



## OPEN ACCESS

## EDITED BY

Zheng Wang,  
Yale University, United States

## REVIEWED BY

Yen-Wen Wang,  
Yale University, United States  
Jossue Mizael Ortiz-Álvarez,  
National Council of Science and Technology  
(CONACYT), Mexico

## \*CORRESPONDENCE

Huifeng Jiang

✉ jiang\_hf@tib.cas.cn

Xiaonan Liu

✉ liuxiaonan@hbut.edu.cn

Jian Cheng

✉ cheng\_j@tib.cas.cn

†These authors have contributed equally to  
this work

RECEIVED 26 September 2025

REVISED 10 December 2025

ACCEPTED 16 December 2025

PUBLISHED 13 January 2026

## CITATION

Shi Y, Zong S, Zhang H, Liu S, Yang J, Lu L,  
Liu X, Cheng J and Jiang H (2026) Evolution  
of cytochrome P450 gene superfamily in  
different cellular organisms.  
*Front. Ecol. Evol.* 13:1713618.  
doi: 10.3389/fevo.2025.1713618

## COPYRIGHT

© 2026 Shi, Zong, Zhang, Liu, Yang, Lu, Liu,  
Cheng and Jiang. This is an open-access article  
distributed under the terms of the [Creative  
Commons Attribution License \(CC BY\)](#). The  
use, distribution or reproduction in other  
forums is permitted, provided the original  
author(s) and the copyright owner(s) are  
credited and that the original publication in  
this journal is cited, in accordance with  
accepted academic practice. No use,  
distribution or reproduction is permitted  
which does not comply with these terms.

# Evolution of cytochrome P450 gene superfamily in different cellular organisms

Yuxia Shi<sup>1,2†</sup>, Shu Zong<sup>2,3†</sup>, Hejian Zhang<sup>2,4†</sup>, Shuo Liu<sup>2,4</sup>,  
Junru Yang<sup>5</sup>, Lina Lu<sup>2</sup>, Xiaonan Liu<sup>1,2\*</sup>, Jian Cheng<sup>2\*</sup>  
and Huifeng Jiang<sup>2\*</sup>

<sup>1</sup>Cooperative Innovation Center of Industrial Fermentation (Ministry of Education & Hubei Province),  
Key Laboratory of Fermentation Engineering (Ministry of Education), Hubei Key Laboratory of  
Industrial Microbiology, National “111” Center for Cellular Regulation and Molecular Pharmaceutics,  
Hubei University of Technology, Wuhan, China, <sup>2</sup>Key Laboratory of Engineering Biology for Low-  
carbon Manufacturing, Tianjin Institute of Industrial Biotechnology, Chinese Academy of Sciences,  
Tianjin, China, <sup>3</sup>College of Life Sciences, Nankai University, Tianjin, China, <sup>4</sup>College of Biotechnology,  
Tianjin University of Science and Technology, Tianjin, China, <sup>5</sup>College of Agronomy and Resources  
and Environment, Tianjin Agricultural University, Tianjin, China

Cytochrome P450 enzymes, a widespread monooxygenase superfamily, are crucial for metabolic diversity and environmental adaptation. However, systematic analysis and comparison of the expansion of the P450 gene superfamily and its potential mechanisms in different biological groups remains limited. Here, we analyzed one million P450 gene sequences from Viridiplantae, Metazoa, Fungi, and Bacteria by integrating protein language models, phylogenetic inference, and gene duplication pattern analysis to resolve their evolutionary trajectories. Our results show that P450 genes expanded primarily through vertical inheritance, but 24 potential cross-group horizontal gene transfer events were also detected. Ecological analyses suggest a potential link between P450 expansion and the complexity of terrestrial niches: P450s are abundant in Viridiplantae and Fungi, with large-scale expansions in Metazoa and Bacteria restricted to specific lineages. Comparative duplication analyses highlight distinct group-specific mechanisms: Viridiplantae are driven by synergistic whole-genome duplication and tandem duplication; Metazoa expand mainly via tandem duplication, facilitating rapid functional diversification; Fungi rely predominantly on dispersed duplication, enhancing metabolic plasticity; and Bacteria maintain their ability to survive under environmental stress by dispersed duplication. At the family level, we identified three typical expansion patterns: high coverage with low copy number (e.g., CYP51), high coverage with high copy number (e.g., CYP71), and low coverage with high copy number (e.g., CYP725). These patterns collectively support core, secondary, and specialized metabolic functions in different biological groups. This study reveals the expansion patterns and driving factors of the P450 gene superfamily across biological groups, offering new genomic insights into its evolutionary diversification.

## KEYWORDS

different biological groups, ecology, evolution, expansion mechanism, P450 gene superfamily

# 1 Introduction

Cytochrome P450 enzymes, a widely distributed superfamily of monooxygenases in nature, play crucial roles in the evolution of metabolic networks (Guengerich, 2002). By catalyzing oxidation reactions such as hydroxylation and epoxidation, these enzymes participate in diverse physiological processes, including hormonal homeostasis, xenobiotic detoxification, and natural product biosynthesis (Liu et al., 2024; Huang et al., 2025). This functional versatility also determines that P450 genes in different biological groups have evolved distinct numbers and functions according to their respective adaptive requirements (Parvez et al., 2016; Esteves et al., 2021). For example, in plants, the P450 gene superfamily has expanded in scale through tandem duplication (TD) and whole-genome duplication (WGD), driving the diversification of secondary metabolites such as terpenoids and alkaloids. These metabolites enable plants to cope with herbivory, pathogen infection, and various environmental stresses. At the level of environmental adaptation, several studies have elucidated the molecular mechanisms linking specific CYP genes to stress responses (Kawai et al., 2014; Wang et al., 2018; Liu et al., 2020b; Hansen et al., 2021; Tang et al., 2024). In Viridiplantae, abscisic acid (ABA) is an important signaling molecule in response to cold stress. Cold stress can induce the expression of PCaP2, which enhances plant membrane stability and osmotic regulation by activating SnRK2.2, SnRK2.3, and their downstream targets ABF2, RD29A, KIN1, and KIN2, as well as by regulating the expression of CBF and COR genes, thereby improving cold tolerance (Wang et al., 2018; Tang et al., 2024). Meanwhile, overexpression of CYP85A1 can promote the expression of the ABA biosynthetic gene NCED1 and increase ABA levels, thereby significantly enhancing cold tolerance in tomato (An et al., 2023). In *Arabidopsis thaliana*, members of the CYP707A family are activated under cold stress and participate in the dynamic regulation of ABA catabolism to maintain endogenous hormonal homeostasis (Baron et al., 2012; An et al., 2023). In Metazoa, the P450 gene superfamily exhibits lineage-specific expansions of families such as CYP2 and CYP3, optimizing xenobiotic metabolism and homeostatic regulation to cope with the challenges posed by complex environmental chemicals (Nelson, 2018). Numerous studies have shown that such expansions are widespread among mammals, being particularly pronounced in species that primarily feed on plants. For example, genomic analysis of the koala revealed a large-scale expansion of the CYP2C family, likely associated with its adaptation to a eucalypt-based diet rich in terpenoids and phenolic compounds (Johnson et al., 2018). In woodrats, the copy numbers of CYP2A, CYP2B, and CYP3A are significantly higher than in other rodents, suggesting that these expansions enhance their ability to metabolize terpenoids from *Juniperus* plants. Similarly, in omnivorous animals such as dogs, bears, and badgers, lineage-specific duplications of CYP2 and CYP3 families have also been observed, which are believed to be related to their broad dietary

habits and the need to detoxify a wide range of plant secondary metabolites (Parvez et al., 2016). For example, Viridiplantae P450 families have undergone large-scale expansion through tandem duplication (TD) and whole-genome duplication (WGD), driving the diversification of secondary metabolites such as terpenoids and alkaloids (Kawai et al., 2014; Costello et al., 2016; Liu et al., 2020b; Kondo et al., 2022). The expansion of P450 genes in fungi not only contributes to the synthesis of secondary metabolites but also significantly enhances their detoxification capacity against natural toxins and environmental pollutants. Such expansions are considered part of the adaptation to new ecological niches. In wood-decaying fungi, both the white-rot fungus *Phanerochaete chrysosporium* and the brown-rot fungus *Postia placenta* have been extensively studied for the biodegradation of various xenobiotics. In *P. chrysosporium*, 33 CYP families are involved in the hydroxylation of polycyclic aromatic hydrocarbons (PAHs), with members of the CYP63 family predicted to participate in the degradation of multiple xenobiotic compounds (Crešnar and Petrič, 2011; Syed and Yadav, 2012; Ichinose, 2013). In natural environments, bacteria are exposed to ecological pressures such as antibiotic-rich conditions or highly competitive microbial communities. To survive and occupy ecological niches under these conditions, bacteria can acquire exogenous resistance genes through horizontal gene transfer (HGT), thereby enhancing their tolerance to antibiotics and overall survival. Thus, the acquisition of foreign genes via HGT is an important mechanism for the development of antibiotic resistance in bacterial lineages. For example, a genomic analysis of 594 soil isolates from 19 *Listeria* species revealed that the abundance of antibiotic resistance genes (ARGs) is closely associated with soil properties and land-use patterns, with aluminum and magnesium content identified as the most influential factors (Goh et al., 2024; Samynathan, 2024). These differences in P450 gene superfamily expansions among different biological groups suggest that the evolutionary trajectories of P450 genes may have been jointly shaped by environmental selective pressures and the demands of metabolic innovation (Crešnar and Petrič, 2011). In Bacteria, P450s are often acquired via HGT, representing an important mechanism for gaining antibiotic resistance (Mokhosoev et al., 2024; Samynathan, 2024). However, current studies mostly focus on single biological groups, and systematic cross-group comparisons remain limited—particularly regarding whether the evolution of P450 gene families is interconnected across different groups.

The expansion of gene families, through functional diversification, is one of the key drivers of adaptive evolution; its expansion mechanisms mainly include horizontal gene transfer (HGT) and vertical evolution (Lynch and Conery, 2000; Moreira, 2011; Boto, 2015). Existing studies have demonstrated that HGT plays a critical role in microbial resistance genes—for example, the plasmid-mediated dissemination of  $\beta$ -lactamase genes—as well as in the environmental adaptation of eukaryotes (Li and Zhang, 2023; Keeling, 2024). However, regarding horizontal gene transfer (HGT) events of P450 genes between different biological groups, only two cases have been reported to date (Shoun et al., 2012; Lamb et al., 2019). In contrast, vertical evolution—through mechanisms such as WGD, TD, and dispersed duplication (DSD)—has been the

**Abbreviations:** HGT, Horizontal gene transfer; WGD, Whole-genome duplication; TD, Tandem duplication; DSD, Dispersed duplication; SEPs, Species Enriched with P450 genes.

predominant driving force for the rapid expansion and functional diversification of gene families (Qiao et al., 2019; Cheng et al., 2021a). For instance, TD has been identified as the major mode of expansion in the plant NBS-LRR disease-resistance gene family (Kang et al., 2012); in soybean and Brassicaceae species, the terpene synthase gene family has expanded mainly through WGD and TD, respectively (Jiang et al., 2019; Li et al., 2024); in the P450 gene superfamily of pepper, 106 tandemly duplicated gene pairs and 23 segmentally duplicated gene pairs have been identified (Hao et al., 2022); in gymnosperms, ancient WGD events and TD-derived gene clusters have enriched ginkgo's chemical and antimicrobial defense pathways (Guan et al., 2016); in addition, TD in the yew genome has driven the evolution of gene families, leading to the emergence of key genes involved in taxol biosynthesis (Cheng et al., 2021b). Nevertheless, studies on the mechanisms underlying P450 gene superfamily expansions in other biological groups remain lacking.

Therefore, although research on the expansion of the P450 gene superfamily has been extensively conducted in many single taxa, a fundamental question remains unresolved: in the macroscopic picture of life evolution, does the evolution of P450 in different biological groups (using representative Viridiplantae, Metazoa, Fungi and Bacteria as models) follow a unified pattern, or have each developed unique strategies? Existing studies, limited to single biological groups, cannot reveal the universal driving forces and group-specific mechanisms shaping the diversity of the P450 superfamily. To fill this systematic gap in our understanding, this study conducted a groundbreaking cross-biological group comparative analysis. First, methodologically, we developed a new workflow based on protein language models (ESM), providing a novel tool for identifying horizontal gene transfer (HGT) events and enabling comparisons of gene families across vast evolutionary distances. Second, regarding scientific questions, we systematically compared the relative contributions of vertical evolution (such as TD and WGD) and HGT to the expansion of P450 in four major taxa within a unified framework for the first time, revealing the core principle that “ecological complexity drives gene abundance, while taxonomic genetic background shapes the expansion mechanism.” Third, in terms of theoretical insights, we proposed three universal models of P450 family expansion, providing a new paradigm for predicting and interpreting the evolutionary trajectories of other gene families. The cross-taxonomic evolutionary blueprint of P450 constructed in this study provides a key case study for understanding the macro-evolutionary laws of gene families, and also provides a theoretical basis and priority screening strategy for targeted discovery of P450 enzymes with specific catalytic functions.

## 2 Methods

### 2.1 Collection and filtration of P450s

First, we downloaded the protein sequences of all genome-sequenced and gene-annotated species available in NCBI. Subsequently, potential P450 sequences were identified using Hmsearch (Eddy, 2011) in the HMMER3 software package,

based on the Pfam ID=PF00067. To retain only complete P450 protein sequences, we excluded those shorter than 250 or longer than 750 amino acids, as among the currently characterized P450 enzymes, the shortest sequence is 343 amino acids, the longest is 718 (excluding multi-domain fusion genes), and the average length is 498. We also removed sequences lacking any of the key conserved motifs (EXXR, FXPXRF, CXG), as these motifs are essential for maintaining the structural integrity and catalytic function of cytochrome P450s: the EXXR and FXPXRF motifs form the E–R–R triad that stabilizes the core fold and locks the heme pocket into position, while the CXG motif (FXXGXXXCXG) contains the axial cysteine ligand required for heme binding and monooxygenase activity (Chen et al., 2014; Wei and Chen, 2018). Multiple sequence alignments were then performed for all P450 sequences together with their 50 closest homologs. Sequences exhibiting continuous gaps or insertions exceeding 30 amino acids at the protein level were discarded. Finally, the proportion of P450 genes (PPG) for each species was defined as the number of complete P450 protein sequences divided by the total number of annotated protein sequences in that species.

### 2.2 Ancestral P450 gene proportion inference

First, we input the names of all species within a given taxonomic group (Viridiplantae, Metazoa, or Fungi) whose genomes have been annotated into TimeTree (Kumar et al., 2022) to obtain the phylogenetic tree of that group. Subsequently, the Proportion of P450 Genes (PPG) for each species was mapped onto the leaf nodes of the phylogenetic tree. Using a probabilistic model of gene family expansion and contraction, we then inferred the P450 gene proportions at the internal nodes of the phylogeny, thereby estimating the ancestral gene ratios. This inference strategy is conceptually similar to the gene family expansion–contraction framework implemented in the CAFE software (Mendes et al., 2021), but our approach primarily focuses on estimating the proportion of genes rather than absolute copy numbers. The lambda value (–lambda) was set to 1.0, and the significance threshold (–pvalue) was set to 0.05 for the analysis. By maximizing the likelihood of the observed data and incorporating a Markov Chain Monte Carlo (MCMC) procedure, we obtained the posterior distribution of P450 gene proportions for each node in the phylogenetic tree, thereby revealing the evolutionary dynamics of this gene family across species.

### 2.3 Enrichment analysis of ecological categories

To assess the distribution patterns of species enriched with cytochrome P450 genes (SEPs) across different ecological groups, we employed a hypergeometric test to evaluate the enrichment significance of each group. The test was based on four parameters: N (the total number of species), K (the number of species within a specific ecological group), n (the number of SEPs), and k (the number of species that both belong to the given ecological group

and are classified as SEPs). In the hypergeometric test analysis, a total of 13,303 bacterial species (including 6,855 rhizosphere bacteria), 1,565 fungal species (including 615 saprotrophic fungi), and 1,586 animal species (including 389 arthropods) were included.

The significance of enrichment was evaluated using the hypergeometric test, calculated as:

$$P(X \geq k) = 1 - \sum_{i=0}^{k-1} \frac{\binom{K}{i} \binom{N-K}{n-i}}{\binom{N}{n}}$$

## 2.4 Analysis of P450 gene duplication patterns

To elucidate the gene expansion mechanisms in SEPs, we developed an automated workflow. First, high-quality genome assemblies were retrieved from the NCBI RefSeq and GenBank databases. For fungal and bacterial genomes with relatively low assembly quality, datasets were further filtered based on the N50 value to meet the analytical requirements. Specifically, for fungal genomes, contig number  $\leq 5,000$  and N50 accounting for  $\geq 0.3\%$  of the total genome length were used as minimum quality thresholds to ensure that subsequent gene duplication analyses were based on genomes with sufficient contiguity for macroevolutionary inference. For bacterial genomes, a quality scoring system was established based on N50, maximum scaffold length, scaffold number, and whether the assembly reached chromosome-level: N50  $\geq 200\text{kb}$ , maximum scaffold  $\geq 1\text{Mb}$ , scaffold number  $\leq 10$ , and chromosome-level assembly were assigned positive scores, while N50  $< 100\text{kb}$  or scaffold number  $> 100$  were assigned negative scores; genomes with a total score of  $\geq 3$  were used for subsequent analysis. After this filtering, we collected genome annotation files and protein sequence files for 186 Viridiplantae, 156 Fungi, 246 Metazoa, and 276 Bacteria species. Protein homology searches were then performed using DIAMOND (Buchfink et al., 2021) to obtain sequence similarity alignments efficiently. Finally, the Duplicate\_gene\_classifier module of MCScanX (Wang et al., 2012) was employed to classify gene duplication types across Viridiplantae, Metazoa, Fungi and Bacteria, including tandem, dispersed, proximal, WGD or segmental, and singleton genes. It is important to note that MCScanX assigns these categories based on a hierarchical algorithm (WGD > Tandem > Proximal > Dispersed) analyzing the current genomic context. Therefore, a gene classified as ‘Tandem’ indicates its current physical state (clustered with a homolog), which may supersede its deeper evolutionary history (e.g., a dispersed gene that subsequently underwent tandem duplication).

## 2.5 P450 family classification

The method for P450 family classification is based on previous study (Liu et al., 2023). First, we performed clustering using OrthoMCL (Li et al., 2003) with a 50% sequence similarity

threshold to remove redundancy and obtain a representative sequence set (P450-50). For each cluster, rather than selecting a representative sequence at random, we chose the sequence closest to the consensus sequence of the cluster to ensure that the representative sequence best reflects the core features of the cluster. Subsequently, sequences were reclustered at a 40% similarity threshold (Class-40), which corresponds to the family-level classification criterion proposed by the International P450 Nomenclature Committee (Nelson et al., 1996). A maximum-likelihood phylogenetic tree was then constructed using FastTree (Price et al., 2010) and rooted in Seaview (Gouy et al., 2010). Notably, no specific outgroup was set; instead, Seaview automatically determined the root based on the tree topology and its algorithm. We traversed the tree from the root to identify family boundaries: when two descendant clades corresponded to different Class-40 clusters, they were defined as distinct families; when members of the same Class-40 cluster accounted for more than 80% of one clade, they were merged into a single family. This threshold was applied to prevent over-splitting due to local phylogenetic inconsistencies, following a principle analogous to the majority-rule consensus approach commonly used in phylogenetic inference (Heled and Drummond, 2010). For interwoven or paraphyletic clusters, if one cluster represented more than 80% of a clade, it was merged; otherwise, the algorithm recursively assessed lower branching levels to resolve family boundaries (Supplementary Figure S15).

## 2.6 P450 group multi-classification model

To construct the P450 protein classification model, each P450–90 sequence—a representative set obtained by clustering sequences at 90% similarity to remove redundancy—was encoded as a 1,280-dimensional vector and labeled according to its taxonomic group: Bacteria, Fungi, Viridiplantae, Metazoa, Archaea, other Eukaryotes, or Viruses. Sequence embeddings were generated using the ESM-1b (Rives et al., 2021) language model. To balance model complexity and training stability, we employed a multilayer perceptron (MLP) with a 1,280–256–256 architecture, which has been widely validated for effectively capturing nonlinear features in protein classification tasks (Elnaggar et al., 2021). The model was trained using cross-entropy loss and optimized with the Adam optimizer (Kingma and Ba, 2017) at a learning rate of  $1 \times 10^{-4}$  for 100 epochs until convergence. To reduce phylogenetic bias, the training and test sets were split in an 8:2 ratio with stratification by taxonomic group to prevent homologous sequence leakage. Stratification here means that, instead of randomly sampling from the entire dataset, we performed separate 8:2 splits within each taxonomic group (e.g., Bacteria, Fungi, Viridiplantae, Metazoa, Archaea, Other Eukaryotes, and Viruses). This ensures that both the training and test sets contain a balanced representation of species from each group. Class weights were introduced during training to mitigate the effects of sample size imbalance across groups. Model performance was evaluated using precision, recall, and F1 score (Supplementary Figure S1, Supplementary Table S1). The model was implemented in PyTorch (Paszke et al., 2019).



## 2.7 Prediction P450 HGT events among different biological groups

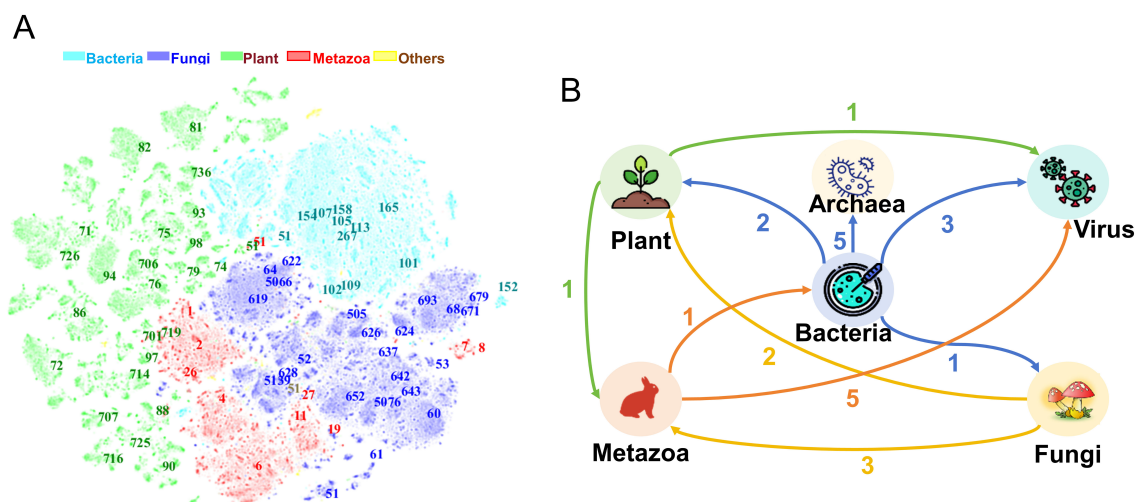
To better represent P450 enzyme sequences, we fine-tuned the protein language model ESM-1b using a dataset of P450 sequences with 90% similarity (only one representative sequence per group), and used the fine-tuned ESM-1b to encode each P450 sequence into a 1280-dimensional numerical representation. We then used these numerical representations to build a multi-class classification model that accurately predicts the phylogenetic group to which a P450 belongs (Supplementary Table S2, The model weights have been uploaded to Zenodo.). Furthermore, we used this classification model to predict the phylogenetic groups of the P450–90 dataset and obtained predicted labels. P450s whose predicted labels did not match the annotated labels were considered potential HGT candidates. For all HGT candidates, we defined that any potential HGT gene must meet three conditions to be considered a true HGT gene: 1) the HGT gene must be located within the evolutionary branch of the donor species when constructing the phylogenetic tree; 2) the HGT gene must be located within the genome, not in a short overlapping sequence; 3) the HGT gene should be present in at least three different species.

## 3 Results

### 3.1 Evolution of the P450 gene superfamily in different biological groups

To investigate the evolutionary patterns of P450 gene families across different biological groups, we retrieved approximately one million P450 gene sequences from public databases (including 552,963 from Viridiplantae, 144,614 from Metazoa, 176,890 from Fungi, 124,010 from Bacteria, and 1,035 from others; Supplementary Figure S2). Since conventional sequence analysis

methods are inadequate for such large-scale comparative analyses, we employed the ESM-1b model to encode each P450 sequence into a 1,280-dimensional numerical representation, followed by dimensionality reduction and visualization using t-distributed stochastic neighbor embedding (t-SNE) (Maaten, 2008) (Figure 1A). In addition, a small fraction of sequences showing classification inconsistencies may have arisen through cross-group horizontal gene transfer (HGT). To further detect such potential HGT events, we developed an ESM-based P450 classification model capable of predicting the biological group of origin for each P450 gene. P450 genes whose predicted biological group conflicted with their annotated origin were considered likely to have arisen through horizontal gene transfer. By integrating model predictions with manual curation, we identified 24 novel HGT events (Figure 1B, Supplementary Table S3) (Of these, 20 events had a Bootstrap score greater than 70, while the remaining four were between 50 and 70.). It has been suggested that the diversity of P450s in Archaea is much lower than in bacteria and eukaryotes, and that Archaea may have originally lacked P450 genes, acquiring them later via lateral transfer from bacteria. For example, CYP119A1 from archaea, well known for its remarkable thermostability, is likely to have originated from a bacterial P450 gene (Supplementary Figure S3). Viral P450s are also likely derived mainly from horizontal gene transfer from animals and bacteria, such as CYP102L1 (Supplementary Table S3). Moreover, studies have shown that giant viruses (e.g., *Mimiviridae* and *Pandoraviridae*) indeed harbor multiple unique P450 genes, even though the corresponding sequences in their hosts (e.g., *Acanthamoeba castellanii*) share less than 25% similarity. This suggests that the origins of these viral P450s may involve early unknown hosts or ancient horizontal gene transfer events. These studies not only reveal the diversity of viral P450s but also broaden our understanding of the evolution of the P450 superfamily (Lamb et al., 2009; Ngcobo et al., 2023). Similarly, the metazoan



**FIGURE 1**  
P450 Genes in horizontal transfer across biological groups. (A) P450 sequences were encoded into a 1280-dimensional numerical representation using the ESM-1b model, followed by dimensionality reduction and visualization via t-SNE. (B) Schematic of horizontal transfer of P450 genes across biological groups. Solid lines and arrows indicate the direction of horizontal transfer, with numbers representing the number of transfer events.

*Rhinocypha anisoptera* appears to have lost its native sterol 14 $\alpha$ -demethylase (CYP51 family) and instead acquired it via horizontal transfer from Fungi (Supplementary Figure S4). These findings suggest that horizontal transfer of P450 genes between different biological groups not only contributes to their evolutionary history but also provides organisms across groups with novel metabolic functions and adaptive potential.

### 3.2 Ecology-driven expansion hypothesis of the P450 gene superfamily

We further compared the distribution and evolutionary characteristics of P450 gene families within different biological groups. First, by examining the species-level distribution of P450 genes (443 Viridiplantae, 1,689 Metazoa, 2,202 Fungi, 70,663 Bacteria, and 2,022 Archaea species), we observed striking differences in the proportion of P450 genes among different biological groups (Figure 2A, Supplementary Figure S5). Notably, Viridiplantae (mean 0.76%) and Fungi (0.69%) exhibited significantly higher genomic proportions of P450 genes compared to Metazoa (0.36%) and Bacteria (0.08%) (Figure 2A). Interestingly, the distribution curves of Viridiplantae and Fungi displayed a bimodal pattern, suggesting finer-scale evolutionary divergences among different phyla. The two peaks in Viridiplantae are primarily driven by green algae (*Chlorophyta*) and angiosperms (*Magnoliopsida*), whereas in Fungi, the bimodal pattern is mainly contributed by the *Ascomycota* and *Zoopagomycota* lineages (Supplementary Figure S6, Supplementary Table S10). Although Metazoa and Bacteria generally showed lower genomic proportions of P450 genes, certain lineages exhibited remarkable expansions (Supplementary Figure S7). For instance, in Metazoa, *Amphiura filiformis* harbors as many as 395 P450 genes, while the bacterium *Kutzneria* sp. CA-103260 carries 100 P450 genes (Supplementary Table S4). We defined species in which P450 genes account for more than 0.6% of the genome as SEPs. Statistical

analyses revealed that SEPs accounted for 75.6% of Viridiplantae, 56.5% of Fungi, 15.2% of Metazoa, and only 1.1% of Bacteria (Supplementary Table S7, Supplementary Table S8). This indicates that P450 genes in different biological groups have their own distinct expansion patterns.

To understand the formation of P450 gene differences among biological groups, we inferred their evolutionary trends using phylogenetic trees (Figure 2B). On a large-scale temporal framework, several waves of P450 gene superfamily expansion show temporal concordance with the emergence of terrestrial ecosystems: following the appearance of land plants approximately 470 million years ago and their gradual occupation of terrestrial niches, significant increases in P450 gene numbers occurred in the lineages of plants, fungi, and arthropods (Supplementary Figures S8, S9; Supplementary Table S9) (Mizutani and Ohta, 2010; Nelson and Werck-Reichhart, 2011; Kapoor et al., 2023; Mills et al., 2023). The proportion of the P450 gene in aerobic bacteria and archaea was 26.6 times and 69.9 times higher than that in anaerobic bacteria and archaea, respectively ( $p < 10^{-30}$ , Mann-Whitney U test; Supplementary Figure S10). This temporal concordance is biochemically plausible: P450s are a class of oxygen-dependent monooxygenases whose oxidative metabolism is more active in oxygen-rich environments. Thus, the increase in atmospheric and near-surface oxygen levels, along with the resulting ecological niche complexity, provided favorable conditions for the expansion of oxidative metabolic functions (Mizutani and Ohta, 2010; Mills et al., 2023). Based on this temporal and functional concordance, we describe this observation as a “suggestive ecological correlation”—where oxidative environmental conditions and ecological complexity created a favorable context for P450 family expansion. Furthermore, this “ecological driving” hypothesis is also reflected in the distribution patterns of extant species: SEPs in Bacteria are significantly enriched in plant rhizosphere soils (Supplementary Table S5,  $P = 1.1 \times 10^{-14}$ ); in Fungi, SEPs are concentrated in saprotrophic *Ascomycota* (Supplementary Table S5,  $P = 1.1 \times 10^{-26}$ ); and in Metazoa, Lepidoptera dominate SEP representation (Supplementary Table S5,

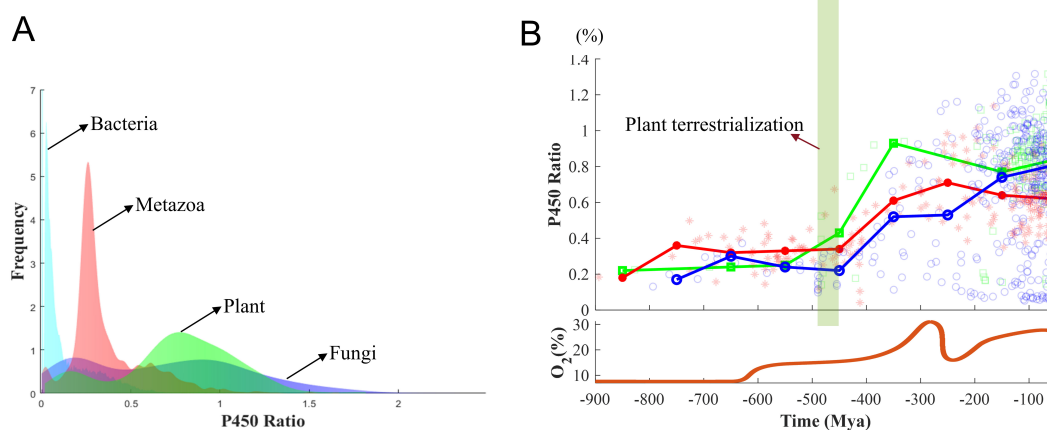


FIGURE 2

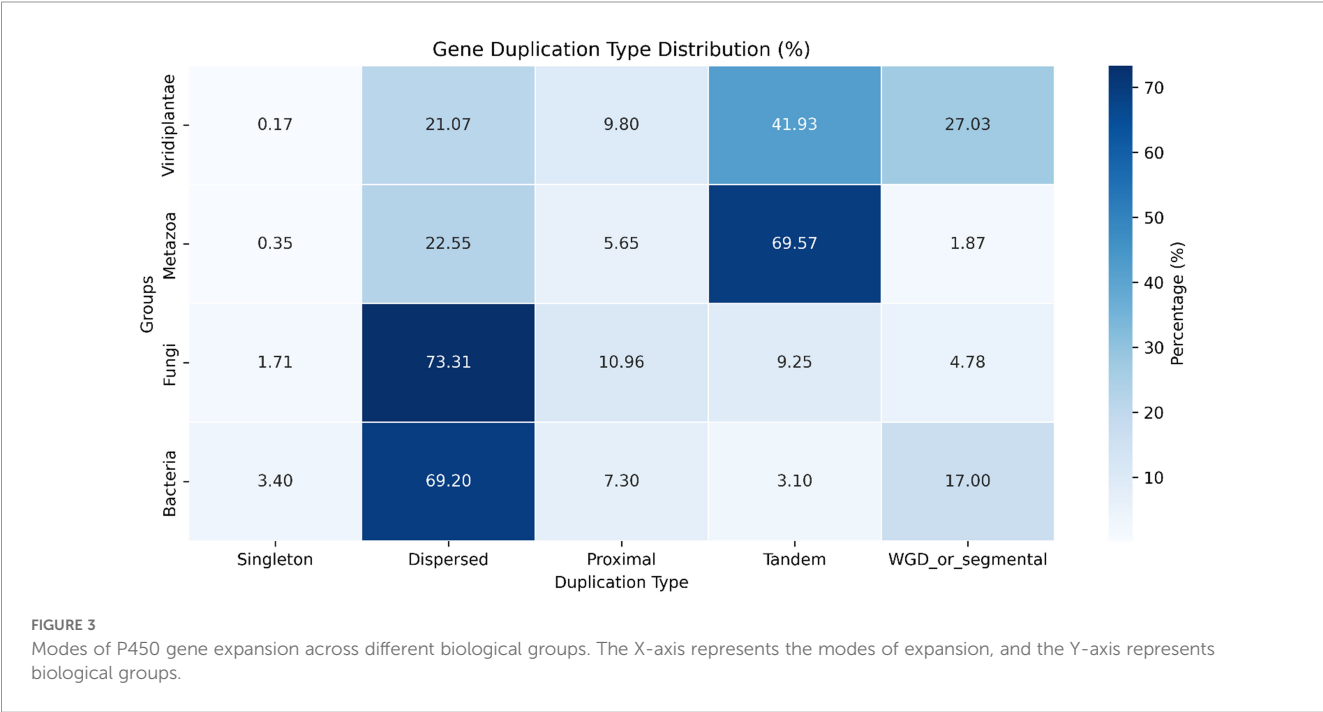
Differences in the distribution of P450 genes across biological groups. (A) Distribution of the proportions of P450 genes in the genomes of Viridiplantae, Fungi, Metazoa, and Bacteria. (B) Evolutionary trajectory of P450 gene numbers. The upper panel shows the evolutionary trend of the genomic proportion of P450 genes over the past 900 million years, while the lower panel depicts the fluctuation curve of atmospheric oxygen concentration.

$P = 7.2 \times 10^{-101}$ ). By contrast, species with low P450 gene content are significantly enriched in anaerobic environments (e.g., methanogens) or obligate parasitism (e.g., *Mycobacterium leprae*) (Warrilow et al., 2009). Strikingly, such ecological selection pressures can reshape the distribution of P450 genes within relatively short evolutionary timescales. For example, within the genus *Mycobacterium*, the rhizosphere-associated *Mycobacterium rhizomassiliense* carries ~70-fold more P450 genes than the obligate parasite *Mycobacterium leprae* (Supplementary Figure S11). This dramatic disparity corroborates a positive correlation between ecological niche complexity and the scale of P450 repertoires, offering a novel molecular evolutionary perspective for understanding environmental adaptation. With the intensification of oxidative conditions at the Earth’s surface and the increasing complexity of ecosystem hierarchies, the expansion of the P450 gene superfamily may have gained evolutionary opportunities during this process. It should be emphasized that current evidence is still primarily based on temporal correlations and distributional statistics, the mechanisms underlying this rapid expansion remain unclear.

3.3 Patterns of P450 gene superfamily expansion

To elucidate the mechanisms of P450 gene expansion across different biological groups, we used the *Duplicate\_gene\_classifier* module of MCScanX to analyze the expansion modes of P450 genes within SEPs. The results revealed significant differences in the composition and dominant types of expansion patterns among biological groups, with each group exhibiting distinct combinations of duplication mechanisms (Figure 3, Supplementary Figure S12). In Viridiplantae, TD and WGD contributed 41.93% and 27.03% of

expansion events, respectively. This “WGD–TD” combinatorial mechanism often promotes stepwise innovation to enhance ecological adaptability: WGD events provide long-term evolutionary material for the P450 family, while TD enables the rapid formation of gene clusters in response to environmental or microbial challenges (Chakraborty et al., 2023; Sahoo et al., 2023). For example, 77.97% of CYP71 genes in *Arabidopsis thaliana* were generated through TD (Supplementary Figure S13, Supplementary Table S11). In contrast, expansion in Metazoa is dominated by TD (69.57%), where gene clusters undergo subfunctionalization to achieve functional innovation. Similarly, TD and DSD account for 39.47% and 38.60%, respectively, of the duplication patterns of CYP2 genes in *Branchiostoma lanceolatum* (Supplementary Figure S14, Supplementary Table S12). Studies have shown that members of the CYP2K gene cluster in fish have differentiated into substrate-specific detoxification functions, while the insect CYP6 family enhances pesticide metabolism through dose-response (Edi et al., 2014; Lee et al., 2018). In Fungi, P450 genes exhibit a unique dispersed duplication pattern (73.31%). Such scattered distribution places genes in relatively independent regulatory contexts, facilitating the evolution of novel catalytic functions and thereby promoting the diversification of metabolic pathways. A representative case is the white-rot fungus *Phanerochaete chrysosporium*, whose large P450 repertoire originates not only from extensive ancestral duplication and genome rearrangements, but also from gene clustering and alternative splicing (Li et al., 2025). Bacteria predominantly exhibit a dispersed replication pattern (69.20%), in which replication can occur at multiple loci across the genome rather than being restricted to the primary chromosomal origin (oriC). This mode of duplication facilitates the movement, amplification, and expression of mobile genetic elements such as integrons, transposons, and insertion sequences,



which often carry antibiotic resistance genes. For example, *Staphylococcus aureus* possesses a single chromosomal replication origin, yet integrons within its genome can initiate localized replication and rearrangements at different loci, forming a dispersed replication pattern. Widely distributed class I integrons can harbor diverse resistance gene cassettes (e.g., *aadA*, *dfrA*), and their ability to replicate and integrate independently allows bacteria to rapidly increase the copy number and expression of resistance genes. This enhanced gene dosage and flexibility in acquiring new resistance traits underlie the survival advantage of bacteria under antibiotic stress (Sabbagh et al., 2021).

### 3.4 Significant expansion of the P450 gene superfamily

Finally, we analyzed which specific P450 gene families underwent significant expansion. Following the standard

classification of P450 enzyme gene families, all P450 enzymes were divided into 3,105 families (Supplementary Table S2). For each family, we calculated two metrics across the four biological groups: (1) species coverage, defined as the proportion of analyzed species in which the family is present; and (2) the maximum proportion within a single species, defined as the highest relative abundance of the family among the species in which it occurs (Figure 4A, Supplementary Table S6). A small number of P450 families display low copy numbers yet are highly conserved. For example, the CYP51 family is present in nearly all eukaryotes and primarily functions in sterol biosynthesis, but it contains only a few copies in each species. Our analysis showed that the CYP51 family exhibits species coverage exceeding 90% in both plants and fungi, with an average copy number maintained between 1 and 3 (Lepesheva and Waterman, 2007) (Supplementary Table S6). CYP73, a core enzyme in the phenylpropanoid metabolic pathway, is retained in almost all land plants, but typically only 1–5 copies exist per genome (Knosp et al., 2024) (Supplementary

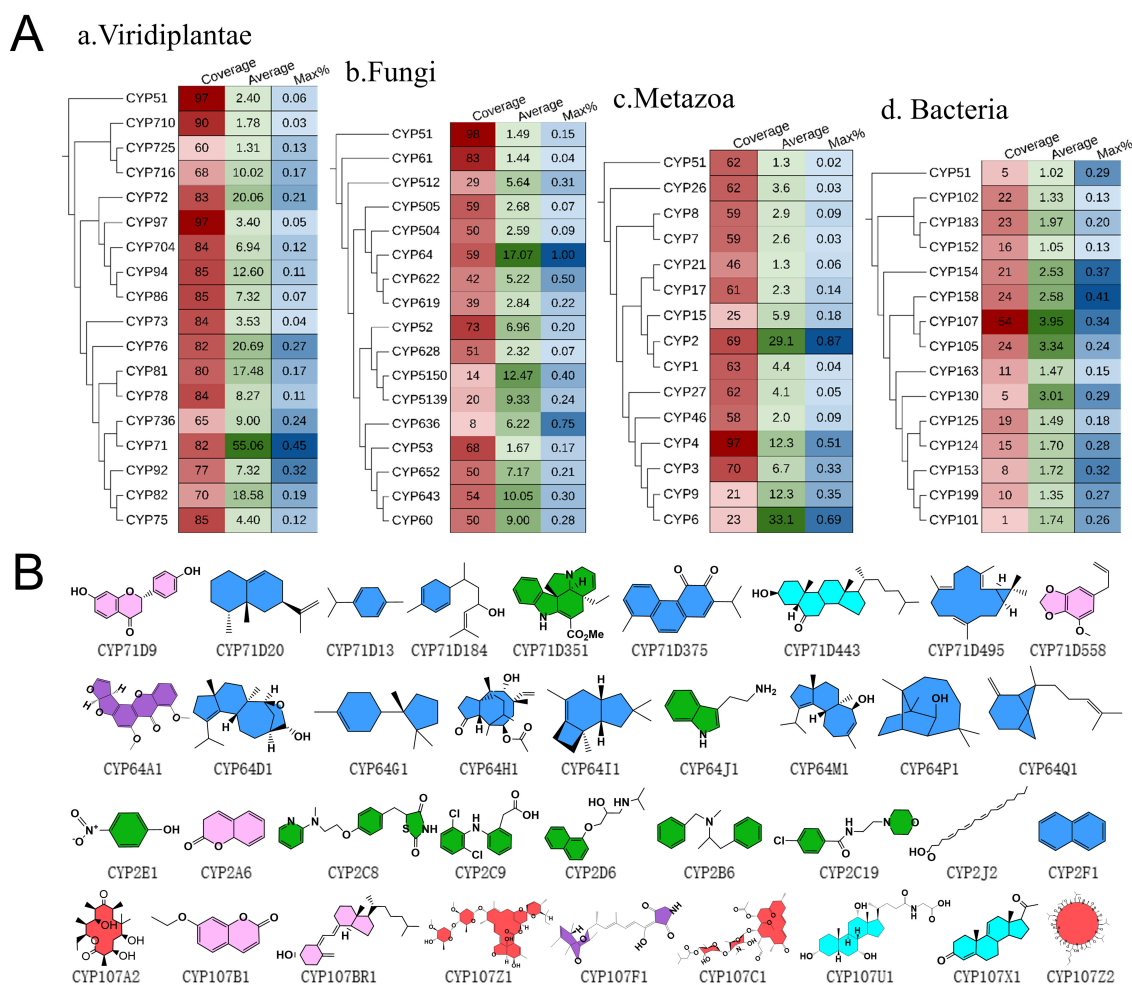


FIGURE 4

Significantly expanded P450 gene families and schematic representations of their substrate molecules. (A) Heatmap showing species coverage of different gene families (red), average copy number across covered species (green), and maximum proportion within a single species (blue). The top ten families for each category were selected (after removing redundancies) to construct the phylogenetic tree–heatmap. a, b, c, and d correspond to the distributions of significantly expanded families in Viridiplantae, Fungi, Metazoa, and Bacteria, respectively. (B) Schematic diagram of catalytic substrates of the four most significantly extended P450 gene families, with different colors representing different substrate categories.



Table S6). Similarly, CYP78 is widely present in angiosperms and is primarily involved in regulating plant development, particularly in controlling the growth of organs such as flowers, fruits, and seeds. This family shows a species coverage of approximately 84% in plants, but its members are also limited in number, with an average of about eight copies per genome (Fang et al., 2012) (Supplementary Table S6). In other words, although these families are broadly distributed at the species level, their functions have been largely fixed as core metabolic or developmental components, and therefore they have not undergone large-scale expansion.

In contrast, the CYP71, CYP64, CYP2, and CYP107 families not only exhibit extremely high species coverage within their respective biological groups but also show very high maximum proportions within individual species (Supplementary Table S6). CYP71 primarily catalyzes the biosynthesis of plant secondary metabolites, such as flavonoids, coumarins, and monoterpenes (Banerjee and Hamberger, 2018; Liu et al., 2020a). CYP64 participates in the oxidation of fungal toxins, antibiotics, and complex hydrocarbons (Shin et al., 2018). CYP2 catalyzes the metabolism of endogenous hormones and xenobiotics in Metazoa (Zanger and Schwab, 2013), whereas CYP107 is responsible for the oxidative modification of bacterial antibiotics and natural products (Kelly and Kelly, 2013) (Figure 4B). Overall, these families have enhanced organismal adaptability to environmental stresses and chemical diversity through gene expansion and functional diversification, representing typical P450 families with both high species coverage and high relative abundance in each biological groups.

In addition to the two types described above, some P450 families exhibit low species coverage but undergo dramatic expansion in a limited number of species. For example, CYP725 is a P450 family specific to a few gymnosperms, with 68 species-specific CYP725 genes identified in the genome of *Taxus yunnanensis* (Song et al., 2021). The CYP725 family has undergone remarkable gene expansion and functional diversification in the genus *Taxus*, which is closely associated with the biosynthesis of the anticancer drug paclitaxel. In particular, members such as CYP725A4 have been experimentally confirmed to catalyze key early hydroxylation reactions in the taxane skeleton, further indicating that the expansion of this family reflects lineage-specific evolution driven by the paclitaxel biosynthetic pathway (Nelson and Werck-Reichhart, 2011; Xiong et al., 2021). Taken together, P450 gene superfamily expansion displays three typical patterns: (1) core families with high species coverage but low copy numbers, maintaining fundamental metabolic or developmental functions; (2) expanded families with high species coverage and high copy numbers, enhancing environmental adaptability through functional diversification; and (3) families with low species coverage but high copy numbers (i.e., specialized metabolism-type expanded families), reflecting pathway-driven or ecological pressure-specific specialization. We define the criteria for a “specialized metabolic expansion family” as gene families with low species coverage and high copy numbers, whose expansion is directly associated with the formation or stabilization of specific metabolic pathways (Hansen et al., 2021).

## 4 Conclusion

This study demonstrates that the evolution of P450 genes is not merely a process of quantitative expansion, but rather a dynamic balance across three dimensions: stability, flexibility, and directionality. Core families maintain basic metabolism and development through low-copy numbers, ensuring system stability; widely expanded families strengthen environmental adaptation through high-copy numbers and functional diversification, showcasing flexibility; specialized families rapidly expand in specific pathways to meet unique metabolic needs, reflecting directionality. This pattern is rooted in the vertical inheritance that sustains core functions and is realized through diversified replication modes driven by ecological pressures and resource complexity.

Although this study has identified different expansion patterns of P450 genes in fungi, bacteria, and Viridiplantae, such as dispersed duplication in fungi and bacteria, the ‘WGD–TD’ combination in Viridiplantae, and TD-dominated clustered expansion in animals, the current results do not provide sufficient quantitative support for a strong coupling between these expansion pathways and ecological niche complexity. Future empirical studies and data validation are needed.

Overall, the evolutionary pattern of P450 genes reveals how gene families balance stability and innovation, thereby shaping life’s ongoing adaptation in complex environments. This understanding not only provides a new explanatory framework for gene family evolution mechanisms but also offers theoretical support for metabolic regulation and biotechnological practices. Furthermore, the study develops a novel analytical approach based on protein language models (ESM) and reveals multiple potential horizontal gene transfer (HGT) events within the P450 gene superfamily, further expanding our understanding of gene family evolutionary diversity and its adaptability.

## Data availability statement

The original contributions presented in the study are publicly available. The source code of the HGT identification model, including the training script, model architecture and sample dataset, can all be obtained at <https://github.com/JiangLab2020/P450HGT>. Link to the refined ESM-1b model for phylogenetic classification of P450 enzymes: <https://zenodo.org/records/17547642>. The corresponding P450 sequences have been deposited in the Plant Cytochrome P450 Database: <http://p450.biodesign.ac.cn/> (Wang et al., 2021).

## Author contributions

YS: Writing – original draft, Writing – review & editing, Investigation, Data curation, Visualization, Methodology. SZ: Writing – original draft, Writing – review & editing, Visualization, Conceptualization, Methodology, Data curation. HZ: Writing – review & editing, Methodology, Software, Data curation, Visualization. SL: Writing – review & editing, Data curation, Investigation, Visualization. JY: Writing – review & editing, Investigation. LL: Writing – review & editing, Resources. XL: Funding acquisition, Resources, Supervision,

Writing – review & editing, JC: Conceptualization, Writing – review & editing, Supervision, Investigation, Visualization, Funding acquisition, Resources, Project administration. HJ: Funding acquisition, Resources, Writing – review & editing.

## Funding

The author(s) declared that financial support was received for this work and/or its publication. This project has received funding from the Strategic Priority Research Program of the Chinese Academy of Sciences XDC0110200, National Natural Science Foundation of China (32571467, 32371499, 12326611), COMSATS Joint Center for Industrial Biotechnology (No. TSBICIP-IJCP-001), Tianjin Synthetic Biotechnology Innovation Capacity Improvement Project (No. TSBICIP-IJCP-002, TSBICIP-CYFH-011, TSBICIP-KJGG-009-02, TSBICIP-KJGG-008-02, TSBICIP-KJGG-018, TSBICIP-PTJJ-012), Hubei University of Technology High-Level Talent Research Startup Fund Program (4301/00960), 2025 Wuhan Natural Science Foundation Exploration Project (Dawn Program) (2025040601020158), major research projects of the Haihe Laboratory of Synthetic Biology (No. 22HHSWSS00005 and 22HHSWSS00004) and Science and Technology Major Project of Guangxi (Project Nos. Guike AA24206048 and AA24206050).

## Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## References

- An, S., Liu, Y., Sang, K., Wang, T., Yu, J., Zhou, Y., et al. (2023). Brassinosteroid signaling positively regulates abscisic acid biosynthesis in response to chilling stress in tomato. *JIPB* 65, 10–24. doi: 10.1111/jipb.13356
- Banerjee, A., and Hamberger, B. (2018). P450s controlling metabolic bifurcations in plant terpene specialized metabolism. *Phytochem. Rev.* 17, 81–111. doi: 10.1007/s11101-017-9530-4
- Baron, K. N., Schroeder, D. F., and Stasolla, C. (2012). Transcriptional response of abscisic acid (ABA) metabolism and transport to cold and heat stress applied at the reproductive stage of development in *Arabidopsis thaliana*. *Plant Sci.* 188–189, 48–59. doi: 10.1016/j.plantsci.2012.03.001
- Boto, L. (2015). “Horizontal gene transfer in evolution,” in *eLS* (Chichester, UK: John Wiley & Sons, Ltd), 1–7. doi: 10.1002/9780470015902.a0022835.pub2
- Buchfink, B., Reuter, K., and Drost, H.-G. (2021). Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat. Methods* 18, 366–368. doi: 10.1038/s41592-021-01101-x
- Chakraborty, P., Biswas, A., Dey, S., Bhattacharjee, T., and Chakrabarty, S. (2023). Cytochrome P450 gene families: role in plant secondary metabolites production and plant defense. *JoX* 13, 402–423. doi: 10.3390/jox13030026
- Chen, W., Lee, M.-K., Jefcoate, C., Kim, S.-C., Chen, F., and Yu, J.-H. (2014). Fungal cytochrome P450 monooxygenases: their distribution, structure, functions, family expansion, and evolutionary origin. *Genome Biol. Evol.* 6, 1620–1634. doi: 10.1093/gbe/evu132
- Cheng, J., Chen, J., Liu, X., Li, X., Zhang, W., Dai, Z., et al. (2021a). The origin and evolution of the diosgenin biosynthetic pathway in yam. *Plant Commun.* 2, 100079. doi: 10.1016/j.xplc.2020.100079
- Cheng, J., Wang, X., Liu, X., Zhu, X., Li, Z., Chu, H., et al. (2021b). Chromosome-level genome of Himalayan yew provides insights into the origin and evolution of the paclitaxel biosynthetic pathway. *Mol. Plant* 14, 1199–1209. doi: 10.1016/j.molp.2021.04.015
- Costello, C. M., Cain, S. L., Pils, S., Frattaroli, L., Haroldson, M. A., and Van Manen, F. T. (2016). Diet and macronutrient optimization in wild ursids: A comparison of grizzly bears with sympatric and allopatric black bears. *PLoS One* 11, e0153702. doi: 10.1371/journal.pone.0153702
- Crešnar, B., and Petrič, S. (2011). Cytochrome P450 enzymes in the fungal kingdom. *Biochim. Biophys. Acta* 1814, 29–35. doi: 10.1016/j.bbapap.2010.06.020
- Eddy, S. R. (2011). Accelerated profile HMM searches. *PLoS Comput. Biol.* 7, e1002195. doi: 10.1371/journal.pcbi.1002195
- Edi, C. V., Djogbénou, L., Jenkins, A. M., Regna, K., Muskavitch, M. A. T., Poupardin, R., et al. (2014). CYP6 P450 enzymes and ACE-1 duplication produce extreme and multiple insecticide resistance in the malaria mosquito *Anopheles Gambiae*. *PLoS Genet.* 10, e1004236. doi: 10.1371/journal.pgen.1004236
- Elnaggar, A., Heinzinger, M., Dallago, C., Rihawi, G., Wang, Y., Jones, L., et al. (2021). ProtTrans: towards cracking the language of life's code through self-supervised deep learning and high performance computing. *IEEE Transactions Pattern Anal. Machine Intelligence*. 44, 7112–7127. doi: 10.48550/arXiv.2007.06225
- Esteves, F., Rueff, J., and Kranendonk, M. (2021). The central role of cytochrome P450 in xenobiotic metabolism—A brief review on a fascinating enzyme family. *JoX* 11, 94–114. doi: 10.3390/jox11030007
- Fang, W., Wang, Z., Cui, R., Li, J., and Li, Y. (2012). Maternal control of seed size by EOD3/CYP78A6 in *Arabidopsis thaliana*. *Plant J.* 70, 929–939. doi: 10.1111/j.1365-3113.2012.04907.x
- Goh, Y.-X., Anupou, S. M. B., Nguyen, A., Zhang, H., Ponder, M., Krometis, L.-A., et al. (2024). Evidence of horizontal gene transfer and environmental selection impacting antibiotic resistance evolution in soil-dwelling *Listeria*. *Nat. Commun.* 15, 10034. doi: 10.1038/s41467-024-54459-9

The author JC declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

## Generative AI statement

The author(s) declared that generative AI was not used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fevo.2025.1713618/full#supplementary-material>

- Gouy, M., Guindon, S., and Gascuel, O. (2010). SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol. Biol. Evol.* 27, 221–224. doi: 10.1093/molbev/msp259
- Guan, R., Zhao, Y., Zhang, H., Fan, G., Liu, X., Zhou, W., et al. (2016). Draft genome of the living fossil Ginkgo biloba. *Gigascience* 5, 49. doi: 10.1186/s13742-016-0154-1
- Guengerich, F. P. (2002). Cytochrome p450 enzymes in the generation of commercial products. *Nat. Rev. Drug Discov.* 1, 359–366. doi: 10.1038/nrd792
- Hansen, C. C., Nelson, D. R., Møller, B. L., and Werck-Reichhart, D. (2021). Plant cytochrome P450 plasticity and evolution. *Mol. Plant* 14, 1244–1265. doi: 10.1016/j.molp.2021.06.028
- Hao, Y., Dong, Z., Zhao, Y., Tang, W., Wang, X., Li, J., et al. (2022). Phylogenomic analysis of cytochrome P450 multigene family and its differential expression analysis in pepper (*Capsicum annuum* L.). *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.1078377
- Heled, J., and Drummond, A. J. (2010). Bayesian inference of species trees from multilocus data. *Mol. Biol. Evol.* 27, 570–580. doi: 10.1093/molbev/msp274
- Huang, Q., Zhang, X., Sun, G., Qiu, R., Luo, L., Wang, C., et al. (2025). Decoding the mechanism of P450-catalyzed aromatic hydroxylation: Uncovering the arene oxide pathway and insights into the regioselectivity. *Chin. J. Catalysis* 70, 420–430. doi: 10.1016/S1872-2067(24)60234-2
- Ichinose, H. (2013). Cytochrome P 450 of wood-rotting basidiomycetes and biotechnological applications. *Biotech. App Biochem.* 60, 71–81. doi: 10.1002/bab.1061
- Jiang, S.-Y., Jin, J., Sarojam, R., and Ramachandran, S. (2019). A comprehensive survey on the terpene synthase gene family provides new insight into its evolutionary patterns. *Genome Biol. Evol.* 11, 2078–2098. doi: 10.1093/gbe/evz142
- Johnson, R. N., O'Meally, D., Chen, Z., Etherington, G. J., Ho, S. Y. W., Nash, W. J., et al. (2018). Adaptation and conservation insights from the koala genome. *Nat. Genet.* 50, 1102–1111. doi: 10.1038/s41588-018-0153-5
- Kang, Y. J., Kim, K. H., Shim, S., Yoon, M. Y., Sun, S., Kim, M. Y., et al. (2012). Genome-wide mapping of NBS-LRR genes and their association with disease resistance in soybean. *BMC Plant Biol.* 12, 139. doi: 10.1186/1471-2229-12-139
- Kapoor, B., Kumar, P., Verma, V., Irfan, M., Sharma, R., and Bhargava, B. (2023). How plants conquered land: evolution of terrestrial adaptation. *J. Evolutionary Biol.* 36, 5–14. doi: 10.1111/jeb.14062
- Kawai, Y., Ono, E., and Mizutani, M. (2014). Expansion of specialized metabolism-related superfamily genes via whole genome duplications during angiosperm evolution. *Plant Biotechnol.* 31, 579–584. doi: 10.5511/plantbiotechnology.14.0901a
- Keeling, P. J. (2024). Horizontal gene transfer in eukaryotes: aligning theory with data. *Nat. Rev. Genet.* 25, 416–430. doi: 10.1038/s41576-023-00688-5
- Kelly, S. L., and Kelly, D. E. (2013). Microbial cytochromes P450: biodiversity and biotechnology. Where do cytochromes P450 come from, what do they do and what can they do for us? *Phil. Trans. R. Soc B* 368, 20120476. doi: 10.1098/rstb.2012.0476
- Kingma, D. P., and Ba, J. (2017). *Adam: A method for stochastic optimization*. (San Diego, CA, USA: International Conference on Learning Representations, ICLR). doi: 10.48550/arXiv.1412.6980
- Knosps, S., Kriegshauser, L., Tatsumi, K., Malherbe, L., Erhardt, M., Wiedemann, G., et al. (2024). An ancient role for CYP73 monooxygenases in phenylpropanoid biosynthesis and embryophyte development. *EMBO J.* 43, 4092–4109. doi: 10.1038/s44318-024-00181-7
- Kondo, M., Ikenaka, Y., Nakayama, S. M. M., Kawai, Y. K., and Ishizuka, M. (2022). Specific gene duplication and loss of cytochrome P450 in families 1–3 in carnivora (Mammalia, laurasiatheria). *Animals* 12, 2821. doi: 10.3390/ani12202821
- Kumar, S., Suleski, M., Craig, J. M., Kasprowitz, A. E., Sanderford, M., Li, M., et al. (2022). TimeTree 5: an expanded resource for species divergence times. *Mol. Biol. Evol.* 39, msac174. doi: 10.1093/molbev/msac174
- Lamb, D. C., Follmer, A. H., Goldstone, J. V., Nelson, D. R., Warrilow, A. G., Price, C. L., et al. (2019). On the occurrence of cytochrome P450 in viruses. *Proc. Natl. Acad. Sci.* 116, 12343–12352. doi: 10.1073/pnas.1901080116
- Lamb, D. C., Lei, L., Warrilow, A. G. S., Lepesheva, G. I., Mullins, J. G. L., Waterman, M. R., et al. (2009). The first virally encoded cytochrome P450. *J. Virol.* 83, 8266–8269. doi: 10.1128/JVI.00289-09
- Lee, B.-Y., Kim, D.-H., Kim, H.-S., Kim, B.-M., Han, J., and Lee, J.-S. (2018). Identification of 74 cytochrome P450 genes and co-localized cytochrome P450 genes of the CYP2K, CYP5A, and CYP46A subfamilies in the mangrove killifish *Kryptolebias marmoratus*. *BMC Genomics* 19, 7. doi: 10.1186/s12864-017-4410-2
- Lepesheva, G. I., and Waterman, M. R. (2007). Sterol 14 $\alpha$ -demethylase cytochrome P450 (CYP51), a P450 in all biological kingdoms. *Biochim. Biophys. Acta (BBA) - Gen. Subj.* 1770, 467–477. doi: 10.1016/j.bbagen.2006.07.018
- Li, H., Zhang, X., Yang, Q., Shanguan, X., and Ma, Y. (2024). Genome-wide identification and tissue expression pattern analysis of TPS gene family in soybean (*Glycine max*). *Front. Plant Sci.* 15. doi: 10.3389/fpls.2024.1487092
- Li, J., Pi, C., Zhang, J., Jiang, F., Bao, T., Gao, L., et al. (2025). Fungal bioconversion of lignin-derived aromatics: Pathways, enzymes, and biotechnological potential. *Biotechnol. Adv.* 83, 108624. doi: 10.1016/j.biotechadv.2025.108624
- Li, L.-G., and Zhang, T. (2023). Plasmid-mediated antibiotic resistance gene transfer under environmental stresses: Insights from laboratory-based studies. *Sci. Total Environ.* 887, 163870. doi: 10.1016/j.scitotenv.2023.163870
- Li, L., Stoeckert, C. J., and Roos, D. S. (2003). OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* 13, 2178–2189. doi: 10.1101/gr.1224503
- Liu, X., Cheng, J., Zhu, X., Zhang, G., Yang, S., Guo, X., et al. (2020a). *De novo* biosynthesis of multiple pinocembrin derivatives in *saccharomyces cerevisiae*. *ACS Synth. Biol.* 9, 3042–3051. doi: 10.1021/acssynbio.0c00289
- Liu, Y., Wang, Q., Liu, X., Cheng, J., Zhang, L., Chu, H., et al. (2023). pUGTdb: A comprehensive database of plant UDP-dependent glycosyltransferases. *Mol. Plant* 16, 643–646. doi: 10.1016/j.molp.2023.01.003
- Liu, S., Yang, S., and Su, P. (2024). Chemo-enzymatic synthesis of bioactive compounds from traditional Chinese medicine and medicinal plants. *Sci. Tradit. Chin. Med.* 2, 95–103. doi: 10.1097/st9.0000000000000027
- Liu, X., Zhu, X., Wang, H., Liu, T., Cheng, J., and Jiang, H. (2020b). Discovery and modification of cytochrome P450 for plant natural products biosynthesis. *Synthetic Syst. Biotechnol.* 5, 187–199. doi: 10.1016/j.synbio.2020.06.008
- Lynch, M., and Conery, J. S. (2000). The evolutionary fate and consequences of duplicate genes. *Science* 290, 1151–1155. doi: 10.1126/science.290.5494.1151
- Maaten, L. (2008). Visualizing Data using t-SNE. *J. Mach. Learn. Res.* 9, 2579–2605.
- Mendes, F. K., Vanderpool, D., Fulton, B., and Hahn, M. W. (2021). CAFE 5 models variation in evolutionary rates among gene families. *Bioinformatics* 36, 5516–5518. doi: 10.1093/bioinformatics/btaa1022
- Mills, B. J. W., Krause, A. J., Jarvis, I., and Cramer, B. D. (2023). Evolution of atmospheric O<sub>2</sub> through the phanerozoic, revisited. *Annu. Rev. Earth Planetary Sci.* 51, 253–276. doi: 10.1146/annurev-earth-032320-095425
- Mizutani, M., and Ohta, D. (2010). Diversification of P450 genes during land plant evolution. *Annu. Rev. Plant Biol.* 61, 291–315. doi: 10.1146/annurev-arplant-042809-112305
- Mokhosoev, I. M., Astakhov, D. V., Terentiev, A. A., and Moldogazieva, N. T. (2024). Cytochrome P450 monooxygenase systems: Diversity and plasticity for adaptive stress response. *Prog. Biophysics Mol. Biol.* 193, 19–34. doi: 10.1016/j.pbiomolbio.2024.09.003
- Moreira, D. (2011). “Horizontal gene transfer: mechanisms and evolutionary consequences,” in *Origins and evolution of life: an astrobiological perspective*. Eds. H. Martin, M. Gargaud and P. López-García (Cambridge University Press, Cambridge), 313–325. doi: 10.1017/CBO9780511933875.022
- Nelson, D. R. (2018). Cytochrome P450 diversity in the tree of life. *Biochim. Biophys. Acta (BBA) - Proteins Proteomics* 1866, 141–154. doi: 10.1016/j.bbapap.2017.05.003
- Nelson, D., and Werck-Reichhart, D. (2011). A P450-centric view of plant evolution. *Plant J.* 66, 194–211. doi: 10.1111/j.1365-3113.2011.04529.x
- Nelson, D. R., Koymans, L., Kamataki, T., Stegeman, J. J., Feyereisen, R., Waxman, D. J., et al. (1996). P450 superfamily: update on new sequences, gene mapping, accession numbers and nomenclature. *Pharmacogenetics* 6, 1–42. doi: 10.1097/00008571-199602000-00002
- Ngcobo, P. E., Nkosi, B. V. Z., Chen, W., Nelson, D. R., and Syed, K. (2023). Evolution of cytochrome P450 enzymes and their redox partners in archaea. *Int. J. Mol. Sci.* 24, 4161. doi: 10.3390/ijms24044161
- Parvez, M., Qhanya, L. B., Mthakathi, N. T., Kgosiemang, I. K. R., Bamal, H. D., Pagadala, N. S., et al. (2016). Molecular evolutionary dynamics of cytochrome P450 monooxygenases across kingdoms: Special focus on mycobacterial P450s. *Sci. Rep.* 6, 33099. doi: 10.1038/srep33099
- Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., et al. (2019). “PyTorch: an imperative style, high-performance deep learning library,” in *Advances in neural information processing systems* (Red Hook, NY: Curran Associates, Inc). Available online at: <https://proceedings.neurips.cc/paper/2019/hash/bdbca288fee7f92f2bfa9f7012727740-Abstract.html> (Accessed December 2025).
- Price, M. N., Dehal, P. S., and Arkin, A. P. (2010). FastTree 2 – approximately maximum-likelihood trees for large alignments. *PloS One* 5, e9490. doi: 10.1371/journal.pone.0009490
- Qiao, X., Li, Q., Yin, H., Qi, K., Li, L., Wang, R., et al. (2019). Gene duplication and evolution in recurring polyploidization–diploidization cycles in plants. *Genome Biol.* 20, 38. doi: 10.1186/s13059-019-1650-2
- Rives, A., Meier, J., Sercu, T., Goyal, S., Lin, Z., Liu, J., et al. (2021). Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences. *Proc. Natl. Acad. Sci. U.S.A.* 118, e2016239118. doi: 10.1073/pnas.2016239118
- Sabbagh, P., Rajabnia, M., Maali, A. M., and Ferdosi-Shahandashti, E. (2021). Integron and its role in antimicrobial resistance: A literature review on some bacterial pathogens. *Iranian J. Basic Med. Sci.* 24, 136–142. doi: 10.22038/ijbms.2020.48905.11208
- Sahoo, B., Nayak, I., Parameswaran, C., Kesawat, M. S., Sahoo, K. K., Subudhi, H. N., et al. (2023). A comprehensive genome-wide investigation of the cytochrome 71 (OsCYP71) gene family: revealing the impact of promoter and gene variants (Ser33Leu) of osCYP71P6 on yield-related traits in indica rice (*Oryza sativa* L.). *Plants* 12, 3035. doi: 10.3390/plants12173035
- Samynathan, R. (2024). Horizontal gene transfer as a direct cause of antibiotic resistance in bacterial pathogens. *J. Student Res.* 13, 1–16. doi: 10.47611/jsrhs.v13i4.8308
- Shin, J., Kim, J.-E., Lee, Y.-W., and Son, H. (2018). Fungal cytochrome P450s and the P450 complement (CYPome) of fusarium graminearum. *Toxins* 10, 112. doi: 10.3390/toxins10030112

- Shoun, H., Fushinobu, S., Jiang, L., Kim, S.-W., and Wakagi, T. (2012). Fungal denitrification and nitric oxide reductase cytochrome P450nor. *Philos. Trans. R Soc. Lond B Biol. Sci.* 367, 1186–1194. doi: 10.1098/rstb.2011.0335
- Song, C., Fu, F., Yang, L., Niu, Y., Tian, Z., He, X., et al. (2021). *Taxus yunnanensis* genome offers insights into gymnosperm phylogeny and taxol production. *Commun. Biol.* 4, 1203. doi: 10.1038/s42003-021-02697-8
- Syed, K., and Yadav, J. S. (2012). P450 monooxygenases (P450ome) of the model white rot fungus *Phanerochaete chrysosporium*. *Crit. Rev. Microbiol.* 38, 339–363. doi: 10.3109/1040841X.2012.682050
- Tang, M., Zhang, W., Lin, R., Li, L., He, L., Yu, J., et al. (2024). Genome-wide characterization of cytochrome P450 genes reveals the potential roles in fruit ripening and response to cold stress in tomato. *Physiologia Plantarum* 176, e14332. doi: 10.1111/ppl.14332
- Wang, Y., Tang, H., DeBarry, J. D., Tan, X., Li, J., Wang, X., et al. (2012). MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* 40, e49. doi: 10.1093/nar/gkr1293
- Wang, H., Wang, Q., Liu, Y., Liao, X., Chu, H., Chang, H., et al. (2021). PCPD: Plant cytochrome P450 database and web-based tools for structural construction and ligand docking. *Synthetic Syst. Biotechnol.* 6, 102–109. doi: 10.1016/j.synbio.2021.04.004
- Wang, X., Wang, L., Wang, Y., Liu, H., Hu, D., Zhang, N., et al. (2018). Arabidopsis PCaP2 plays an important role in chilling tolerance and ABA response by activating CBF- and snRK2-mediated transcriptional regulatory network. *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.00215
- Warrilow, A. G. S., Jackson, C. J., Parker, J. E., Marczylo, T. H., Kelly, D. E., Lamb, D. C., et al. (2009). Identification, Characterization, and Azole-Binding Properties of *Mycobacterium smegmatis* CYP164A2, a Homolog of ML2088, the Sole Cytochrome P450 Gene of *Mycobacterium leprae*. *Antimicrob. Agents Chemother.* 53, 1157–1164. doi: 10.1128/AAC.01237-08
- Wei, K., and Chen, H. (2018). Global identification, structural analysis and expression characterization of cytochrome P450 monooxygenase superfamily in rice. *BMC Genomics* 19, 35. doi: 10.1186/s12864-017-4425-8
- Xiong, X., Gou, J., Liao, Q., Li, Y., Zhou, Q., Bi, G., et al. (2021). The *Taxus* genome provides insights into paclitaxel biosynthesis. *Nat. Plants* 7, 1026–1036. doi: 10.1038/s41477-021-00963-5
- Zanger, U. M., and Schwab, M. (2013). Cytochrome P450 enzymes in drug metabolism: Regulation of gene expression, enzyme activities, and impact of genetic variation. *Pharmacol. Ther.* 138, 103–141. doi: 10.1016/j.pharmthera.2012.12.007