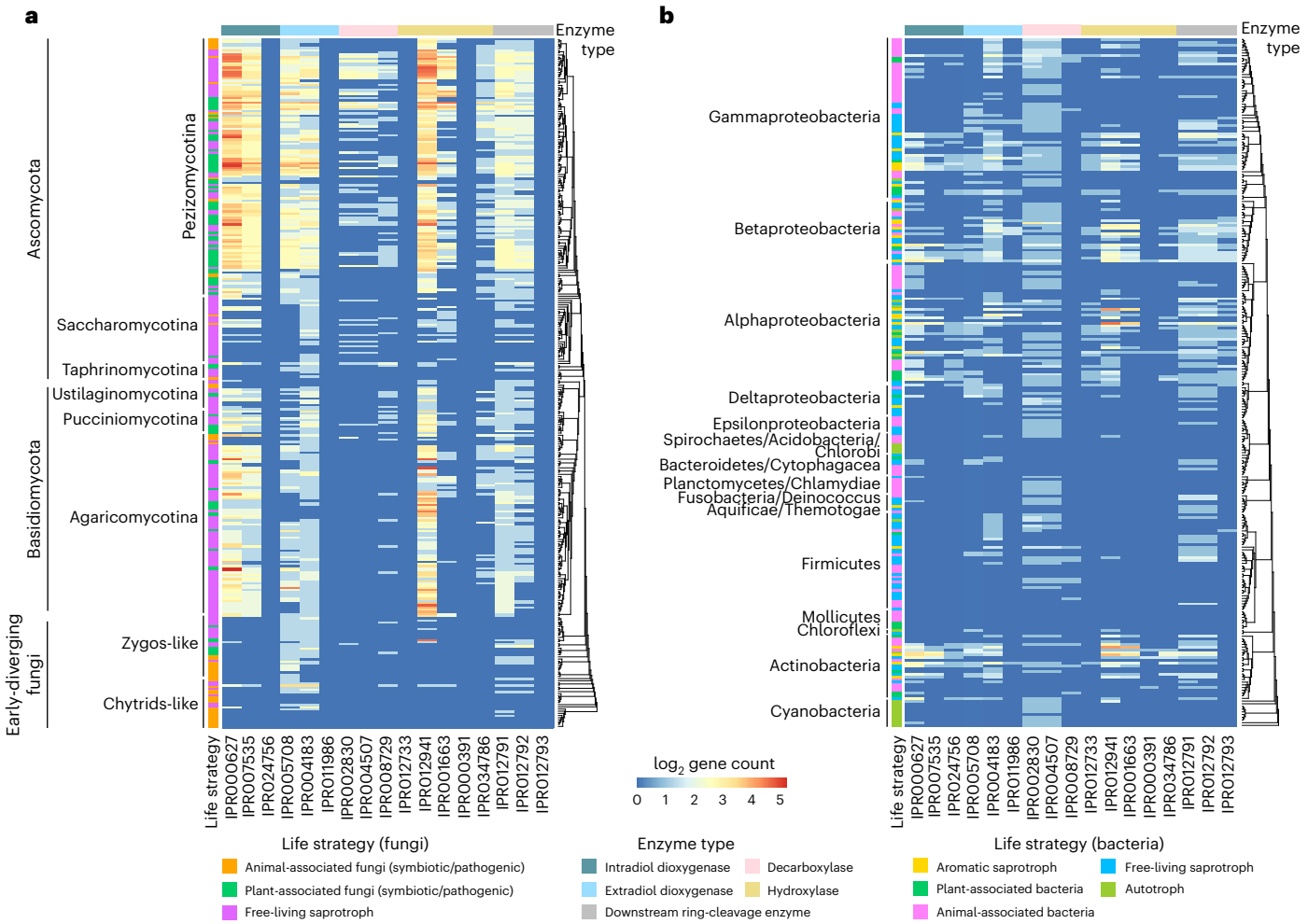

Fig. 2 | Distribution of proxy IPR domains of putative enzymes associated with the conversion of lignin-related compounds in fungi and bacteria. **a**,**b**, The distributions of proxy IPR domains for fungi (**a**) and bacteria (**b**) are shown. The colours in the heatmap cells indicate $\log_2$ abundance of annotated proteins that contain the proxy IPR domain. Rows represent each independent fungal or bacterial genome and the colour keys indicate the nutritional mode. Columns include each proxy IPR domain and colour keys indicate the type of activity. Bacterial and fungal lineages are shown on the left of the heatmaps in **a** and **b**; phylogenetic maps for bacteria or fungi are shown on the right. The list of genomes as well as their abbreviations can be found in Supplementary Tables 1 and 2. A figure with higher resolution and labels for the bacterial and fungal genomes can be found in Supplementary Data 1.

Gammaproteobacteria and Actinobacteria—although their capacity is known to be limited compared to fungal ones[31]. In fungi, domains for the selected lignin-oxidizing enzymes were primarily found in Pezizomycotina (filamentous Ascomycota)—mainly laccases and DyPs (other hemoperoxidases identified (IPR001621) are non-ligninolytic generic peroxidases as revealed by the absence of catalytic sites featured in LiPs, MnPs and VP, after manual revision of their amino acid sequences)—and Agaricomycotina (filamentous Basidiomycota)—including all types of lignin-oxidizing enzymes (Supplementary Figs. 21 and 22). Among Agaricomycetes, these domains are mostly found in WRF (Supplementary Fig. 23). Correlation analyses between the abundance of proxy IPR domains and lignin-oxidizing enzymes in bacteria and fungi (Supplementary Figs. 24 and 25) showed strong correlations (Spearman $r > 0.7$) among dioxygenases, precleavage enzymes (hydroxylases and decarboxylases) and enzymes downstream ring cleavage in fungi. These sets of enzymes were also positively correlated, albeit more weakly ($r < 0.7$), with the genomic occurrence of lignin-oxidizing enzymes. However, weak or non-existent correlations were found among enzyme types within Agaricomycetes that represent different wood-decaying strategies (Supplementary Fig. 26). Positive but moderate correlations ($r < 0.7$) were observed among all sets of enzymes in the bacterial dataset. These findings suggest that the microbial ability for lignin depolymerization and the catabolism of lignin-related aromatic compounds are either uncoupled or only weakly coupled in wood-degrading fungi.

Specificity of intradiol dioxygenases is shaped by multiple evolutionary processes

We initiated gene tree phylogenetic reconstruction, among putative enzymes involved in the catabolism of aromatic compounds, with aromatic ring-cleavage enzymes (Fig. 3a,b, Supplementary Text 2 and Supplementary Table 21). While it is known that intradiol and extradiol dioxygenases have different origins[32], the evolutionary processes within each group and between kingdoms are less understood. Intradiol dioxygenases are involved in the cleavage of aromatic catabolic intermediates (for example, catechol, protocatechuate and hydroxyquinol; Fig. 1 and Table 1) derived from the conversion of G- and H-lignin units. We identified three major clades, all containing the proxy domain IPR000627 (Fig. 3a). The first clade (Fig. 3a, purple outer arc) includes enzymes that only contain the proxy domain IPR000627. This clade featured a single reference enzyme, the intradiol dioxygenase SACTE_2871 from the bacterium *Streptomyces* sp. SirexAA-E, known to have activity on catechol[33] (Figs. 1 and 3a). Most enzymes in this clade contain a signal peptide (97% for bacterial enzymes and 65% for

https://doi.org/10.1038/s41559-025-02785-6

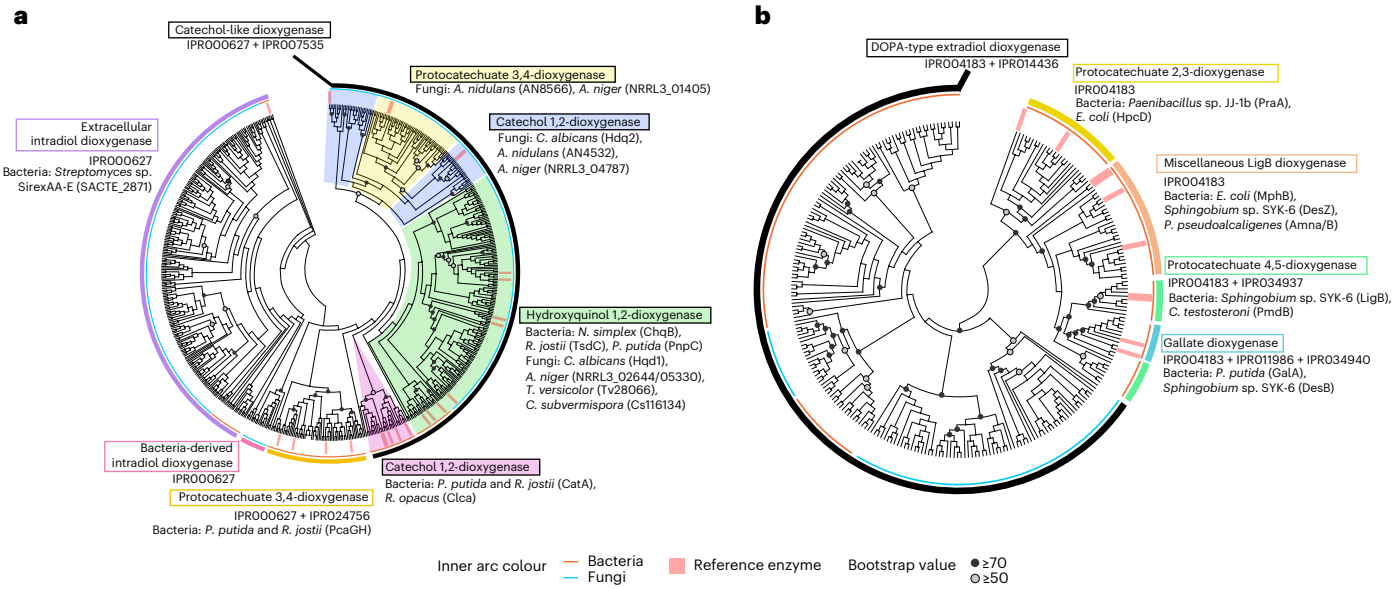

Fig. 3 | Gene tree phylogeny and evolutionary relationships of putative and characterized intradiol dioxygenases and LigB-type extradiol dioxygenases in bacteria and fungi. a, Gene COGs for bacteria (@bactNOG) and fungi (@fuNOG) that contain the IPR domain IPR000627 (intradiol dioxygenase C-terminal) and reference enzymes for intradiol dioxygenases from bacteria and fungi (Supplementary Table 6) are included in the analysis. **b**, Gene COGs for bacteria (@bactNOG) and fungi (@fuNOG) that contain the IPR domain IPR004183 (extradiol dioxygenase subunit B) and reference enzymes for LigB-type extradiol dioxygenases in bacteria (Supplementary Table 6) are included in the analysis. Circles on nodes indicate bootstrap support values. Inner arcs indicate origins of tip sequences (bacterial or fungal). Outer arcs highlight enzyme clades with predicted activities/substrate specificities based on the presence of reference enzymes in the clades. The figure with higher resolution and labels for the bacterial/fungal sequences can be found in Supplementary Data 1. Sequence names are the combination of orthogroup ID (Supplementary Table 20), paralogue number (if applicable) and species abbreviation (Supplementary Tables 1 and 3). Full species names not mentioned in the main text: *Aspergillus nidulans*, *Candida albicans*, *Nocardioides simplex*, *Pseudomonas pseudoalcaligenes*, *Rhodococcus opacus*.

fungal enzymes). Specifically, 422 proteins (372 from fungi and 50 from bacteria) contain a signal peptide while 13 of them (3 from fungi and 10 from bacteria) have one transmembrane helix with the catalytic site on the extracellular leaflet (Supplementary Data 2), suggesting an extracellular function. In addition, these extracellular intradiol dioxygenases (EIDs) were detected in the extracellular proteomes of *Pleurotus eryngii* (Joint Genome Institute (JGI) ID 20454)[34] and *Pleurotus ostreatus* (JGI ID 1102184)[35]. EIDs are rarely found in bacteria (27 out of 255 genomes; Supplementary Fig. 27) but are common in fungi (169 out of 317 genomes; Supplementary Fig. 28) including WRF (Fig. 4) and their function is unknown. There is also a small distinct clade of fungal intradiol dioxygenases that contain only IPR000627 but lack signal peptides (a magenta outer arc in Fig. 3a; Supplementary Fig. 29). In fungi, these dioxygenases have an eggNOG assignment for bacterial orthogroups (0626B@bactNOG), which suggests that these are enzymes derived from bacteria through horizontal gene transfer.

The second clade of intradiol dioxygenases (Fig. 3a, yellow outer arc) includes exclusively bacterial protocatechuate 3,4-dioxygenases. Branching patterns indicate the distinction of alpha and beta subunits, corresponding to PcaG and PcaH in *Pseudomonas putida* (Supplementary Fig. 27). The beta subunit always contains the proxy domain IPR024756, while the alpha subunit may contain the domain IPR012786 (Supplementary Table 6 and Supplementary Fig. 27). Topologies of both subclades show a pairing relationship between subunits (Fig. 3a and Supplementary Fig. 27). Lastly, the third clade (Fig. 3a, black outer arc) contains catechol-like dioxygenases and harbours the proxy domains IPR000627 and IPR007535. Bacterial enzymes are placed as basal to fungal enzymes in the tree (Fig. 3a) and are divided into two subclades on the basis of substrate preference (for catechol and chlorocatechol (pink shading) and for hydroxyquinol (green shading)). Fungal enzymes are distributed in three subclades (Fig. 3a and Supplementary Fig. 29). The first subclade includes hydroxyquinol dioxygenases (Fig. 3a, green shading), which is conserved in Ascomycota and Basidiomycota. On the basis of the distribution of reference enzymes in this subclade, hydroxyquinol seem to be a common catabolic aromatic intermediate in fungi, and mainly in wood decay fungi (Fig. 4). The second and third subclades contain 3,4-protocatechuate dioxygenases and catechol 1,2-dioxygenases (Fig. 3a, yellow and blue shadings, respectively). Interestingly, 3,4-protocatechuate dioxygenases are only found in Ascomycota, suggesting that ring cleavage of protocatechuate and downstream conversion to βKA is absent in Basidiomycota unless protocatechuate conversion proceeds via oxidative decarboxylation towards the hydroxyquinol pathway (Fig. 1). Based on these results and tree topology, we propose that fungal and bacterial 1,2-catechol dioxygenases have independent origins, and that fungal protocatechuate 3,4-dioxygenases later emerged from Ascomycota-specific catechol dioxygenases (Fig. 3a). This indicates that fungal and bacterial protocatechuate dioxygenases had different origins which is reflected by the number of enzyme subunits (two heterosubunits in bacteria and only one in fungi). Differently, hydroxyquinol dioxygenases share the same origin in bacteria and fungi. Some of these evolutionary processes were suggested in a study that included intradiol dioxygenases from the fungi *A. niger* and their enzyme homologues[36]. The current analysis demonstrates these evolutionary processes at the bacterial and fungal kingdom levels.

Fungal EIDs resemble bacterial hydroxyquinol-cleaving EID

Phylogenetic reconstruction of intradiol dioxygenases exposed a large clade of putative EIDs with undescribed functions in wood decay fungi. Therefore, we conducted experimental and computational analyses to compare the reference bacterial enzyme SACTE_2871 (Fig. 3) with one from the WRF *P. ostreatus* (OPIXU_PleosPC15_1) from that clade (Fig. 5, Supplementary Fig. 28 and Supplementary Data 2). We also included another intracellular intradiol dioxygenase (CatA1, which

https://doi.org/10.1038/s41559-025-02785-6

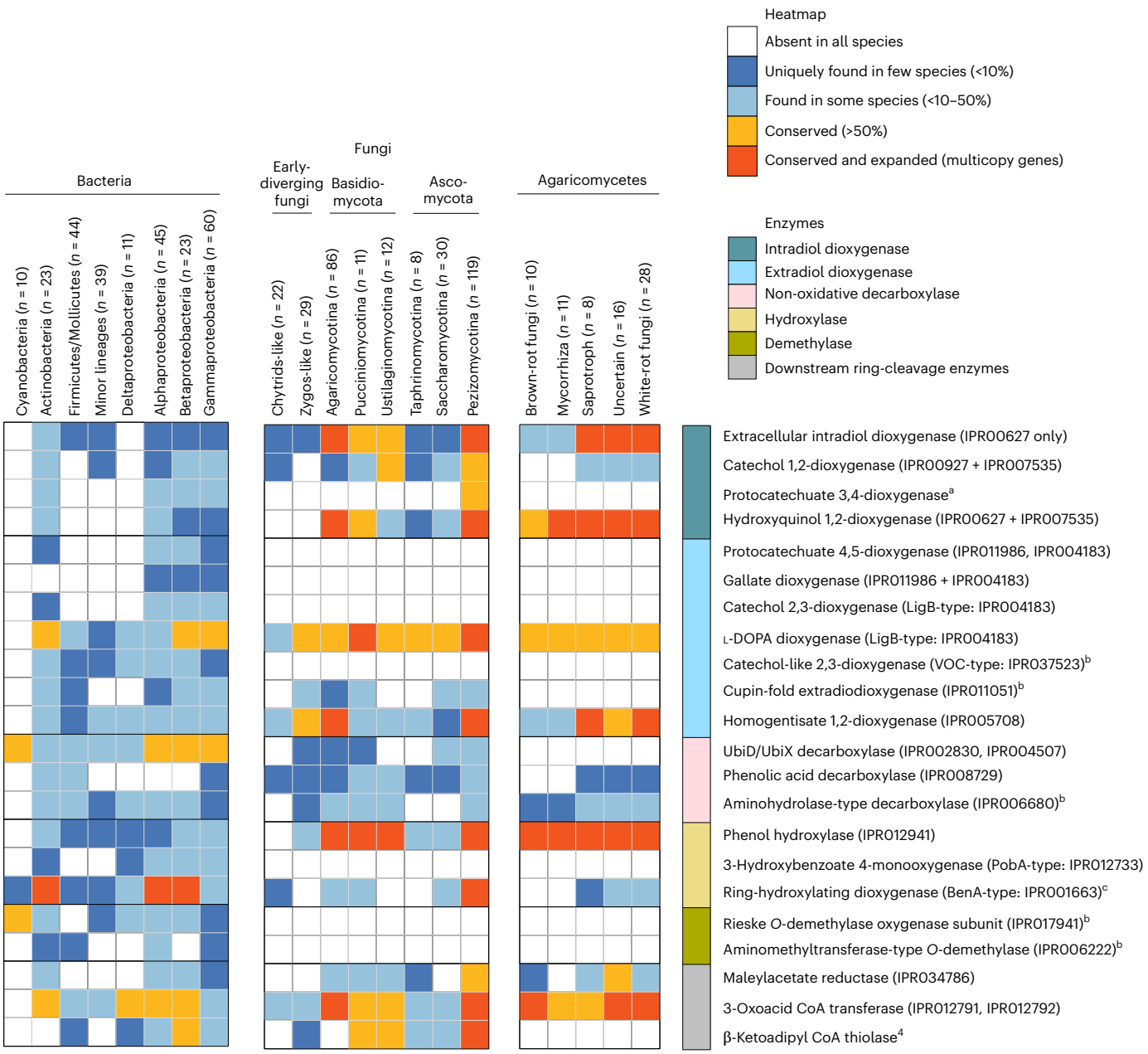

Fig. 4 | Summarized IPR domain conservation of putative enzymes associated to the catabolism of lignin-related aromatic compounds in species of bacteria, fungi and Agaricomycetes fungi. Agaricomycetes include wood decay fungi (WRF, brown-rot fungi and uncertain (that is, wood decay fungi without a consensus in the nutritional mode; Supplementary Table 5). Colours in cells indicate the level of enzyme conservation based on (1) the presence of proxy IPR domains (Fig. 2), (2) gene orthology (identical @bactNOG and @fuNOG) with the reference enzymes from literature (Supplementary Tables 6–10) and (3) phylogenetic placement from gene tree analyses (Fig. 4 and Supplementary Figs. 27–46). Value *n* indicates the number of genomes included in the analyses for each lineage/nutritional mode. Superscripts are as follows: [a]Protocatechuate 3,4-dioxygenases have different origins between bacteria and fungi. Bacterial enzymes have two different subunits (alpha subunit with IPR000627 and beta subunit with IPR000627 + IPR024756), while fungal enzymes have one subunit (IPR000627 + IPR007535). [b]Inference is only based on gene homology with reference enzymes, not the presence of the proxy IPR domains. [c]There are two groups of BenA-type enzymes that harbour IPR001663 in bacteria. The first group pairs with BenB-type enzymes that harbour IPR000391. The second group does not have a pairing with BenB-type enzymes. There is no BenB-type enzyme in fungi. [d]In bacteria, the prevalence is determined by the presence of a domain for the β-ketoadipyl CoA thiolase family (IPR012793). In fungi, the prevalence is determined by the presence of orthologues (0PHTR@fuNOG) for the β-ketoadipyl CoA thiolase Oct1, characterized in *C. parapsilosis*.

cleaves catechol, from *P. putida* KT2440) to strengthen the analyses. The purification of 0PIXU was unsuccessful. Purified SACTE_2871 was active on catechol, protocatechuate and caffeate (as previously reported)[33]. However, we also identified hydroxyquinol ring cleavage via oxygen consumption assays (Fig. 5a) and identification products via an enzymatic chain reaction (βKA from hydroxyquinol) (Fig. 5b). This reaction was positive at pH 7 for both SACTE_2871 and CatA1—relevant pH for intracellular enzymes—and at pH 5 for SACTE_2871—relevant pH for extracellular enzymes in wood decay environments.

Further structural analyses comparing the docked structures of CatA1, SACTE_2871 and 0PIXU were performed to better rationalize the substrate specificity of 0PIXU. 0PIXU structurally matches SACTE_2871 (PDB 4ILT, root mean square deviation of 1.453 Å; Fig. 5c) with an absolute conservation of the intradiol-type ferric ion core which is cradled

https://doi.org/10.1038/s41559-025-02785-6

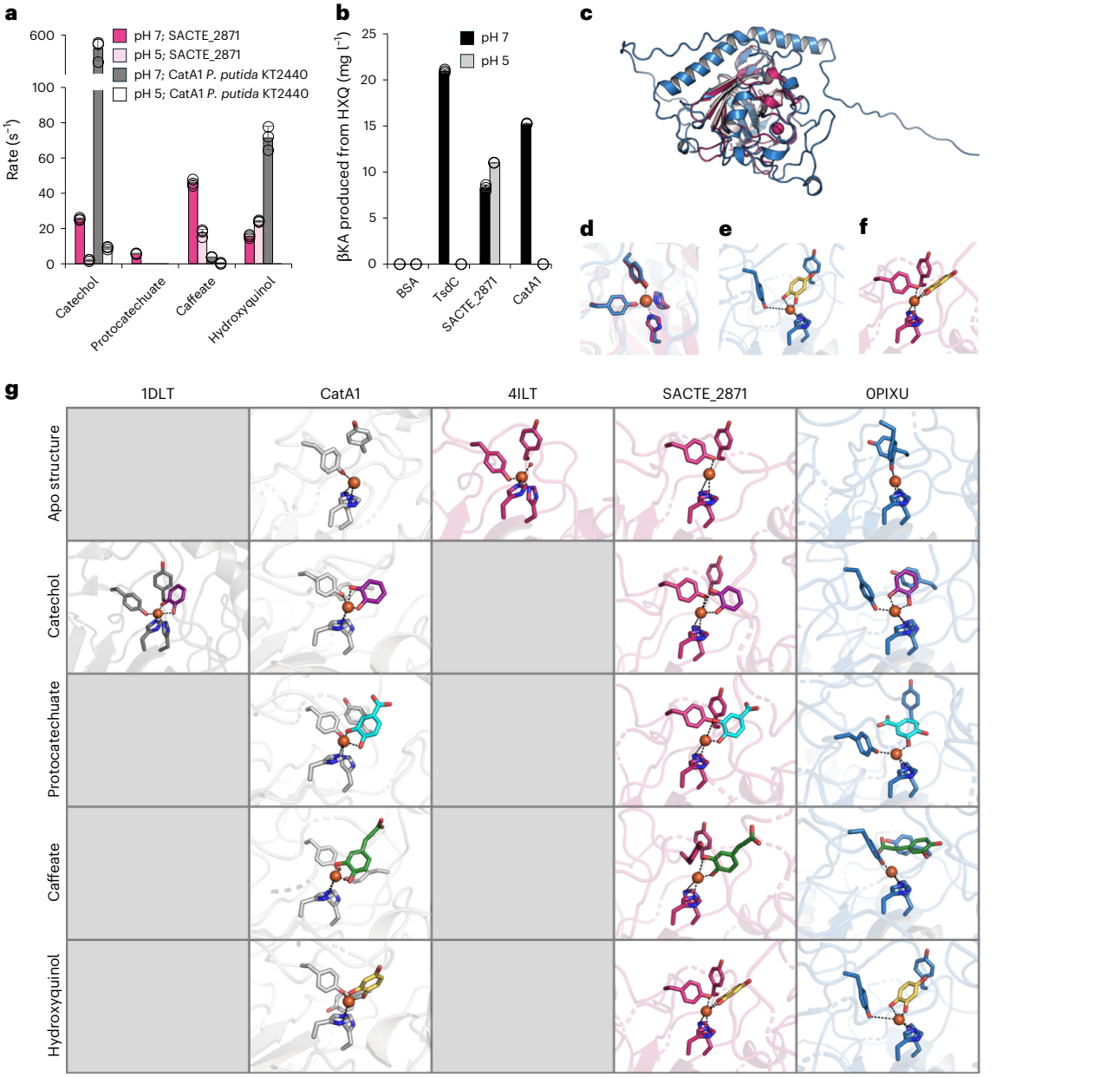

Fig. 5 | Experimental and computational analyses of intracellular and EIDs. a, Specific activity measurements of SACTE_2871 from *Streptomyces* sp. SirexAA-E (magenta) and CatA1 from *P. putida* KT2440 (blue) on catechol, protocatechuate, caffeate and hydroxyquinol following oxygen consumption in 0.1 M potassium phosphate at pH 7 (darker shade) and pH 5 (lighter shade). Bar graphs represent the average of three biological replicates and the error bars show standard deviation from these replicates. Dot plots display the individual datapoints. **b**, Conversion of hydroxyquinol (HXQ) to βKA at pH 5 and 7 via a chain enzymatic reaction with the dioxygenase SACTE_2871 or CatA1 and a subsequent reductase TsdD (TsdD reported previously[21]). TsdC is a functional dioxygenase on HXQ and used as a positive control. BSA is bovine serum albumin and used as negative control. Bar graphs represent the average of three biological replicates and the error bars show standard deviation from these replicates. Dot plots display the individual datapoints. **c**,**d**, Overall enzyme structure shown as ribbon representation (**c**) and metal-binding coordination shown as licorice representation (**d**) of an AlphaFold model of 0PIXU from *P. ostreatus* (blue) and SACTE_2871 (Protein Data Bank (PDB) 4ILV, magenta). **e**,**f**, NeuralPLexer-based docking of HXQ (yellow) with 0PIXU (blue) (**e**) and SACTE_2871 (magenta) (**f**). **g**, Comparison of best-docked substrate-bound complexes. The docked complexes were generated de novo using NeuralPLexer using the primary sequences of CatA1 (white), SACTE_2871 (magenta) and 0PIXU (blue) in combination with the simplified molecular input line entry system (SMILES) codes of catechol (purple), protocatechuate (cyan), caffeate (green) and HXQ (yellow). The crystal structures of catechol in complex with CatA1 from *Acinetobacter* sp. ADP1 (PDB 1DLT, grey) and apo-structure of SACTE_2871 (PDB 4ILT, magenta) are shown for comparisons. Source data for **a** and **b** are shown in Supplementary Data 3.

by two-His and two-Tyr metal-binding ligands (Fig. 5d). Evaluation of these models suggests smaller substrates such as catechol and hydroxyquinol can readily be bound productively in the active sites of CatA1, SACTE_2871 and 0PIXU (Fig. 5e–g). In contrast, the result of the docking exercise was not as conclusive for larger substrates such as protocatechuate or caffeate (Fig. 5g). Taken together, 0PIXU can be expected to function similarly to SACTE_2871 specifically towards smaller catecholic substrates considering that both are secreted proteins with similar overall architecture and active site.

To get further insights into the origin and evolution of EIDs in wood-degrading fungi, we conducted a computational analysis of gene family evolution (CAFE) in 52 species of Agaricomycetes. Out of 87 enzyme families examined (that is, including EIDs, extracellular enzymes for plant cell wall degradation and enzymes for amino acid

metabolism), only 10 exhibited a significantly high evolutionary rate, including EIDs (Supplementary Fig. 30). No correlations were found between Agaricomycete lifestyles and EIDs. The common ancestor of the 52 analysed species had 2 EIDs and the subsequent evolutionary paths varied among fungal groups. Generally, EIDs were lost during the evolution of Polyporales and in the ancestor of Boletales. However, Agaricales (which include 0PIXU from *P. ostreatus*) showed greater diversity, with a remarkable increased EID number throughout the evolution of the Physalacriaceae family (Supplementary Fig. 30). It is noteworthy that many Agaricales species containing EIDs lack intracellular intradiol dioxygenases (Supplementary Fig. 30). Interestingly, it appears that the lack of EIDs in Agaricomycetes is compensated by the presence of intracellular intradiol dioxygenases.

Extradiol dioxygenases are insufficiently characterized in fungi

We continued with gene tree phylogeny reconstruction for extradiol dioxygenases. Extradiol ring cleavage is well documented in bacteria but limited in fungi. To our knowledge, the only recently characterized extradiol dioxygenases in fungi associated with the catabolism of lignin-related compounds catalyse the cleavage of 2-methoxyhydroquinone[37,38] by MhdA in *A. niger* and the cleavage of 2-methoxyhydroquinone and dimethoxyhydroquinone by PcHGD1 and PcHGD2 in *Phanerochaete chrysosporium*[38]. This reaction enables ring cleavage without prior demethylation in downstream catabolic pathways of G- and S-type lignin-related compounds (Fig. 1). These fungal enzymes contain the domain IPR005708 (homogentisate-1, 2-dioxygenase), which is highly conserved in Pezizomycotina and Agaricomycotina—including in WRF—while being less conserved in bacteria (Table 1 and Figs. 2 and 4). Gene tree phylogeny of IPR005708-containing enzymes reveals that these enzymes are duplicated in several species of both Pezizomycotina and Agaricomycotina, but their duplication patterns are distinct (Supplementary Fig. 31). Multicopy homogentisate-type dioxygenases in Agaricomycotina species result from recent duplication (as seen in *P. chrysosporium*). However, multicopy enzymes in many Pezizomycotina species are found in multiple clades (as seen in *A. niger*), suggesting ancient duplication and functional divergence (Supplementary Fig. 31). In bacterial pathways, extradiol dioxygenases are critical for the catabolism of aromatic compounds derived from S-type lignin units (Fig. 1 and Table 1) and there are three major types: LigB domain-containing, vicinal oxygen chelate (VOC) and cupin-fold[39]. Apart from the well-known bacterial clades (Fig. 3b, Supplementary Text 3, and Supplementary Figs. 32 and 33), we only identified a clade of LigB-type extradiol dioxygenases in fungi (annotated as DOPA-type extradiol dioxygenases; Fig. 3b, black outer arc), which may represent the ancestral state before the divergence of LigB-type enzymes (Figs. 3b and 5 and Supplementary Fig. 33). DOPA-type dioxygenases are known to be involved in the biosynthesis of secondary metabolites[40] and are conserved in WRF and other wood decay fungi (Fig. 4). Related to VOC and cupin-fold dioxygenases (Supplementary Text 3 and Supplementary Table 20), large clades without reference enzymes were identified (Supplementary Figs. 34 and 35), representing an opportunity to explore meta-cleavage of lignin-related compounds.

Bacterial-type non-oxidative decarboxylation and demethylation is not conserved in wood decay fungi

Decarboxylation, hydroxylation and demethylation are critical reactions that modify aromatic lignin-related compounds before ring cleavage[23] (Fig. 1). Only a few non-oxidative decarboxylases can be found in fungi and their presence is not widespread compared to bacteria (Fig. 4, Supplementary Figs. 36–38, Supplementary Data 1 and Supplementary Text 4). An example of these enzymes is the ferulic acid decarboxylase Fdc1 from *Saccharomyces cerevisiae*, which is the only UbiD-containing decarboxylase characterized[41] and converts the carboxylic acid group into vinyl derivates (Fig. 1). Oxidative decarboxylases have been recently biochemically characterized in WRF[21] and these enzymes (as well as some hydroxylases) contain the FAD-binding domain 3 (IPR002938), which is not considered a proxy IPR domain (Table 1 and Supplementary Table 20). However, on the basis of domain composition and phylogenetic placements, we have identified closely related putative decarboxylases and hydroxylases that warrant investigation to continue elucidating fungal aromatic catabolic pathways (Supplementary Figs. 39–41 and Supplementary Text 5). Lastly, we did not detect orthologous enzymes of bacterial *O*-demethylases in fungi (Fig. 4, Supplementary Figs. 42–45 and Supplementary Text 6). This suggests that *O*-demethylation requires a different set of enzymes (for example, cytochrome P450s)[42] or that methoxylated aromatic compounds are catabolized through an alternative pathway (ring cleavage without demethylation), as described above[37,38,37].

Gaps in the downstream pathway from hydroxyquinol cleavage in wood decay fungi

Downstream biochemical reactions from hydroxyquinol ring cleavage towards central carbon metabolism include enzymes such as maleylacetate reductases, 3-oxoacid CoA transferases and β-ketoadipyl CoA thiolases (Fig. 1 and Table 1). In general, these enzymes are interspersed across the bacterial and fungal trees (Supplementary Figs. 46–48). However, β-ketoadipyl CoA thiolases are currently not detected in early-diverging fungi and Agaricomycotina, suggesting that enzymes with still unknown domains/orthologues may conduct this catabolic step in WRF (Supplementary Text 7).

Discussion

Here, we integrate genomic and phylogenetic analyses to gain insight into the metabolic capacity of bacteria and fungi to catabolize lignin-derived aromatic compounds, with a particular focus on wood decay fungi in the Agaricomycetes class. Evolution has favoured a limited number of enzymes (such as peroxidases and laccases) and organisms (such as WRF) for the efficient degradation of lignin. In contrast to lignin depolymerization capabilities, this study reveals that the capacity to modify and/or catabolize monomeric lignin-related aromatic compounds is widely distributed across both the fungal and bacterial kingdoms and involves a diverse enzyme array which contributes to deadwood carbon turnover and/or carbon sequestration as microbial biomass. Additionally, our results show that bacteria have a broader diversity of IPR domains associated with aromatic catabolism while fungi exhibit fewer but more highly conserved domains. Regarding lignin composition, G-type lignin is typically more resistant than S-type lignin to decay[43,44]. Considering that WRF are known to mineralize lignin to CO_2 and H_2O (ref. 45) and given that extradiol ring cleavage is probably necessary for the catabolism of S-type lignin monomers (Fig. 1), we hypothesize that WRF may have evolved various pathways for the catabolism of this particular lignin unit. For instance, fungal extradiol dioxygenases may directly cleave dimethoxylated aromatic compounds (that is, dimethoxyhydroquinone), as recently described[38]; however, the preceding and necessary oxidative decarboxylation of S-type units remains to be elucidated. In addition, this study highlights fungal enzyme families, such as DOPA-type and VOC-type extradiol dioxygenases (Fig. 4b), which warrant further investigation to understand S-type lignin catabolism.

WRF, the main clade of lignin-degrading fungi, are the most recent common ancestor in the Agaricomycetes class[6] and currently account for 94% of the 1,600–1,700 species of wood-rotting fungi in North America alone[46]. Brown-rot fungi, which are highly efficient at using plant polysaccharides, evolved from ancestral WRF and lost the capability to appreciably degrade lignin[47]. Even though the nutritional mode in the Agaricomycetes class was not correlated with the distribution of putative aromatic catabolic enzymes (Supplementary Fig. 20), brown-rot fungi showed, in general, a lower level of gene conservation for intradiol

dioxygenases, decarboxylases and downstream ring-cleavage enzymes compared to WRF (Fig. 4). In fact, brown-rot fungi lack certain domains relative to WRF corresponding to catechol dioxygenases, phenolic acid decarboxylases and ring-hydroxylating decarboxylases. These results suggest that brown-rot fungi may also have a reduced capacity to modify aromatic compounds compared to WRF. However, based on the presence of hydroxyquinol dioxygenases and downstream ring-cleavage enzymes (Fig. 4), we hypothesize that brown-rot fungi may use certain aromatic compounds as carbon sources, which has not yet been experimentally demonstrated.

Understanding the substrate specificity of intradiol and extradiol dioxygenases is crucial for identifying upstream and downstream ring-cleavage pathways towards central carbon metabolism[48] (Fig. 1). This study highlights the expanded substrate specificity of aromatic ring-cleavage enzymes during fungal evolution, suggesting an enhanced metabolic capacity for the intracellular catabolism of lignin-related compounds—and, therefore, the use of lignin as a carbon source—contributing to carbon sequestration in forests as fungal biomass. The significance of hydroxyquinol as a key catabolic intermediate during lignin decay is also underscored. This is evidenced by the widespread distribution and high gene copy number of intracellular hydroxyquinol 1,2-dioxygenases in the fungal kingdom, particularly among wood decay fungi (Fig. 4). Notably, this pathway requires one fewer enzyme to reach βKA from hydroxyquinol compared to the canonical βKA pathway[20,21] from protocatechuate and catechol. Furthermore, this study reveals a diversity of EIDs in fungi (including some from WRF) that are probably involved in hydroxyquinol cleavage, which broadens the spatial location of ring-cleavage reactions and products in wood decay environments.

Moving forward, addressing limitations of this study (Supplementary Text 8 and Supplementary Table 21), developing high-throughput enzyme production methods (for example, cell-free expression) and screening fungal enzymes from the clades highlighted in this study on the lignin-related compounds shown in this study and additional lignin decay products will be key to advancing gene-function assignment and improving fungal genome annotation. Evaluating these enzymes in the context of fungal secondary metabolites with aromatic structures is also recommended. Additionally, identifying enzymes and fungal hosts capable of funnelling a variety of lignin-related compounds into single value-added products is crucial for advancing the use of lignin-rich biorefinery streams and strengthening the bioeconomy[26,49]. Furthermore, linking fungi to the biochemical reactions studied here and the forests they inhabit will be crucial to enhance our understanding in global carbon cycling and carbon sequestration in soils.

Methods

Selection of reference enzymes

We retrieved reference enzymes related to aromatic catabolic pathways and lignin degradation from literature and the selection criteria included demonstration and validation of biochemical activities and closed homology to characterized enzymes. Reference enzymes included aromatic ring-modifying activities such as decarboxylases (oxidative and non-oxidative), hydroxylases, *O*-demethylases and dioxygenases (intradiol and extradiol) as well as downstream ring-cleavage activities belonging to the hydroxyquinol pathway (maleylacetate reductases, β-ketoadipate (βKA) succinyl-CoA transferases and β-ketoadipyl CoA thiolases) (Fig. 1). Within the dioxygenase family, we identified another subtype of dioxygenases referred to as cupin-fold dioxygenases, which are also considered extradiol dioxygenases in some literature[39] and we selected seven of these. Details for all these reference enzymes are shown in Supplementary Tables 6–10. Lignin-oxidizing enzymes were also included in the analyses, such as LiPs, MnPs, VPs, dye-colourizing peroxidases and laccases. Details for these enzymes are included in Supplementary Table 11.

Taxon sampling

We sampled 225 bacterial (Supplementary Tables 1 and 2) and 317 fungal genomes (Supplementary Tables 3 and 4). Protein models have been used as a representation for genomic data. Protein model sequences for most bacterial genomes were downloaded from the National Center for Biotechnology Information (NCBI) genome database. The database from the JGI, integrated microbial genomes and microbiomes, was used to retrieve protein models for additional bacterial genomes[50] and the JGI MycoCosm database to retrieve protein models for fungal genomes[51].

For bacterial genomes, we followed the sampling array from a previous bacterial pan-genome study[52] and we randomly selected one or two representative genomes from each bacterial genus. For lineage and nutritional assignment, we adopted the description from the genome homepage. The lineages were categorized by bacterial phyla (Supplementary Table 1). The nutritional modes were categorized into five groups (Supplementary Table 2): aromatic saprotrophs (which use aromatic compounds as a carbon source), autotroph (which use inorganic carbon sources), animal-associated bacteria (mutualist, commensal or parasite having an animal host), plant-associated bacteria (mutualist, commensal or parasite having a plant host) and environmental saprotroph (saprotrophs that are not known to use aromatic compounds).

For fungal genomes, we attempted to sample at least one genome from each fungal family wherever a reference genome was available (Supplementary Table 3). Lineage and nutritional assignments for each genome were assigned according to the description from the genome homepage. The coarse lineages were categorized into three groups: early-diverging fungi, Ascomycota and Basidiomycota. Fine-scale lineages were subclassified from the three major lineages: Chytrids-like fungi and Zygos-like fungi for early-diverging fungi, Taphrinomycotina, Saccharomycotina and Pezizomycotina for Ascomycota and Agaricomycotina, Pucciniomycotina and Ustilaginomycotina for Basidiomycota. The nutritional modes were categorized into three groups: animal-associated fungi (mutualist, commensal or parasite having an animal host), plant-associated fungi (mutualist, commensal or parasite having a plant host) and environmental saprotroph (Supplementary Table 4). Of 317 fungal genomes, 73 genomes belong to Agaricomycetes, which is a lineage known to include wood-decaying fungi. We assigned nutritional modes to the fungi included in Agaricomycetes by using a consensus from several publications[6,9,12,53–55]. The nutritional modes were categorized into five groups: WRF, brown-rot fungi, uncertain wood-degrading fungi (including 'grey' rot fungi or no consensus from literature), mycorrhizal fungi and saprotrophs (which do not include wood-degrading fungi) (Supplementary Table 5).

Genome-wide protein domain and gene orthology searches

Protein models of the analysed genomes and protein sequences of the reference enzymes were subjected to protein domain searches through InterProScan 5.44-79 (ref. 56). Only IPR identifiers were retained for subsequent analyses. Gene orthology was determined using eggNOG mapper 1.0.3 and eggNOG 4.5.1 (ref. 57). Only clusters of orthogroup (COGs) for bacteria and fungi, COG/bactNOG and fuNOG respectively, were used for further analyses.

Organismal phylogeny

For bacterial phylogeny, we used a bacterial core gene set for phylogenomics from a recent publication[58]. Prokaryotic COGs identifiers from the core gene set were used to retrieve single-copy orthologues from the 225 bacterial genomes. Ninety-two core genes were included in phylogenomic analyses. We used MAFFT 6.903 for protein sequence alignment of each gene[59]. The alignments were trimmed for informative sites with GBLOCKS 0.91 (ref. 60). The trimmed alignments of all the 92 core genes were subsequently concatenated, resulting in 10,082 informative sites for phylogenetic reconstruction. RAxML

v.8.2.9 (ref. 61) was used for maximum-likelihood tree estimation with PROTGAMMAILG as a protein substitution model and 1,000-replicate bootstrapping (Supplementary Fig. 2).

Fungal genomes often present issues related to the annotation of intron–exon junctions which causes a substantial loss of information for protein sequence alignments in numerous fungal species. We attempted to process the fungal data under the same pipeline as that used for the bacterial dataset, but less than 1,000 informative sites in a concatenated sequence were found, which is not adequate for reconstructing a tree of 317 genomes. Therefore, we used gene presence/absence patterns as input data for phylogenetic reconstruction. According to eggNOG annotation, 34,542 assigned fuNOG orthogroups were found among our sampled genomes. We transformed these data into a gene presence/absence (PA) matrix, with rows as a list of fungal genomes and columns as a list of orthogroups. Any genomes containing at least one gene copy of an orthgroup are assigned as presence (valued as 1), otherwise assigned as absence (valued as 0). Then, the PA matrix was used as binary data to reconstruct a species tree using RAxML 8.2.9 with BINCAT as a binary data model and 1,000-replicate bootstrapping (Supplementary Fig. 3).

Selection of proxy IPR domains

Unique IPR identifiers found in each type of enzymes were reported (Supplementary Tables 6–10) and used to construct heatmaps across bacterial and fungal genomes (Supplementary Figs. 4–18) to visualize the abundance of each IPR domain in each studied genome. We used a customized Perlscript to parse the data from genome-wide protein domain searches into the IPR distribution table (Supplementary Data 2). Tables include IPR identifiers, bacterial or fungal genomes and number of proteins with detected IPR identifiers in each genome. The IPR distribution for each type of enzyme is visualized as a heatmap with rearranged rows according to phylogenetic relatedness. The heatmap was generated through pheatmap v.1.0.12 (ref. 62) and phytools v.0.7.47 (ref. 63), run under R 3.6.2 and RStudio 1.2.5033 platforms.

We then down selected IPR unique identifiers for each enzyme to generate Fig. 2, which we denominate proxy domains. Specifically, (1) we maintained IPR identifiers with descriptions related to the catabolism of aromatic compounds (as proxies for aromatic catabolic enzymes), (2) we selected IPR identifiers classified as 'domain' type when two or more IPR IDs had similar descriptions, (3) we excluded IPR identifiers when an IPR domain had low prevalence (<10%) in the overall list of bacterial and fungal genomes (in less than 22 bacterial genomes and 31 fungal genomes) (Supplementary Table 12). The selected proxy IPR domains for each type of enzyme are shown in Table 1. We note that, even though some of the IPR descriptions from our validated enzymes in WRF are not specific or related to aromatic compounds (IPR002938), these domains are included in Table 1 because these results are also discussed. Because the IPR domain of β-ketoadipyl CoA thiolase (key enzyme in downstream ring cleavage) was absent in fungi, we used homology-based evidence to identify putative β-ketoadipyl CoA thiolases in fungi (the cluster of orthogroup 0PHTR@fuNOG from eggNOG database) based on a previous study that identified a β-ketoadipyl CoA thiolase from *Candida parapsilosis*[64].

For lignin-degrading enzymes, we selected proxy domains on the basis of the specific description of the enzyme of interest (Supplementary Table 11). The proxy IPR domains for each type of enzyme are as follows: IPR001621 (fungal ligninase) for LiP, MnP and VP; IPR006314 (DyP-type peroxidase family) for DyP; and IPR011707 (multicopper oxidase-like, N-terminal) for laccases. Heatmaps showing the distribution of IPR domains for lignin-degrading enzymes across bacterial and fungal genomes can be found in Supplementary Figs. 21–23.

Associations with microbial lineages and nutritional modes

First, we examined the overall distribution of the proxy IPR domains among different lineages and nutritional modes of bacteria and fungi. The abundance data of each proxy domain were combined in a single heatmap with rearranged rows according to phylogenetic relatedness of bacteria and fungi, columns representing the proxy domains and colour keys indicating nutritional modes (Fig. 2). The presence/absence patterns of the proxy domains are summarized for bacterial and fungal lineages in Supplementary Tables 13 and 15.

The association between nutritional modes and/or lineages and the distribution of the selected enzymes was also evaluated. Only proxy IPR domains were selected for this study and genomes without any proxy IPR domain were discarded from the analyses. For each proxy IPR domain, we ran a negative binomial generalized linear model with the assigned formula 'IPR count ~ nutritional mode', meaning that an enzyme count with an IPR of interest is dependent on a nutritional mode. The multcomp v.1.4-14 package[65] was used for statistical analyses, as well as the Tukey's post hoc comparison through the 'glht' function. The distribution of proxy domains was also visualized by non-metric multidimensional scaling (NMDS) plots and tested for association with lineages/nutritional modes by the Adonis test performed in vegan v.2.5.6 (ref. 66).

Finally, the correlation between proxy domains belonging to aromatic catabolic enzymes and ligninolytic enzymes was also tested. For this analysis, we categorized enzymes into four groups: ring-cleavage dioxygenases (intradiol and extradiol dioxygenases), pre-ring-cleavage enzymes (non-oxidative decarboxylases and hydroxylases), downstream ring-cleavage enzymes and ligninolytic enzymes. The abundances of enzymes having proxy IPR domains were used for scatterplots (Supplementary Figs. 24–26). Spearman's rank correlation was tested for each pair of enzyme groups.

Gene tree phylogeny and sequence analyses

We retrieved gene orthologues of the proxy IPR domains to reconstruct the gene tree phylogeny (Supplementary Table 20 and Supplementary Data 2): IPR000627, IPR007535 and IPR024756 for intradiol dioxygenase (Fig. 4a and Supplementary Figs. 27–29), IPR005708 for homogentisate-type extradiol dioxygenase (Supplementary Fig. 31), IPR004183 and IPR011986 for LigAB-type extradiol dioxygenase (Fig. 4b and Supplementary Figs. 32 and 33), IPR002830 and IPR004507 for UbiD–UbiX-type non-oxidative decarboxylase (Supplementary Fig. 36), IPR008729 for Pad-type phenolic acid non-oxidative decarboxylase (Supplementary Fig. 37), IPR012733 for 4-hydroxybenzoate 3-monooxygenase (Supplementary Fig. 40), IPR012941 for phenol hydroxylase (Supplementary Fig. 41), IPR001663 and IPR000391 for Rieske-type ring-hydroxylating dioxygenases (Supplementary Fig. 43), IPR034786 for maleylacetate reductase (Supplementary Fig. 46), IPR012791 and IPR012792 for 3-oxoacid CoA transferase (Supplementary Fig. 47), and IPR012793 for β-ketoadipyl CoA thiolase (Supplementary Fig. 48). As the domain for β-ketoadipyl CoA thiolase (IPR012793) was not detected in fungi, we used the protein orthologues (0PHTR@fuNOG) of the β-ketoadipyl CoA thiolase Oct1 in *C. parapsilosis*[64] for gene tree phylogeny. We checked the annotation of each gene orthologue identifier assigned from eggNOG mapper. Any orthologue identifier that had different annotations from what it was described for an IPR domain were discarded for subsequent analyses (Supplementary Table 21).

Gene tree phylogeny was reconstructed separately for each type of enzyme. To achieve this, all orthologous protein sequences, together with the reference enzymes, were aligned using the MUSCLE algorithm performed in MEGA-X[67]. The alignments were then visually inspected for trimming. Any protein sequence with a length of less than 50% of the average length of all the reference enzymes were removed. Extra hinges at both N terminus and C terminus regions were also trimmed off. After that, the alignments were used for the maximum-likelihood tree reconstruction using RAxML 8.2.9 with PROTGAMMAILG as a protein substitution model and 500-replicate bootstrapping. We attempted to include both bacterial and fungal sequences in one analysis. However,

some cases, such as intradiol dioxygenases and LigB-type extradiol dioxygenases, presented over 1,000 sequences at the same time, which complicates sequence alignment and phylogenetic analyses. In these cases, we ran the analyses separately for fungal sequences. Then, we sampled representative sequences from each clade from the gene tree phylogeny (when using only fungal sequences), combined them with the bacterial sequences and the reference enzymes and re-ran the analyses. Gene trees were visualized in MEGA-X and graphical representation was created through Inkscape 0.92.

We also conducted phylogenetic analyses for aromatic catabolic enzymes that do not have the proxy IPR domains but other IPR domains belonging to the reference enzymes (Supplementary Tables 6–9). Exemplars of these enzymes are VOC domain-containing extradiol dioxygenases (Supplementary Fig. 34), cupin-fold extradiol dioxygenases (Supplementary Fig. 35), aminohydrolase-type non-oxidative decarboxylases (Supplementary Fig. 38), FAD-binding domain 3-containing proteins for oxidative decarboxylases and hydroxylases (Supplementary Fig. 39), Rieske domain-containing *O*-demethylases (Supplementary Figs. 42 and 44) and aminomethyltransferase folate-binding domain-containing *O*-demethylases (Supplementary Fig. 45). In these cases, we selected only gene COGs for bacteria (@bactNOG) and fungi (@fuNOG) that have the same orthogroup IDs with each reference enzyme for phylogenetic analyses. For VOC domain-containing proteins, FAD-binding domain 3-containing proteins, Rieske domain-containing proteins and aminomethyltransferase folate-binding domain-containing proteins, we also included several sequences from each orthogroup ID for the analyses. For the orthogroups that are not similar to the reference enzymes, we sampled representative sequences by selecting bactNOG from species that are known to catabolize aromatic compounds and selecting fuNOG from species that are known as genetic models.

To predict localization of aromatic catabolic enzymes, we analysed protein sequences for signal peptides using SignalP-5.0[68]. In case there is a signal peptide in a protein of interest, we further analysed for transmembrane domains using TMHMM 2.0 (https://services.healthtech.dtu.dk/services/TMHMM-2.0/).

Gene family evolution analysis

The CAFE v.4.1 program[69] was used to analyse expansions and contractions in the family of EIDs throughout evolution in the context of the Agaricomycetes lifestyles. For that, the time-calibrated tree of 52 Agaricomycetes species and the EID gene numbers identified in these species were used as input data, together with the gene numbers of 62 additional plant cell wall degrading enzyme families, CBM1 and 24 families participating in amino acid metabolism as background gene set[10]. The gene copy numbers of the EID enzymes at ancestral nodes reconstructed with CAFE are presented in Supplementary Fig. 30. The maximum likelihood value of the birth and death parameter (k) describing the probability that any gene will be gained or lost, over a time interval, was estimated for the whole tree. Gene families with a family-wide $P < 0.05$ were considered to have a significantly higher rate of evolution. These included the EID family. Then, a Viterbi $P < 0.05$ was used to determine the branches in the species tree in which this fast-evolving family underwent notable contractions or expansions.

Modelling and docking of EIDs

The structures for CatA1 and SACTE_2871 were generated in silico and docked using AlphaFold2, respectively, in conjunction with a generative algorithm model, NeuralPLexer. The resulting docked structures were manually curated to account for both productive metal-binding ligand coordination as well as the bi-dentate binding of the catecholic substrate to the ferric ion. Specifically, NeuralPlexer was used to generate in silico binding poses[70]. The original implementation was forked to conduct a minor bug fix, and can be accessed at https://github.com/EvanKomp/NeuralPlexer and Zenodo[71]. The program was given crystal or AlphaFold2 Apo protein structures as templates, and ligands represented as SMILES strings. The following additional optional parameters were used: '--task=batched_structure_sampling --n-samples 10-- chunk-size 1 --num-steps 100 --cuda --sampler=langevin_simulated_annealing --use-template' resulting in an ensemble of 10 holo protein-ligand predicted poses, each the result of 100 diffusion denoising steps. The resulting ensembles were curated manually to avoid steric clashes and improbable poses, producing structures shown in Fig. 5.

Chemicals

All chemicals used in this work are of analytical grade or higher and used without further purification.

Cloning of EIDs

DNA manipulation and propagation were performed using standard protocols. Plasmid for heterologous expression of SACTE_2871 (pEUK159) was subcloned from pAJB016 into pET15b using Gibson Assembly master mix (NEB) and inserted between NdeI and BamHI sites using the following primer pairs GCCGCGCGGCAGCCATATGCC GCTCGTGGCAGGG and GTTAGCAGCCGGATCCCTAACCCCGCTCC CACAGTGCC. The fidelity of the construct was evaluated by Sanger sequencing (Azenta).

Protein production and purification

The codon optimized sequences of 0PIXU_PleosPC15_1, CatA1_from_pEE083 and SACTE_2871_from_pAJB025 are shown in Supplementary Table 22. SACTE_2871 and CatA1 were produced heterologously in *Escherichia coli* BL21(DE3) transformed with pEUK159 or pEE083. The cell was grown in lysogeny broth (LB Miller) supplemented with 100 mg l^{-1} of ampicillin. Freshly transformed cells were used to inoculate a starter culture (40 ml of media in 125-ml flask) and grown to saturation at 37 °C at 200 r.p.m. overnight. The starter culture was used to seed the main culture (1 l of media in 2.5-l flask) at 1% (v/v) which is grown at 37 °C at 200 r.p.m. Upon reaching optical density OD_{600} of 0.7, the culture was induced by the addition of 1 mM IPTG and 0.2 mM of ferrous ammonium sulfate. The induced culture was grown for an additional 16 h at 18 °C. The resulting biomass was harvested by centrifugation and frozen at −80 °C until further processing.

Protein purification was performed by two chromatography steps: immobilized-metal affinity and anion exchange chromatography. The thawed biomass was resuspended in Buffer A (20 mM HEPES, 100 mM NaCl, pH 7.5; 10 ml per 5 g of wet biomass) with trace amount of DNase I. The cell was lysed by French Press and clarified using the combination of centrifugation and passaging through a 0.45-µm filter. The clarified lysate was applied to high-density NiNTA beads (GoldBio) and washed with Buffer A containing up to 40 mM imidazole. The protein was eluted using Buffer A with 300 mM imidazole. The eluate was concentrated and diluted with imidazole-free Buffer A and applied to Source-15Q resin (Cytiva) operated using ÄKTA Pure fast protein chromatography system. The protein was eluted with a gradient of Buffer A containing 1 M NaCl. Fractions of interest, identified by ultraviolet traces and confirmed by SDS–polyacrylamide gel electrophoresis, were pooled and concentrated using centrifugal spin filter with 10 kDA cut-off (Millipore Sigma) frozen over a bath of liquid N_2. Frozen proteins were stored at −80 °C until further use.

Enzyme specific activity measurements

Enzymatic activity of purified SACTE_2871 and CatA1 were evaluated by monitoring the consumption of O_2 using OXYG1+ oxygraph (Hansatech). The instrument was calibrated daily following the manufacturer instruction using air-saturated water and sodium dithionite. The measurements were performed using 0.1 M potassium phosphate at pH 5 and 7 at 25 °C. Typical reaction contains 0.5 mM aromatic substrates (stock concentration of 100 mM in dimethylsulfoxide) and initiated

by the addition of the purified dioxygenase. One exception is that 0.1 µM of CatA1 was used when in conjunction with catechol due to the upper limit of the oxygraph. The observed initial velocities were collected over at least one minute and corrected by the background rates before the enzyme addition. Enzyme assays were conducted in triplicate (biological replicates) and standard deviation is reported as error bars in graphs.

Identification of products from enzyme reactions

Samples for high-performance liquid chromatography (HPLC)-based analysis were prepared as described for specific activity measurements with the addition of TsdD from *Rhodococcus jostii* RHA1 and NADPH to couple the production of ring-opened product of hydroxyquinol to the production of βKA as reported before[21]. The reaction setup follows the one described above for specific activity measurements and the reaction was incubated for 16 h at room temperature. βKA was analysed via HPLC[72] following the next steps. βKA was converted to levulinic acid by adding 10 µl of 10 N H_2SO_4 to 1 ml of sample volume. The sample was heated at 50 °C for at least 8 h. Quantitation of levulinic acid was conducted using an Agilent 1200 Series system (Agilent Technologies) equipped with a refractive index detector. Separations were performed on a HPX-87H 7.8 × 300 mm internal diameter, 9-µm column (BioRad) with an isocratic flow of 5 mM H_2SO_4 at 0.6 ml min^{-1} for a total run time of 27 min. Standards and samples were injected onto the column at a volume of 20 µl, while the temperature of the column and detector were maintained at 55 °C. Enzyme assays were conducted in triplicate (biological replicates) and standard deviation is reported as error bars in graphs.

Reporting summary

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

Data availability

All data are available in the article, Supplementary Information and Supplementary Data 1–3. The protein sequences of bacterial and fungal genomes used in this study were retrieved from public databases including DOE-JGI MycoCosm and NCBI Genome database. Reference genomes and respective publications are detailed in Supplementary Information. Raw protein sequences of studied genomes, raw protein sequences of studied enzymes, genome-wide results of InterProScan and eggNOG mapper, selected sequences for phylogenetic analyses, as well as other customized scripts (in Perl and R) for data retrieval and data analyses are available on figshare at https://doi.org/10.6084/m9.figshare.28541660.v1 (ref. 73). Source data are provided with this paper.

Code availability

Analyses in this study were conducted using customized scripts written in R. The scripts are available on figshare at https://doi.org/10.6084/m9.figshare.28541660.v1 (ref. 73). NeuralPlexer was used to generate in silico binding poses and the original implementation was forked to conduct a minor bug fix, which is available on Zenodo at https://doi.org/10.5281/zenodo.15587299 (ref. 71).

References

1. McCarl, B. A., Metting, F. B. & Rice, C. Soil carbon sequestration. *Clim. Change* **80**, 1–3 (2007).
2. Siegenthaler, U. & Sarmiento, J. L. Atmospheric carbon dioxide and the ocean. *Nature* **365**, 119–125 (1993).
3. Horwath, W. in *Carbon Cycling and Formation of Soil Organic Matter* (ed. Paul, E. A.) 303–339 (Academic Press, 2007).
4. Pan, Y. et al. A large and persistent carbon sink in the world's forests. *Science* **333**, 988–993 (2011).
5. Boerjan, W., Ralph, J. & Baucher, M. Lignin biosynthesis. *Annu. Rev. Plant Biol.* **54**, 519–546 (2003).
6. Floudas, D. et al. The Paleozoic origin of enzymatic lignin decomposition reconstructed from 31 fungal genomes. *Science* **336**, 1715–1719 (2012).
7. Martínez, A. T. et al. Biodegradation of lignocellulosics: microbial, chemical, and enzymatic aspects of the fungal attack of lignin. *Int. Microbiol.* **8**, 195–204 (2005).
8. Ayuso-Fernández, I., Rencoret, J., Gutiérrez, A., Ruiz-Dueñas, F. J. & Martínez, A. T. Peroxidase evolution in white-rot fungi follows wood lignin evolution in plants. *Proc. Natl Acad. Sci. USA* **116**, 17900–17905 (2019).
9. Riley, R. et al. Extensive sampling of basidiomycete genomes demonstrates inadequacy of the white-rot/brown-rot paradigm for wood decay fungi. *Proc. Natl Acad. Sci. USA* **111**, 9923–9928 (2014).
10. Ruiz-Dueñas, F. J. et al. Genomic analysis enlightens Agaricales lifestyle evolution and increasing peroxidase diversity. *Mol. Biol. Evol.* **38**, 1428–1446 (2020).
11. Hage, H. et al. Gene family expansions and transcriptome signatures uncover fungal adaptations to wood decay. *Environ. Microbiol.* **23**, 5716–5732 (2021).
12. Nagy, L. G. et al. Comparative genomics of early-diverging mushroom-forming fungi provides insights into the origins of lignocellulose decay capabilities. *Mol. Biol. Evol.* **33**, 959–970 (2015).
13. del Río, J. C., Rencoret, J., Gutiérrez, A., Kim, H. & Ralph, J. Unconventional lignin monomers—extension of the lignin paradigm. *Adv. Bot. Res.* **104**, 1–39 (2022).
14. Bugg, T. D., Ahmad, M., Hardiman, E. M. & Rahmanpour, R. Pathways for degradation of lignin in bacteria and fungi. *Nat. Prod. Rep.* **28**, 1883–1896 (2011).
15. Kamimura, N. et al. Bacterial catabolism of lignin-derived aromatics: new findings in a recent decade: update on bacterial lignin catabolism. *Environ. Microbiol. Rep.* **9**, 679–705 (2017).
16. Fuchs, G., Boll, M. & Heider, J. Microbial degradation of aromatic compounds—from one strategy to four. *Nat. Rev. Microbiol.* **9**, 803–816 (2011).
17. Lubbers, R. J. M. et al. A comparison between the homocyclic aromatic metabolic pathways from plant-derived compounds by bacteria and fungi. *Biotechnol. Adv.* **37**, 107396 (2019).
18. del Cerro, C. et al. Intracellular pathways for lignin catabolism in white-rot fungi. *Proc. Natl Acad. Sci. USA* **118**, e2017381118 (2021).
19. Duran, K. et al. From ^{13}C-lignin to ^{13}C-mycelium: *Agaricus bisporus* uses polymeric lignin as a carbon source. *Sci. Adv.* **10**, eadl3419 (2024).
20. Harwood, C. S. & Parales, R. E. The beta-ketoadipate pathway and the biology of self-identity. *Annu. Rev. Microb.* **50**, 553–590 (1996).
21. Kuatsjah, E. et al. Biochemical and structural characterization of enzymes in the 4-hydroxybenzoate catabolic pathway of lignin-degrading white-rot fungi. *Cell Rep.* **43**, 115002 (2024).
22. de, O. G. S. C. et al. Polyphenol oxidase activity on guaiacyl and syringyl lignin units. *Angew. Chem. Int. Ed. Engl.* **63**, e202409324 (2024).
23. Erickson, E. et al. Critical enzyme reactions in aromatic catabolism for microbial lignin conversion. *Nat. Catal.* **5**, 86–98 (2022).
24. Seibold, S. et al. The contribution of insects to global forest deadwood decomposition. *Nature* **597**, 77–81 (2021).
25. Lustenhouwer, N. et al. A trait-based understanding of wood decomposition by fungi. *Proc. Natl Acad. Sci. USA* **117**, 11551–11558 (2020).

26. Beckham, G. T., Johnson, C. W., Karp, E. M., Salvachúa, D. & Vardon, D. R. Opportunities and challenges in biological lignin valorization. *Curr. Opin. Biotechnol.* **42**, 40–53 (2016).
27. Mitchell, A. L. et al. InterPro in 2019: improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Res.* **47**, D351–D360 (2019).
28. Lubbers, R. J. M. et al. Discovery of novel *p*-hydroxybenzoate-m-hydroxylase, protocatechuate 3,4 ring-cleavage dioxygenase, and hydroxyquinol 1,2 ring-cleavage dioxygenase from the filamentous fungus *Aspergillus niger*. *ACS Sust. Chem. Eng.* **7**, 19081–19089 (2019).
29. Voigt, K. et al. Early-diverging fungal phyla: taxonomy, species concept, ecology, distribution, anthropogenic impact, and novel phylogenetic proposals. *Fungal Divers.* **109**, 59–98 (2021).
30. Sigoillot, J. C. et al. in *Lignins: Biosynthesis, Biodegradation and Bioengineering* Vol. 61 (eds Jouanin, L. & Lapierre, C.) 263–308 (Academic Press, 2012).
31. Linde, D. et al. Comparing ligninolytic capabilities of bacterial and fungal dye-decolorizing peroxidases and class-II peroxidase-catalases. *Int. J. Mol. Sci.* **22**, 2629 (2021).
32. Eltis, L. D. & Bolin, J. T. Evolutionary relationships among extradiol dioxygenases. *J. Bacteriol.* **178**, 5930–5937 (1996).
33. Bianchetti, C. M. et al. Fusion of dioxygenase and lignin-binding domains in a novel secreted enzyme from cellulolytic *Streptomyces* sp. SirexAA-E. *J. Biol. Chem.* **288**, 18574–18587 (2013).
34. Peña, A. et al. A multiomic approach to understand how *Pleurotus eryngii* transforms non-woody lignocellulosic material. *J. Fungi* **7**, 426 (2021).
35. Fernández-Fueyo, E. et al. A secretomic view of woody and nonwoody lignocellulose degradation by *Pleurotus ostreatus*. *Biotechnol. Biofuels* **9**, 49 (2016).
36. Semana, P. & Powlowski, J. Four aromatic intradiol ring cleavage dioxygenases from *Aspergillus niger*. *Appl. Environ. Microb.* **85**, e01786–01719 (2019).
37. Lubbers, R. J. M. et al. Vanillic acid and methoxyhydroquinone production from guaiacyl units and related aromatic compounds using *Aspergillus niger* cell factories. *Microb. Cell Fact.* **20**, 151 (2021).
38. Kato, H. et al. Identification and characterization of methoxy- and dimethoxyhydroquinone 1,2-dioxygenase from *Phanerochaete chrysosporium*. *Appl. Env. Microb.* **90**, e01753–01723 (2024).
39. Vaillancourt, F. H., Bolin, J. T. & Eltis, L. D. The ins and outs of ring-cleaving dioxygenases. *Crit. Rev. Biochem. Mol. Biol.* **41**, 241–267 (2006).
40. Khan, M. I. & Giridhar, P. Plant betalains: chemistry and biochemistry. *Phytochemistry* **117**, 267–295 (2015).
41. Mukai, N., Masaki, K., Fujii, T., Kawamukai, M. & Iefuji, H. PAD1 and FDC1 are essential for the decarboxylation of phenylacrylic acids in *Saccharomyces cerevisiae*. *J. Biosci. Bioeng.* **109**, 564–569 (2010).
42. Mallinson, S. J. B. et al. A promiscuous cytochrome P450 aromatic *O*-demethylase for lignin bioconversion. *Nat. Commun.* **9**, 2487 (2018).
43. van Erven, G. et al. Structural motifs of wheat straw lignin differ in susceptibility to degradation by the white-rot fungus *Ceriporiopsis subvermispora*. *ACS Sust. Chem. Eng.* **7**, 20032–20042 (2019).
44. Schwarze, F. W. M. R. Wood decay under the microscope. *Fungal Biol. Rev.* **21**, 133–170 (2007).
45. Camarero, S., Bockle, B., Martínez, M. J. & Martínez, A. T. Manganese-mediated lignin degradation by *Pleurotus pulmonarius*. *Appl. Env. Microb.* **62**, 1070–1072 (1996).
46. Blanchette, R. A. Delignification by wood-decay fungi. *Ann. Rev. Phytopathol.* **29**, 381–403 (1991).
47. Hibbett, D. S. & Donoghue, M. J. Analysis of character correlations among wood decay mechanisms, mating systems, and substrate ranges in homobasidiomycetes. *Syst. Biol.* **50**, 215–242 (2001).
48. Johnson, C. W. & Beckham, G. T. Aromatic catabolic pathway selection for optimal production of pyruvate and lactate from lignin. *Metab. Eng.* **28**, 240–247 (2015).
49. Ragauskas, A. J. et al. Lignin valorization: improving lignin processing in the biorefinery. *Science* **344**, 1246843 (2014).
50. Markowitz, V. M. et al. IMG 4 version of the integrated microbial genomes comparative analysis system. *Nucleic Acids Res.* **42**, D560–D567 (2014).
51. Grigoriev, I. V. et al. MycoCosm portal: gearing up for 1000 fungal genomes. *Nucleic Acids Res.* **42**, D699–D704 (2014).
52. Lapierre, P. & Gogarten, J. P. Estimating the size of the bacterial pan-genome. *Trends Genet.* **25**, 107–110 (2009).
53. Floudas, D. et al. Evolution of novel wood decay mechanisms in Agaricales revealed by the genome sequences of *Fistulina hepatica* and *Cylindrobasidium torrendii*. *Fungal Genet. Biol.* **76**, 78–92 (2015).
54. Kohler, A. et al. Convergent losses of decay mechanisms and rapid turnover of symbiosis genes in mycorrhizal mutualists. *Nat. Genet.* **47**, 410–415 (2015).
55. Oghenekaro, A. O. et al. Genome sequencing of *Rigidoporus microporus* provides insights on genes important for wood decay, latex tolerance and interspecific fungal interactions. *Sci. Rep.* **10**, 5250 (2020).
56. Jones, P. et al. InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236–1240 (2014).
57. Huerta-Cepas, J. et al. eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Res.* **44**, D286–D293 (2016).
58. Na, S. I. et al. UBCG: up-to-date bacterial core gene set and pipeline for phylogenomic tree reconstruction. *Korean J. Microbiol.* **56**, 280–285 (2018).
59. Katoh, K. & Toh, H. Recent developments in the MAFFT multiple sequence alignment program. *Brief. Bioinform.* **9**, 286–298 (2008).
60. Castresana, J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* **17**, 540–552 (2000).
61. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
62. Kolde, R. Package 'pheatmap'. R package version 1.0.12 (2019).
63. Revell, L. J. Phytools: an R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* **3**, 217–223 (2012).
64. Gérecová, G. et al. Metabolic gene clusters encoding the enzymes of two branches of the 3-oxoadipate pathway in the pathogenic yeast *Candida albicans*. *FEMS Yeast Res.* **15**, fov006 (2015).
65. Hothorn, T. et al. multcomp. R package version 1.4-14 (2020).
66. Oksanen, J. et al. vegan. R package version 2.5.6 (2019).
67. Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **35**, 1547–1549 (2018).
68. Almagro Armenteros, J. J. et al. SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat. Biotechnol.* **37**, 420–423 (2019).
69. Han, M. V., Thomas, G. W., Lugo-Martinez, J. & Hahn, M. W. Estimating gene gain and loss rates in the presence of error in genome assembly and annotation using CAFE 3. *Mol. Biol. Evol.* **30**, 1987–1997 (2013).
70. Qiao, Z., Nie, W., Vahdat, A., Miller, T. F. & Anandkumar, A. State-specific protein-ligand complex structure prediction with a multiscale deep generative model. *Nat. Mach. Intell.* **6**, 195–208 (2022).

71. Morehead, A., Qiao, Z., Komp, E. & Madsen, J. EvanKomp/NeuralPlexer: used in publication (v1.0). *Zenodo* https://doi.org/10.5281/zenodo.15587299 (2025).
72. Rorrer, N. A. et al. Production of β-ketoadipic acid from glucose in *Pseudomonas putida* KT2440 for use in performance-advantaged nylons. *Cell Rep. Phys. Sci.* **3**, 100840 (2022).
73. Kijpornyongpan, T. Cross-kingdom comparative genomics and phylogenetics of aromatic catabolic enzymes in bacteria and fungi. *figshare* https://doi.org/10.6084/m9.figshare.28541660.v1 (2025).
74. Díaz, E., Ferrández, A. & García, J. L. Characterization of the *hca* cluster encoding the dioxygenolytic pathway for initial catabolism of 3-phenylpropionic acid in *Escherichia coli* K-12. *J. Bacteriol.* **180**, 2915–2923 (1998).
75. Gallego-Giraldo, L., Escamilla-Trevino, L., Jackson, L. A. & Dixon, R. A. Salicylic acid mediates the reduced growth of lignin down-regulated plants. *Proc. Natl Acad. Sci. USA* **108**, 20814–20819 (2011).
76. Qualley, A. V., Widhalm, J. R., Adebesin, F., Kish, C. M. & Dudareva, N. Completion of the core β-oxidative pathway of benzoic acid biosynthesis in plants. *Proc. Natl Acad. Sci. USA* **109**, 16383–16388 (2012).

Acknowledgements

This material is based upon work supported by the US Department of Energy, Office of Science, Office of Biological and Environmental Research under the Early Career Research Program. This work was mainly authored by the National Renewable Energy Laboratory, operated by Alliance for Sustainable Energy for the US Department of Energy (DOE) under contract no. DE-AC36-08GO28308. The work by F.J.R.-D. was supported by the Lig2Plast Project PID2021-126384OB-I00 funded by MICIU/AEI/10.13039/501100011033/ and by European Regional Development Fund in the European Union. J.E.E. was supported by the 1,000 Fungal Proteins project (10.46936/staf.proj.2024.61269/60012534) from the Environmental Molecular Sciences Laboratory, a DOE Office of Science User Facility sponsored by the Biological and Environmental Research program under contract no. DE-AC05-76RL01830. We thank M. C. Aime, F. Martin, M.-N. Rosso and L. Nagy for permission to use their unpublished fungal genomes at the time the study began. The computing resource for this study was through the Purdue research computing cluster facility, with access granted by M. C. Aime. A. J. Borchert is thanked for providing pAJB016. E. Erickson is thanked for providing pEE083. We thank K. J. Ramirez and M. A. Ingraham for conducting β-ketoadipate analysis via HPLC. The views expressed in the article do not necessarily represent the views of the DOE or the US Government. The US Government retains and the publisher, by accepting the article for publication, acknowledges that the US Government retains a non-exclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this work, or allow others to do so, for US Government purposes.

Author contributions

Conceptualization, funding acquisition, project administration and supervision were undertaken by D.S. Formal analysis was done by T.K. Data curation was by T.K., E. Kuatsjah and F.J.R.-D. Methodology was developed by T.K., E. Kuatsjah, E. Komp, F.J.R.-D., J.E.E. and D.S. Investigation was carried out by T.K., E. Kuatsjah and D.S. Visualization was carried out by T.K., E. Kuatsjah and D.S. The original draft was written by T.K. and D.S. All authors were involved in reviewing and editing the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41559-025-02785-6.

Correspondence and requests for materials should be addressed to Davinia Salvachúa.

Peer review information *Nature Ecology & Evolution* thanks Ronnie Lubbers and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

nature portfolio

Corresponding author(s): Davinia Salvachua

Last updated by author(s): Jun 3, 2025

Reporting Summary

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

Statistics

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

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

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Interproscan 5.44-79 was used for genome-wide protein domain searching. eggNOG 4.5 and eggNOG-mapper 1.0 were used for genome-wide gene annotation. Data fetching and formatting from Interproscan and eggNOG-mapper results were performed using customized Perl scripts, which are available in the FigShare repository, https://figshare.com/articles/dataset/Cross-kingdom_comparative_genomics_and_phylogenetics_of_aromatic_catabolic_enzymes_in_bacteria_and_fungi/28541660
Data analysis	Analyses in this study were conducted using customized scripts written in R. The scripts are available in the FigShare repository, with DOI 10.6084/m9.figshare.28541660.v1 The CAFE ver. 4.1 program was used to analyze expansions and contractions in the family of extracellular intradiol dioxygenases throughout evolution in the context of the Agaricomycetes lifestyles. References CAFE program: https://doi.org:10.1093/bioinformatics/btl097; https://doi.org:10.1093/molbev/mst100. NeuralPlexer was used to generate in silico binding poses. The original implementation was forked to conduct a minor bug fix, and can be accessed at https://github.com/EvanKomp/NeuralPlexer and at Zenodo with DOI 10.5281/zenodo.15587298. The program was given crystal or AlphaFold2 Apo protein structures as templates, and ligands represented as SMILES strings. The following additional optional parameters were used: `--task=batched_structure_sampling --n-samples 10 --chunk-size 1 --num-steps 100 --cuda --sampler=langevin_simulated_annealing --use-template` resulting in an ensemble of 10 holo protein-ligand predicted poses, each the result of 100 diffusion denoising steps.

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

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

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

The protein sequences of bacterial and fungal genomes utilized in this study were retrieved from public databases including DOE-JGI MycoCosm and NCBI Genome database. List of reference genomes and respective publications were detailed in Supplementary Materials. Raw protein sequences of studied genomes, raw protein sequences of studied enzymes, genome-wide results of Interproscan and eggNOG mapper, selected sequences for phylogenetic analyses, as well as other customized scripts (in Perl and R) for data retrieval and data analyses will be available in the FigShare data repository, of which the DOI will be provided whenever the article is published. A temporary private link is provided here during a peer-review process: https://figshare.com/s/331f89699c7dab859d5a

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	This information has not been collected
Reporting on race, ethnicity, or other socially relevant groupings	This information has not been collected
Population characteristics	This information has not been collected
Recruitment	This information has not been collected
Ethics oversight	This information has not been collected

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

Field-specific reporting

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

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

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

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

Sample size	No sample-size calculation was performed. We attempted to sample representatives from every lineage of bacteria and fungi wherever reference genomes are available. For bacterial datasets, we performed subsampling from the previous bacterial pan-genome study (Lapierre and Gogarten, Trends in genetics 2009). Their study utilized 573 sequenced bacterial genomes. For subsampling in our study, we selected at least one genome for each bacterial genus. For genera with multiple species, we randomly selected 2 – 5 representative genomes, depending on the availability. Genomes that are not randomly but intentionally sampled are from bacterial species capable to catabolize aromatic compounds. For fungal datasets, we sampled representative genomes from the established fungal tree of life depicted in DOE-JGI MycoCosm genome database. We sampled at least two representative genomes for each fungal order wherever the reference genomes are openly accessible. Genomes from fungal species with known capability for catabolizing aromatic compounds, as well as white-rot/brown-rot fungi model, were intentionally sampled. These genomes included in this study have been published, otherwise we did requests to principal investigators to utilize unpublished genomes.
Data exclusions	We excluded any genomes that we have troubles in retrieving protein sequences and/or genomes that do not have a same number of protein sequences to the genome metadata. This results in the final numbers of 255 bacterial genomes and 317 fungal genomes included in this study.
Replication	All of the in silico analyses were performed with specified software/packages, defined parameters, and customized scripts to ensure reproducibility. We re-ran analyses at least once, including phylogenomics, phylogenetics, association analyses, and sequence prediction, to ensure that the results were consistent.
Randomization	The genomic study is empirical. Random allocation is not applicable since each genome has defined lineage and nutritional mode based on species identity. We expected that phylogenetic distance (i.e. bacterial/fungal lineage) is a major confounding factor for our analyses. Therefore, we ran additional analyses using lineages as a grouping variable, and displayed enzyme distribution results under a phylogenetic topology.

Blinding | The genomic study is empirical, thus random allocation and blinding are not applicable.

Reporting for specific materials, systems and methods

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

Materials & experimental systems

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

Methods

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

Plants

Seed stocks | *Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.*

Novel plant genotypes | *Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.*

Authentication | *Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosiacism, off-target gene editing) were examined.*