GENETICS

Convergent evolution of berberine biosynthesis

Zhichao Xu[1,2]*†, Ya Tian[1,2]†, Jing Wang[1,2]†, Yuwei Ma[1,2], Qi Li[1,2], Yuanze Zhou[3], Wanran Zhang[1,2], Tingxia Liu[4], Lingzhe Kong[1,2], Yifan Wang[1,2], Ziyan Xie[1,2], Zhoujie An[1,2], Baojiang Zheng[1,2], Yuhong Zhang[5], Chang Cao[5], Chengwei Liu[1,2], Lixia Tian[6], Chengpeng Fan[7], Jiushi Liu[8], Hui Yao[8], Jingyuan Song[8], Baozhong Duan[9], Haitao Liu[8]*, Ranran Gao[4,10]*, Wei Sun[4]*, Shilin Chen[11]*

Berberine is an effective antimicrobial and antidiabetic alkaloid, primarily extracted from divergent botanical lineages, specifically *Coptis* (Ranunculales, early-diverging eudicot) and *Phellodendron* (Sapindales, core eudicot). In comparison with its known pathway in *Coptis* species, its biosynthesis in *Phellodendron* species remains elusive. Using chromosome-level genome assembly, coexpression matrix, and biochemical assays, we identified six key steps in berberine biosynthesis from *Phellodendron amurense*, including methylation, hydroxylation, and berberine bridge formation. Notably, we discovered a specific class of *O*-methyltransferases (NOMT) responsible for *N*-methylation. Structural analysis and mutagenesis of PaNOMT9 revealed its unique substrate-binding conformation. In addition, unlike the classical FAD-dependent berberine bridge formation in Ranunculales, *Phellodendron* uses a NAD(P)H-dependent monooxygenase (PaCYP71BG29) for berberine bridge formation, originating from the neofunctionalization of tryptamine 5-hydroxylase. Together, these findings reveal the convergence of berberine biosynthesis between *Coptis* and *Phellodendron* and signify the role of the convergent evolution in plant specialized metabolisms.

Copyright © 2024 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works. Distributed under a Creative Commons Attribution NonCommercial License 4.0 (CC BY-NC).

INTRODUCTION

Plant-derived benzylisoquinoline alkaloids (BIAs) are a resource of medicines. Numerous BIAs, such as berberine (*1*–*4*), morphine (*5*, *6*), and sanguinarine (*7*, *8*), have been used in the treatment of various diseases. Many investigations for the BIA isolation and distribution in the plant kingdom have been completed. On the one hand, it is interesting that a large proportion of BIAs is limited to specific angiosperm orders, such as early-diverging eudicots (e.g., Ranunculales and Proteales) and Magnoliids (e.g., Magnoliales, Laurales, and Piperales) (*9*–*12*). On the other hand, the occurrence of a few BIAs, such as berberine, follows different patterns. Berberine was initially separated from the bark of *Zanthoxylon clava*, a core eudicot plant (Sapindales). Continuous investigations have shown that *Phellodendron* spp. also produce berberine (*13*). To date, berberine production primarily relies on botanical sources like *Berberis* and *Coptis* from Ranunculales (*14*) and *Phellodendron* in Sapindales.

Berberine is an effective and generic antimicrobial and antidiabetic medicine. The past investigations have characterized the main biosynthetic steps of berberine in the roots and rhizomes in *Coptis* species (*15*). Its biosynthetic pathway starts with the coupling of dopamine (**1**) and 4-hydroxyphenylacetaldehyde (4HPAA; **2**) into (*S*)-norcoclaurine (**3**) catalyzed by norcoclaurine synthase (NCS), a member of PR10/Bet v1 family (*16*, *17*). Subsequent modifications including *O*/*N*-methylation by *O*-/*N*-methyltransferases (MTs) and hydroxylation by cytochromes P450 shape the formation of (*S*)-reticuline (**7**) (*18*–*20*), an essential intermediate of the BIA biosynthetic pathway. The berberine bridge enzyme (BBE) from flavin adenine dinucleotide (FAD)–dependent oxidase family catalyzes the conversion of **7** to the tetra-cyclic berberine backbone (*S*)-scoulerine (**8**) (*21*). Last, 9OMT (*22*), CYP719A (*23*, *24*), and STOX (BBE like) (*25*) are responsible for three consecutive steps—the *O*-methylation, methylenedioxy bridge formation, and four-electron oxidation reactions—to convert **8** to berberine (**11**) (Fig. 1A). With the elucidation of berberine biosynthesis, recent attention has focused on metabolic engineering and environmentally sustainable production in vitro. Although the de novo microbial synthesis of berberine has been achieved, current yields have not met industrial production requirements due to the limitation of catalytic efficiency and enzyme promiscuity (*26*, *27*). The findings of alternative biosynthetic enzymes will provide essential genetic elements conducive to achieving high-efficiency production of BIAs.

Despite the abovementioned pathway understood in *Coptis* spp., whether distantly related *Phellodendron* species rich in berberine share the same one remains unknown. On the basis of the fact of the different lineages between *Phellodendron* spp. and *Coptis* spp., we hypothesize that *Phellodendron* species might have evolved different mechanisms to biosynthesize berberine and other BIAs. To test this hypothesis, we used an integrative approach of genome assembly, transcriptomics, mutagenesis, genome comparison, phylogenetic analysis, protein crystallization, enzymatic assay, and metabolomics to dissect the biosynthesis pathway of berberine in *Phellodendron amurense*. The resulting data not only determined six biosynthetic steps of berberine in this lineage but also uncovered previous unknown enzymatic functions and mechanisms. OMT enzymes from *Phellodendron* catalyze the *N*-methylation of BIA intermediates via

[1]Key Laboratory of Saline-alkali Vegetation Ecology Restoration (Northeast Forestry University), Ministry of Education, Harbin 150040, China. [2]College of Life Science, Northeast Forestry University, Harbin 150040, China. [3]Department of Hematology, Southwest Hospital, Third Military Medical University (Army Medical University), Chongqing 400038, China. [4]Key Laboratory of Beijing for Identification and Safety Evaluation of Chinese Medicine, Institute of Chinese Materia Medica, China Academy of Chinese Medical Sciences, Beijing 100700, China. [5]College of Chemistry, Chemical Engineering and Resource Utilization, Northeast Forestry University, Harbin 150040, China. [6]School of Pharmaceutical Sciences, Guizhou University, Guiyang 550025, China. [7]School of Basic Medical Sciences, Wuhan University, Wuhan 430072, China. [8]Key Lab of Chinese Medicine Resources Conservation, State Administration of Traditional Chinese Medicine of the People's Republic of China, Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100193, China. [9]College of Pharmaceutical Science, Dali University, Dali 671003, China. [10]Artemisinin Research Center, Institute of Chinese Materia Medica, China Academy of Chinese Medical Sciences, Beijing 100700, China. [11]Institute of Herbgenomics, Chengdu University of Traditional Chinese Medicine, Chengdu 611137, China.
*Corresponding author. Email: zcxu@nefu.edu.cn (Z. Xu); htliu@implad.ac.cn (H.L.); rrgao1991@icmm.ac.cn (R.G.); wsun@icmm.ac.cn (W.S.); slchen@cdutcm.edu.cn (S.C.)
†These authors contributed equally to this work.

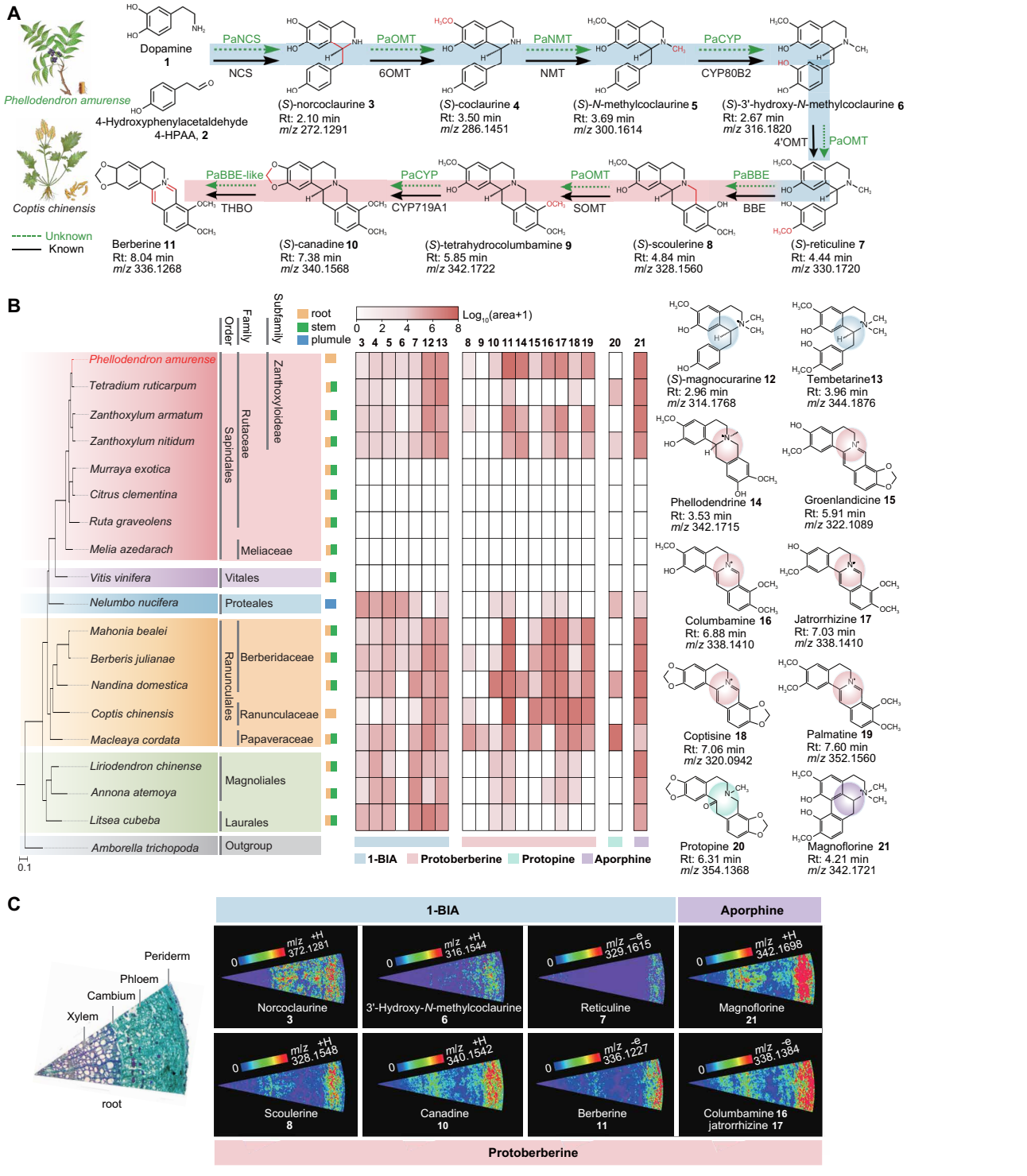

Fig. 1. Protoberberine and aporphine-type BIAs produced in *P. amurense*. (**A**) The biosynthetic pathway of berberine (**11**). The black arrows represent the enzymes reported from *Coptis*, and the dashed arrows mean that these steps in *P. amurense* remain elusive. **1**, dopamine; **2**, 4HPAA; **3**, (*S*)-norcoclaurine; **4**, (*S*)-coclaurine; **5**, (*S*)-*N*-methylcoclaurine; **6**, (*S*)-3′-hydroxy-*N*-methylcoclaurine; **7**, (*S*)-reticuline; **8**, (*S*)-scoulerine; **9**, (*S*)-tetrahydrocolumbamine; **10**, (*S*)-canadine; and **11**, berberine. Rt, retention time. (**B**) Phylogenetic tree and BIA metabolite profiles of tested species. The phylogenetic tree was constructed using single-copy orthologs from *P. amurense* and other 18 species, based on transcriptome or genome data (table S1). Metabolites were extracted from roots (yellow) and/or stems (green) of 17 species, as well as lotus plumule of *N. nucifera* (blue). The heatmap was developed with the BIA levels indicated by the values of $\log_{10}$(peak area + 1). BIA types: 1-BIA, protoberberine, protopine, and aporphine. **12**, (*S*)-magnocurarine; **13**, tembetarine; **15**, phellodendrine; **15**, groenlandicine; **16**, columbamine; **17**, jatrorrhizine; **18**, coptisine; **19**, palmatine; **20**, protopine; and **21**, magnoflorine. (**C**) A microscopic image shows a cross section of *P. amurense* roots, and MALDI-MSI images generated with heatmaps show the accumulation patterns of metabolites.

different mechanisms. Unlike the FAD-dependent BBE in *Coptis*, lineage-specific CYP450s were identified to catalyze the berberine bridge formation in *Phellodendron*. These findings characterize a divergent biosynthesis of berberine differentially evolved in *Coptis* and *Phellodendron*. The divergence of the berberine biosynthesis expands our understanding of BIA metabolism in plants and provides notable insights into metabolic engineering.

RESULTS

Abundance of berberine in *P. amurense*

To understand the distribution of BIAs in *P. amurense* and 17 other plant species in six different orders (Magnoliales, Laurales, Ranunculales, Proteales, Vitales, and Sapindales), liquid chromatography–mass spectrometry (LC-MS) analysis was completed to profile 19 BIA metabolites in various tissues (table S1). The results showed that these BIAs were mainly detected from plants in Magnoliids (Magnoliales and Laurales) and early-diverging eudicots. These data also uncovered that protoberberine- and aporphine-type BIAs, especially berberine [**11**, $[M + H]^+$ mass/charge ratio (*m*/*z*) 336.1268], phellodendrine (**14**, $[M + H]^+$ *m*/*z* 342.1715), columbamine (**16**, $[M + H]^+$ *m*/*z* 338.1410), jatrorrhizine (**17**, $[M + H]^+$ *m*/*z* 338.1410), palmatine (**19**, $[M + H]^+$ *m*/*z* 352.1560), and magnoflorine (**21**, $[M + H]^+$ *m*/*z* 342.1721), were highly accumulated in the roots or/and stems of Zanthoxyloideae species (Rutaceae and Sapindales), such as *P. amurense*, *Tetradium ruticarpum*, *Zanthoxylum armatum*, and *Zanthoxylum nitidum* (Fig. 1B and figs. S1 and S2). In contrast with these four species, BIAs were not detected in other Sapindales species and *Vitis vinifera*. In the order Ranunculales, species such as *Mahonia bealei*, *Berberis julianae*, *Nandina domestica*, and *Coptis chinensis* also accumulated high levels of protoberberine-type BIAs, similar to *P. amurense* (Fig. 1B). These findings confirm that protoberberine-type BIAs, especially for berberine, have an independent and lineage-specific distribution among species in distantly related Sapindales (*Phellodendron*) and Ranunculales (*Coptis*). The biosynthetic pathway of berberine in Ranunculales, especially in *Coptis*, has been fully elucidated (Fig. 1A). On the basis of the similarities of the BIA profiles in *Phellodendron* and *Coptis*, we proposed that the two plants in these two genera might share the similar steps of the berberine pathway. In addition, a matrix-assisted laser desorption/ionization–mass spectrometry imaging (MALDI-MSI) was completed to spatially localize the BIAs in the roots of *P. amurense*. The resulting images showed that the berberine pathway metabolites and berberine derivatives were localized in the periderm and phloem of roots (Fig. 1C and fig. S3).

Genome assembly and WGD of *P. amurense*

To identify the crucial genes responsible for the berberine biosynthesis in *P. amurense*, we generated 137.52 Gb of Illumina sequencing data, 73.14 Gb of PacBio Sequel II long-read sequencing data, and 324.14 Gb of Hi-C paired-end reads, from which a chromosome-level haplotype genome was assembled (Fig. 2A, fig. S4, and table S2). Genome survey, initial assembly, and chromosome anchoring analyses constructed 38 pseudo-chromosomes, which covered 2.81-Gb genome size with a scaffold N50 length of 76.63 Mb (fig. S5 and tables S3 and S4). This assembled genome reached a Benchmarking Universal Single-Copy Orthologs (BUSCO) completeness value of 97.7% (table S5). Furthermore, ~87.03% of the *P. amurense* genome consisted of transposable elements, in which the long terminal repeats retrotransposons (LTR-RTs) were the most abundant class of repetitive sequences (table S6). A total of 41,414 protein-coding genes were predicted in the *P. amurense* genome with a BUSCO completeness of 97.60%, indicating the high integrity of the genome assembly and gene prediction (table S5). The analysis of phylogenomics and dating revealed that two BIA-producing Zanthoxyloideae species (*P. amurense* and *Z. armatum*) were sister to the Aurantioideae and Rutoideae species that did not accumulate BIAs (Fig. 2B and fig. S6), and the divergence time between *P. amurense* and *Z. armatum* was 24.90 million years ago (Ma), with a 95% confidence interval (CI) of 18.33 to 31.53 Ma (fig. S6).

An intragenomic collinearity analysis revealed remnants of at least two whole-genome duplication (WGD) events in the *P. amurense* genome (Fig. 2A and figs. S7 and S8). An intergenomic collinearity indicated that four paralogous segments in the *P. amurense* and *Z. armatum* genomes corresponded to one orthologous region in the *Citrus clementina* genome (Fig. 2, C and D), suggesting that two round WGD events in the *P. amurense* and *Z. armatum* genomes might have occurred after the divergence with the *C. clementina* genome. The genomes of *P. amurense* and *Z. armatum* were, respectively, mapped into the *C. clementina* genome. This mapping phylogenetically divides them into four subgenomes (PamuA, PamuB, PamuC, and PamuD; and ZarmA, ZarmB, ZarmC, and ZarmD; Fig. 2E and fig. S9). To further investigate the phylogenetic placement of the *P. amurense*s WGDs, the K_s (synonymous substitutions per synonymous site) distribution of paralogs from pairwise subgenomes showed that two rate-adjusted K_s peaks: 0.16 for the PamuAC paralogs and 0.18 for the PamuBD paralogs. This result confirmed the shared two round WGDs before the split of *P. amurense* and *Z. armatum* (Fig. 2, F and G, and fig. S10). These results indicated that the widespread WGD events and complicated chromosome evolution might have occurred in Rutaceae species.

Elucidation of the berberine pathway in *Phellodendron*

To identify genes involved in berberine biosynthesis in *P. amurense*, our previously sequenced transcriptomes from 18 tissues and berberine accumulation patterns were integrated for Pearson correlation coefficient (*r*) analysis. The resulting data revealed top-ranked candidate genes, which correlated with the production of berberine ($r > 0.80$). Of these genes, included are two *PaOMTs*, one *PaNMT*, and four *PaCYPs* genes, which were selected for functional analysis (Fig. 3A).

First, to determine the *O*- and *N*-methylations involved in the biosynthesis of berberine, recombinant *PaOMT1*, *PaOMT2*, and *PaNMT1* were induced for in vitro test of their methylation activities with four substrates, (*S*)-norcoclaurine (**3**), (*S*)-coclaurine (**4**), (*S*)-3′-hydroxy-*N*-methylcoclaurine (**6**), and (*S*)-scoulerine (**8**). These tests determined that PaOMT1 and PaOMT2 had methylation activity, while PaNMT1 did not. PaOMT1 catalyzed successive *O*-methylation of C-2 and C-8 hydroxyl groups of **8** to form (*S*)-tetrahydrocolumbamine (**9**) (*m*/*z* 342) and (*S*)-tetrahydropalmatine (**26**) (*m*/*z* 356), which presents the same function as Cc9OMT from *C. chinensis* (Fig. 3B and fig. S11). Unexpectedly, PaOMT2 did not present any *O*-methylated function but, unexpectedly, exhibited *N*-methylation activity of **4** to produce (*S*)-*N*-methylcoclaurine (**5**, *m*/*z* 300) and (*S*)-magnocurarine (**12**, *m*/*z* 314) through single and double *N*-methylations, respectively, which is equivalent catalytic activity to CcNMT from *C. chinensis* (Fig. 3B and fig. S12). Furthermore, PaOMT2 also catalyzed the *N*-methylation and *N*-demethylation of unsaturated *NH* or *NCH3* groups on intermediates **3**, **4**, and **8**, exhibiting

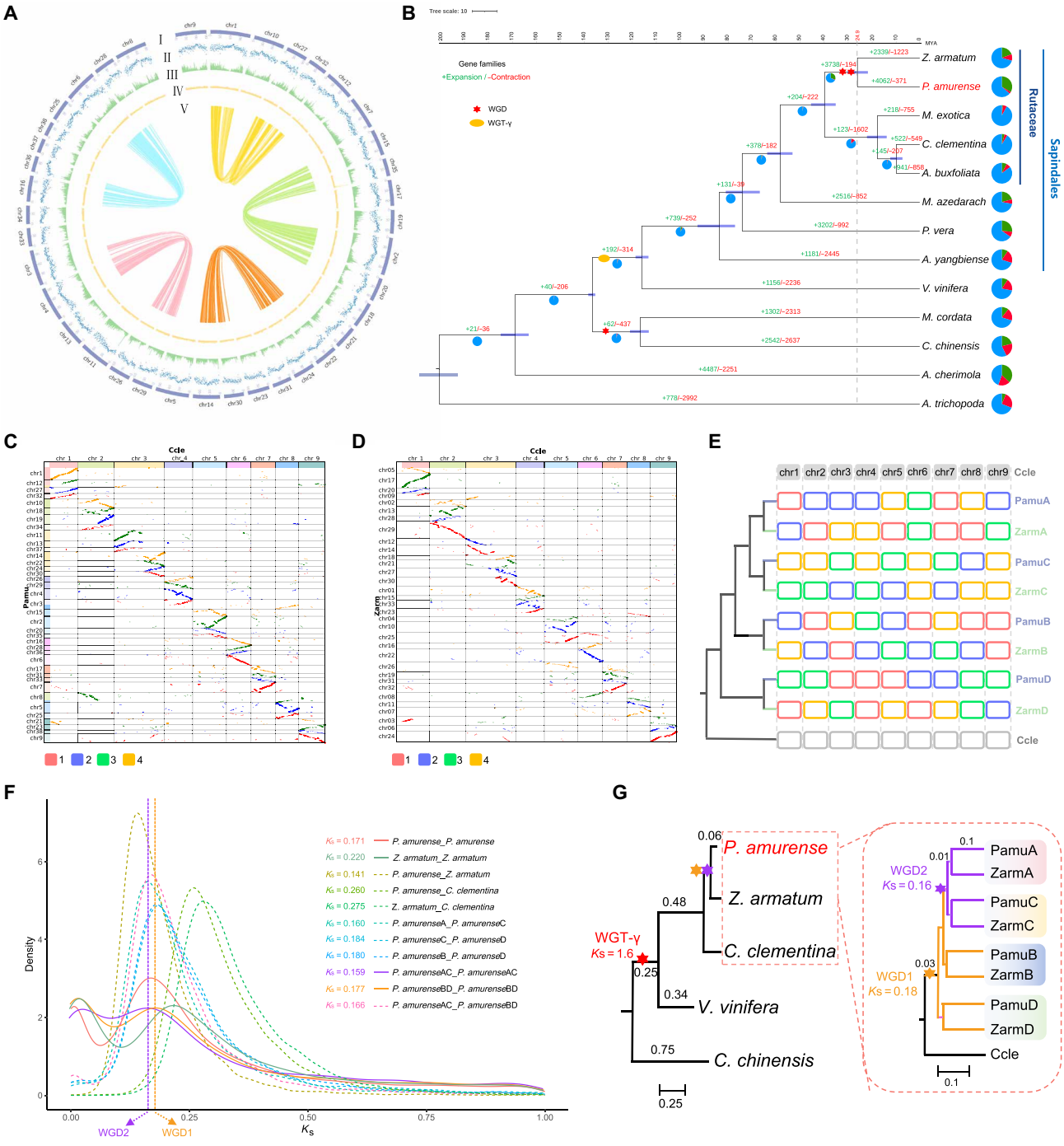

Fig. 2. Genomic features, phylogenetic tree, and WGD events of the *P. amurense* genome. (**A**) Genome characteristics of *P. amurense*. The circos plot from the outmost to the inmost circle visualizes 38 pseudo-chromosomes (I), gene density represented as number of genes per Mb (II), GC content in 1 Mb windows (III), density of repeat sequences (IV), and each linking line in the center of the circos plot showing a pair of homologous genes (V). (**B**) A phylogenetic tree developed with single-copy orthologs from *P. amurense* and 12 other candidate species. Green and red colored numerals highlighted at each branch line mean the number of the expanded and contracted gene families, respectively. Pie charts show the proportions of gene families that have undergone contraction (red) or expansion (green). Divergence timings with a 95% CI were highlighted with horizontal purple bars at the internodes. WGT-γ means the gamma triplication event. (**C**) Collinearity between the *P. amurense* and *C. clementina* genomes. The *C. clementina* genome was used as reference to identify synteny blocks and orthologous gene pairs for each chromosome in the genome of *P. amurense*. (**D**) Collinearity between the *Z. armatum* and *C. clementina* chromosomes. (**E**) Phylogenetic trees were, respectively, constructed using the homologous genes from different chromosomes of *P. amurense* subgenomes and *Z. armatum* subgenomes according to the collinearity mapping between candidate species and *C. clementina* chromosomes (Chr1 to Chr9). The red, blue, green, and orange boxes in *P. amurense* and *Z. armatum* represent four types of synteny blocks with *C. clementina*. (**F**) K_S distribution for orthologous and paralogous genes among *P. amurense*, *Z. armatum*, *P. amurense* subgenomes, *Z. armatum* subgenomes, and *C. clementina*. (**G**) Predicted WGD of *P. amurense* and *Z. armatum*. WGD1 and WGD2 represent the WGD events shared between the *P. amurense* and *Z. armatum* genomes.

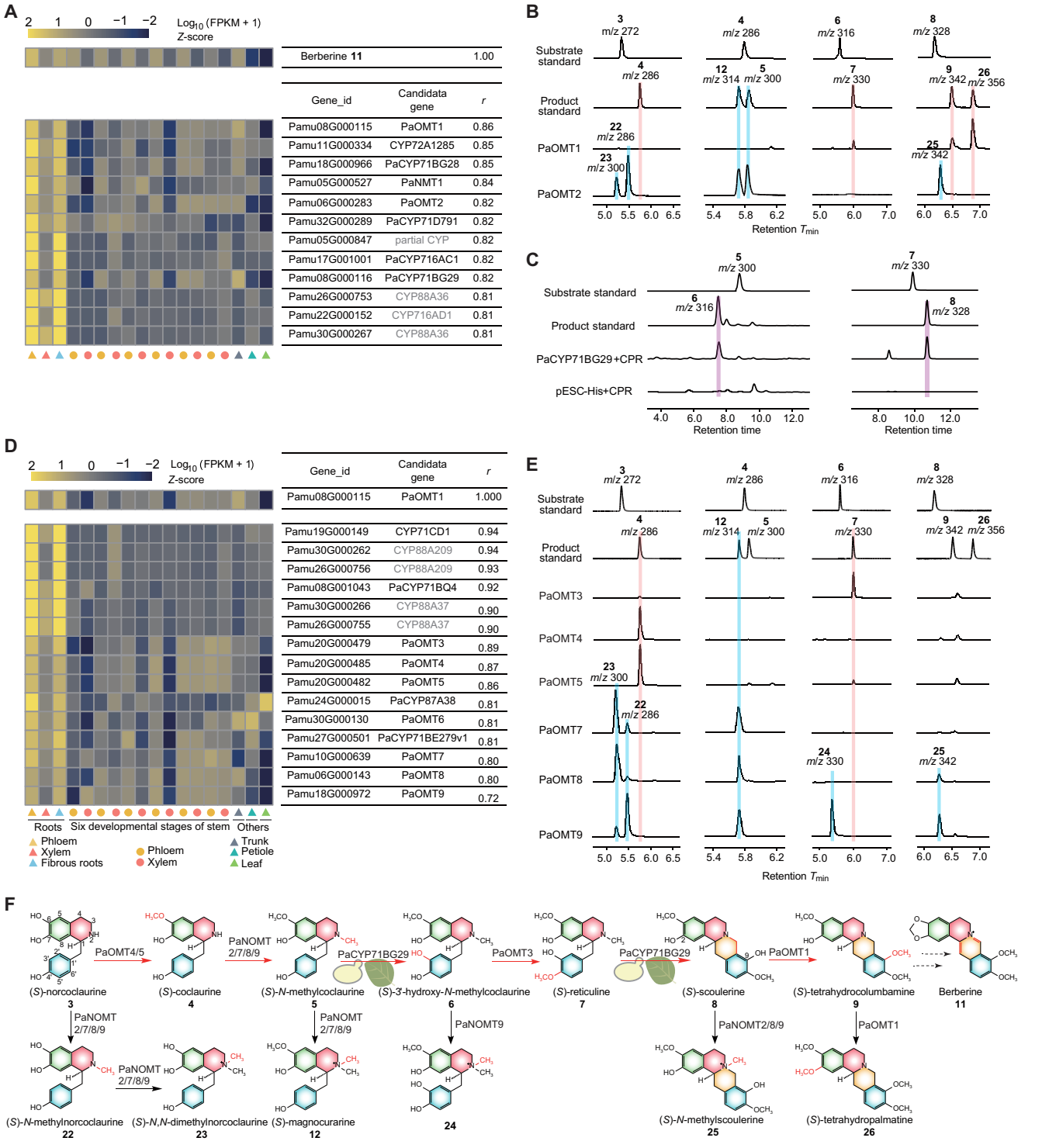

Fig. 3. Elucidation of the berberine biosynthetic pathway in *Phellodendron*. (**A**) Pearson correlation coefficient between the *P. amurense* transcriptome data and berberine content ($r > 0.80$). The heat map displays the *z*-score calculated from $\log_{10}$ (FPKM + 1). CYP450s, highlighted in gray, are unsuccessfully cloned, partial genes, or the homologous genes of reported limonin biosynthesis (fig. S21). FPKM, Fragments Per Kilobase of exon model per Million mapped fragments. (**B**) Extracted ion chromatograms of compounds catalyzed by purified PaOMT1 and PaOMT2 using **3**, **4**, **6**, and **8** as substrates. (**C**) Extracted ion chromatograms of compounds catalyzed by PaCYP71BG29 microsomes using **5** and **7** as substrates. (**D**) Pearson correlation coefficient between the *P. amurense* transcriptome data and *PaOMT1* gene. Gray-highlighted genes are unsuccessfully cloned or the homologous genes of reported limonin biosynthesis. (**E**) Extracted ion chromatograms of compounds catalyzed by purified PaOMT4/5/7/8/9 using **3**, **4**, **6**, and **8** as substrates. (**F**) This scheme summarizes these catalytic steps that have been verified with enzymatic data provided in this report. The catalytic activities of PaCYP71BG29 were doubled confirmed using yeast and tobacco expression systems. *N*-methylated OMT members were renamed as NOMTs. T_{min}, time in minutes.

obvious substrate promiscuity (Fig. 3F). These previously unidentified activities enhance understanding the *N*-methylation in the BIA biosynthesis.

Second, the oxidation of BIAs was tested with CYP proteins. On the basis of the reports that CYP450 enzymes CYP80B1/2 (*24*, *28*, *29*), CYP80G1 (*29*, *30*), and CYP719A1 (*24*, *31*) have been shown to catalyze C-3′ hydroxylation of **5**, C-C phenol coupling of **7**, and the methylenedioxy bridge formation of (*S*)-tetrahydrocolumbamine (**9**) in Ranunculales, respectively, four candidate PaCYPs—*PaCYP71BG28*, *PaCYP71BG29*, *PaCYP71D791*, and *PaCYP716AC1*—were investigated in vitro for their catalytical function using **5**, **7**, and **9** as substrates. The results demonstrated that catalytic activities of PaCYP71BG29 to **5** and **7**, while that of others was not detected (fig. S13). PaCYP71BG29 exhibited hydroxylase activity at the C-3′ position of substrate **5** to product **6** (*m/z* 316) (Fig. 3C and fig. S14). In addition, PaCYP71BG29 catalyzed the berberine bridge formation of **7** to produce **8** (*m/z* 328) (Fig. 3C and fig. S15), which is the same as the BBE activity of FAD-dependent oxidase in *Coptis* spp. In addition, the transient expression of PaCYP71BG29 in *Nicotiana benthamiana* also confirmed its activity of hydroxylation toward **5** (fig. S16) and berberine bridge formation toward **7** (fig. S17).

Last, a coexpression analysis using *PaOMT1* as a bait gene was completed to elucidate the remanent biosynthetic steps of berberine in *Phellodendron* and selected seven OMTs (*PaOMT3* to *PaOMT9*) and three CYPs (*PaCYP71BQ4*, *PaCYP71BE279v1*, and *PaCYP87A38*) based on their highly similar expression to that of PaOMT1 (Fig. 3D). The in vitro biochemical assays showed three types of methylation activities. (i) As described above in PaOMT2 tests, PaOMT7, PaOMT8, and PaOMT9 catalyzed the *N*-methylated activity and had a broad substrate promiscuity (Fig. 3E and fig. S12). (ii) PaOMT4 and PaOMT5 methylated the C-6 hydroxyl group of **3** to form **4** (Fig. 3E and fig. S18). (iii) PaOMT3 methylated the C-4′ hydroxyl group of **6** to produce **7** (Fig. 3E and fig. S19). In addition, no activities were detected for from catalytic assays of PaOMT6 and three PaCYP450s with different substrates (figs. S13 and S20). In summary, these biochemical data elucidate 6-enzymatic steps of the berberine pathway in *Phellodendron* (Fig. 3F).

Crystal structure and catalytic mechanism of PaNOMT9

According to the previously unidentified catalytic characterization, the *N*-methylated OMT members have been renamed as the NOMT group (Fig. 3F). To understand the mechanism of *N*-methylation catalyzed by OMTs in *P. amurense*, we solved the crystal structure of the ternary complex PaNOMT9/*S*-adenosylmethionine (SAM)/**4** [Protein Data Bank (PDB) ID: 8ZN0] at a resolution of 1.8 Å (Fig. 4A). The protein structure of PaNOMT9 formed a dimer (labeled A and B). Each of these two distinct monomers constitutes an N-terminal dimerization domain (1 to 165 amino acids mainly consist of α1 to α10, β1 to β3, and η1) and a SAM/*S*-adenosylhomocysteine (SAH)–binding domain (199 to 359 amino acids, α12 to α16, β4 to β10, and η2) in the C-terminal Rossmann fold, two of which are connected by the α11 helices (166 to 198 amino acids) (Fig. 4A). A total of 17 residues within 5 Å around substrate **4** were localized at the interface among the dimerization domain (α6 and α10), the SAM/SAH-binding domain (α15 and η2), and helix α11 of PaNOMT9 (figs. S22A and S23). The protein-ligand interaction analysis revealed that H127 formed hydrogen bonds with the C6-methoxy and C7-hydroxyl groups, while Y159 formed hydrogen bonds with the NH group (Fig. 4B). The core substrate **4** was surrounded by the hydrophobic residues F123, F126, W262, I265, M312, I315, G316, V320, and L325 and interacted with these residues. In addition, three residues distant from the active site—D185, D293, and L295—were identified through sequence alignment as differing between *N*-methylated and *O*-methylated OMT proteins (Fig. 4C and figs. S22B and S23).

The abovementioned 20 residues of PaNOMT9 were substituted with alanine (A) one by one to generate 20 variants, and then the activity of each variant was tested with **4** as substrate. The contents of methylated (**5**) or dimethylated (**12**) products from the reactions were estimated with LC-MS analysis (Fig. 4D). On the one hand, the results showed that the replacement of each M312, I315, and L319 with alanine (A) almost abolished the *N*-methylation activity (Fig. 4D). On the other hand, four variants—F126A, M176A, I265A, and L325A—exhibited obviously higher **5**/**12** (methylation/dimethylation) ratios than the wild-type one did (Fig. 4D). These results indicate that the hydrophobic interactions are essential for the construction of substrate binding pocket (Fig. 4E). In addition, the D293A and L295 mutants, which are positioned far from the active site, exhibited a nearly complete loss of *N*-methylation activity, likely attributable to structural perturbations within the protein. The H127A variant of PaNOMT9 could only catalyze an ineffective *N*-methylation of **4**, suggesting that the hydrogen bonds between H127 and substrate **4** were critical for the formation of substrate-binding conformation (Fig. 4, D and E). We further generated 11 additional H127 mutants, and the results demonstrated that the conservative substitution (H127N) and substitutions with similar properties (H127K and H127R) retained wild-type *N*-methylation activity toward **4**. In contrast, the H127P mutant exhibited a near-complete loss of *N*-methylation activity (fig. S24A). The H127A and H127P mutants, which are positioned farther from the C6-methoxy and C7-hydroxyl groups, resulted in a larger cavity within the active pocket, indicating that H127 plays a crucial role in the substrate-binding conformation (fig. S24B).

A comparison of crystal structure was completed to understand the mechanism of *N*-methylation by PaNOMT9. Tf6OMT (PDB ID: 5ICE) from *Thalictrum flavum* and CjCNMT (PDB ID: 6GKV) from *Coptis japonica* were reported to catalyze *O*-methylation and *N*-methylation, respectively, and their crystal structures were resolved in details (*32*, *33*). On the basis of these, we compared the protein structures of Tf6OMT and CjCNMT with PaNOMT9 (fig. S22). On the one hand, their overall structures between Tf6OMT and PaNOMT9 are similar, and their SAM and SAH positions are relatively close; however, both structures have distinct protein folds (fig. S22C). The structures of Tf6OMT/norlaudanosoline/SAH and PaNOMT9/**4**/SAM complexes were overlaid to compare the active pockets and substrate-binding conformations (Fig. 4F and fig. S22C). The conformation comparison revealed that the orientations of substrates in the active site pockets of Tf6OMT and PaNOMT9 were different (Fig. 4F). The NH group of the substrate **4** was positioned relatively closer to the SAM donor (4.5 Å) than other bound ligands in the ternary complex PaNOMT9/**4**/SAM, while the C-6 OH group was positioned closer to the SAM donor (3.4 Å) in the Tf6OMT complex. In addition, the docking simulation with PaNOMT9 and substrates **3**, **6**, and **8** exhibited a similar conformation to **4** (fig. S25). On the other hand, compared with the protein structure of CjCNMT (PDB ID: 6GKV), the nonhomologous PaNOMT9 presents a completely different overall structure, including the conformation of SAM and substrate binding regions, hydrogen bonds,

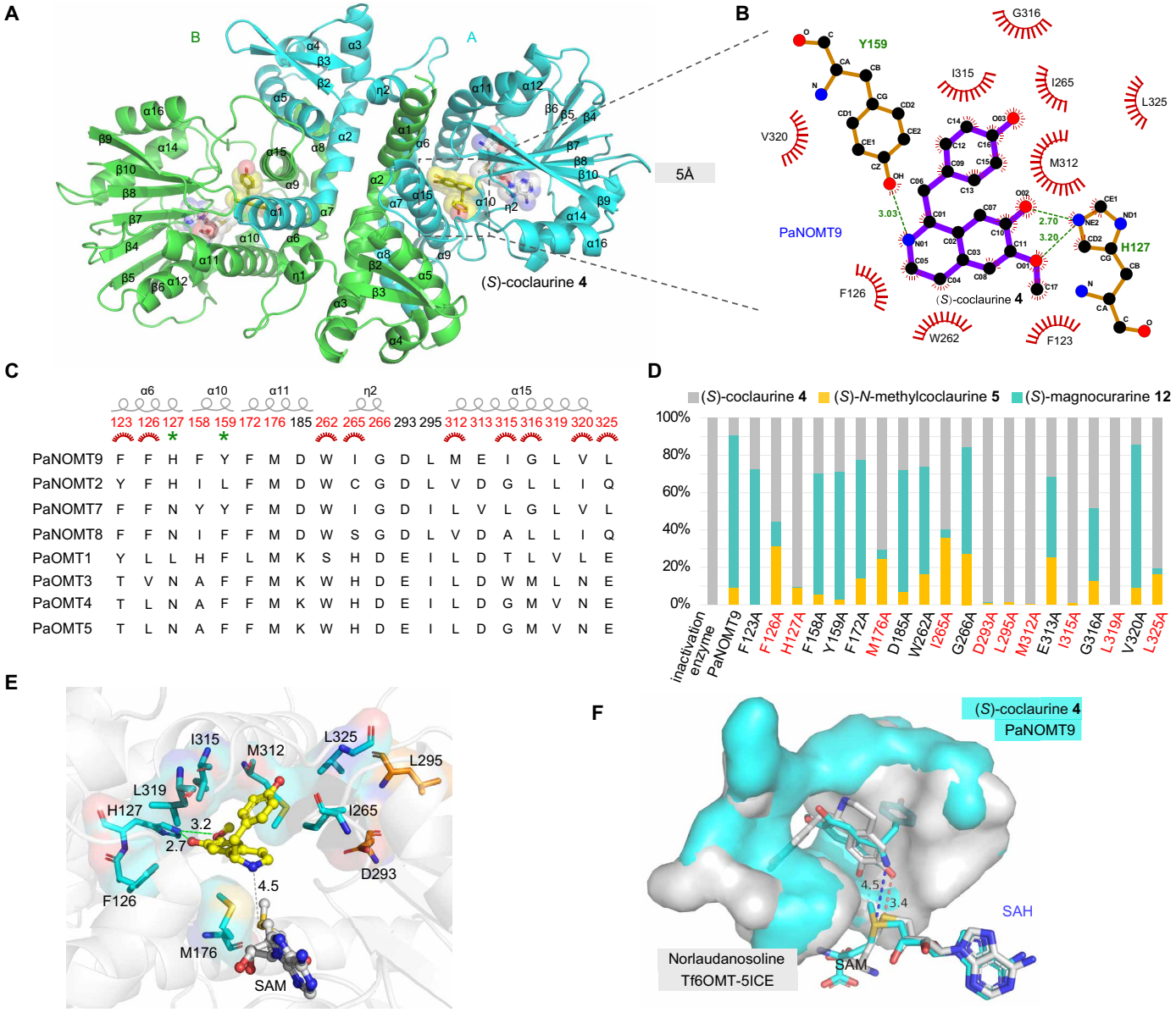

Fig. 4. Crystal structure and catalytic mechanism of PaNOMT9. (**A**) An overall view of the crystal structure of PaNOMT9/**4**/SAM. (**B**) Protein-ligand interaction diagram between PaNOMT9 and the ligand **4**. Red marks signify the hydrophobic residues. (**C**) The sequence alignment of amino acids among *N*-methylation active OMT proteins PaNOMT9/2/7/8 and *O*-methylation OMT proteins PaOMT1/3/4/5. (**D**) Catalytic activities of PaNOMT9 and 20 variants resulted from targeted mutations are shown with stacked bars. (**E**) Crystal structure of ternary complex formed by PaNOMT9/**4**/SAM, in which the mutations of those crucial residues reduce the catalytic activities. The residues highlighted with cyan color indicate the proximity to the substrate within 5 Å, while the residues colored with orange are outside this range. The dotted gray line means the distance between the *N* atom of **4** and the methylation group of SAM is 4.5 Å. (**F**) Overlay of the crystal structures of two ternary complexes: Tf6OMT/norlaudanosoline/SAH and PaNOMT9/**4**/SAM. The protein-ligand binding conformation differences of these two ternary complexes lead to the function divergence of *O*-/*N*-methylation.

and hydrophobic residues, although both proteins exhibit similar *N*-methylated activity. These results illustrate that the conformational flexibility and plasticity of PaNOMT9 at and around these regions enables the *N*-methylation of substrates.

Neofunctionalization and catalytic mechanism of PaCYP71BG29 in the berberine bridge formation

Homologous mining, phylogenetic analysis, and functional assays were carried out to trace the origination and evolution of berberine bridge formation catalyzed by PaCYP71BG29. The phylogenetic analysis showed that *PaCYP71BG29* and five other genes from *P. amurense*, nine genes in *Z. armatum*, and two genes in *C. clementina* clustered in the same clade as *OsCYP71P1* from *Oryza sativa* (with 46 to 58% sequence identity) (Fig. 5A, fig. S26, and tables S7 and S8). A further comparison with amino acid sequences showed that PaCYP71BG29 exhibits 47% identity with OsCYP71P1 [a tryptamine 5-hydroxylase (T5H)] (*34*).

The phylogenetic tree of CYP71BG subfamily members was divided into seven nodes (node 1 to node 7). On the basis of the phylogenetic tree, nine *CYP71BG* genes, and their two ancestral

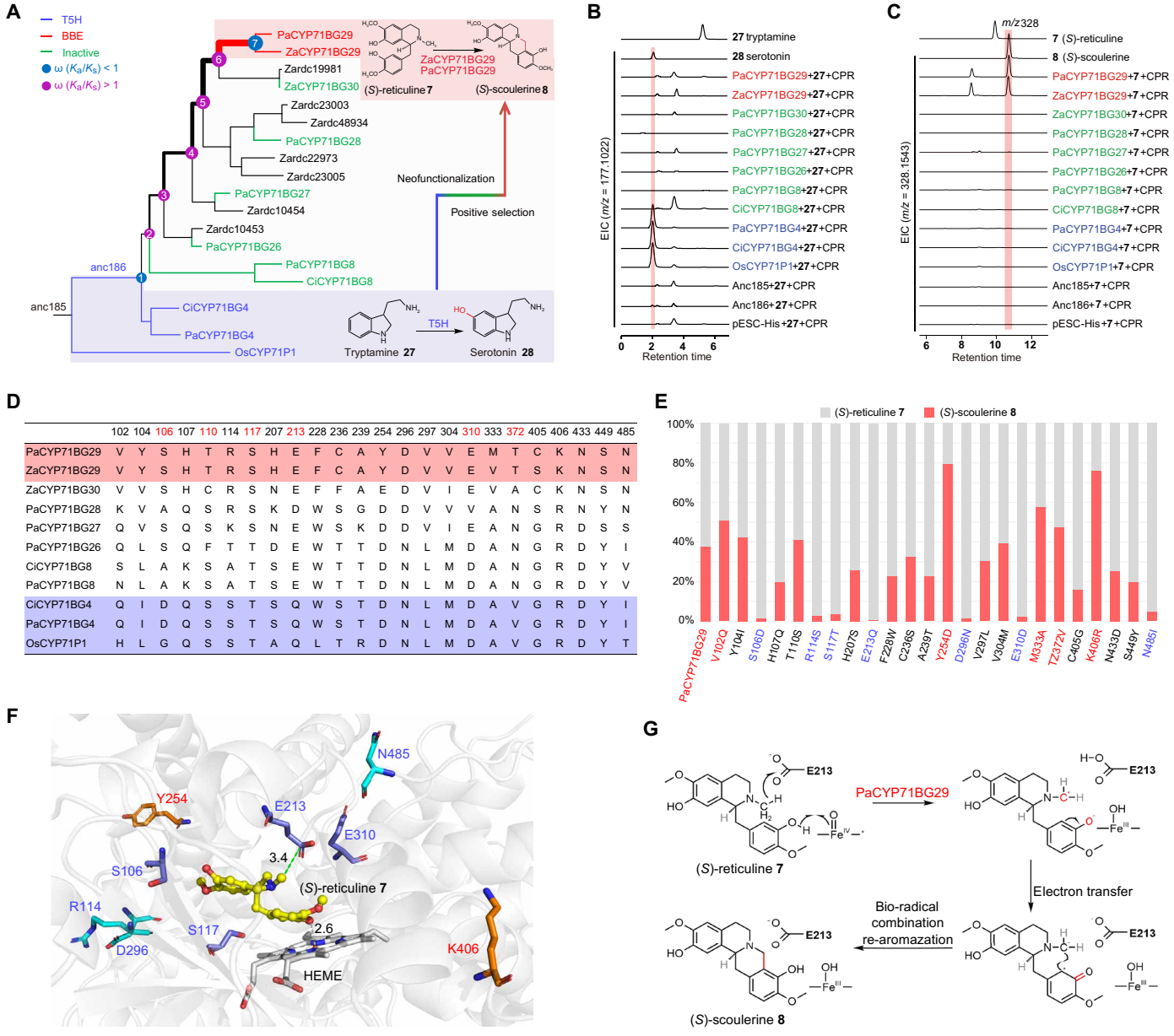

Fig. 5. Determination of the neofunctionalization of PaCYP71BG29 and its catalytic mechanism. (**A**) A maximum-likelihood phylogenetic tree was built with the PaCYP71BG subfamily members from *P. amurense*, *Z. armatum*, and *C. clementina*. OsCYP71P1 was chosen as the outgroup. K_a/K_s, nonsynonymous (K_a) with synonymous (K_s) substitution rates. (**B** and **C**) Extracted ion chromatograms (EIC) characterized compounds formed in the enzymatic reactions consisting of candidate CYP microsomes and two substrates-tryptamine **27** and (*S*)-reticuline **7**. (**D**) Sequence alignment of PaCYP71BG29, ZaCYP71BG29, and other CYP71BG subfamily members. Twenty-four amino acids were presented, and six variation sites colored with red are within 5 Å from **7**. (**E**) Stacked bars show catalytic preference of PaCYP71BG29 and its 24 variants toward two substrate: (*S*)-reticuline **7** and (*S*)-scoulerine **8**. (**F**) Molecular docking model of PaCYP71BG29/HEME/**7** indicates those crucial residues, which enhance (red) and reduce (blue) the catalytic activities. The distances among between substrate **7** and crucial residue E213 and HEME are 3.4 and 2.6 Å. (**G**) A mechanism is proposed for the formation of the berberine bridge catalyzed by PaCYP71BG29.

genes (*anc185* and *anc186*) were selected to induce recombinant proteins for activity tests (Fig. 5, B and C). First, according to the catalytic activity of OsCYP71P1, tryptamine (**27**) was used as the substrate for in vitro enzymatic assays, and OsCYP71P1 was chosen as positive control. The results showed that PaCYP71BG4, CiCYP-71BG4, and OsCYP71P1 converted tryptamine **27** to serotonin **28** (Fig. 5B and fig. S27). In addition, the assays showed that the ancestor CYP71BG genes, anc186, also encoded a T5H, although its catalytic activity was moderate. By contrast, the CYP719BG candidate genes in node 2, which resulted from gene expansion in the *P. amurense* and *Z. armatum* genomes, showed no hydroxylation activity toward tryptamine **27** (Fig. 5, A and B). Second, reticuline **7** was selected as a substrate to test berberine bridge activity. The catalytic assays showed that PaCYP71BG29 and ZaCYP71BG29 in node 7 exhibited a berberine bridge activity, converting (*S*)-reticuline **7** to (*S*)-scoulerine **8** (Fig. 5, A and C). By contrast, other duplicated CYP71BG genes in

P. amurense and *Z. armatum* could not catalyze the conversion of **7** to **8**. These results indicate that neofunctionalization and divergence have occurred in *PaCYP71BG29* and *ZaCYP71BG29* during plant evolution. Last, the two-ratio branch model was used to detect the positive selection. The results show that the branches from node 2 to node 6 had ω values greater than 1. These values also suggest that the neofunctionalization has occurred in *PaCYP71BG29* and *ZaCYP-71BG29*. The acquired catalysis is responsible for the formation of the berberine bridge in *P. amurense* and *Z. armatum* (Fig. 5A).

A docking simulation via AlphaFold2 was completed to understand the mechanism of the cyclase activity of PaCYP71BG29 in the berberine biosynthesis, using (*S*)-reticuline **7** as the ligand (fig. S28). Residues substantially differing among active and inactive CYP71BG enzymes were further selected via sequence alignment (Fig. 5D). A total of 24 various residues were selected using alignment; among them, six residues, including S106, T110, S117, E203, E310, and T372, localized around **7** within 5 Å (Fig. 5D and fig. S29). These residues of PaCYP71BG29 were subjected to a series of substitutions toward the T5H function, and the activities of mutagenesis using **7** as substrate were examined (with three replicates; Fig. 5E). Several mutants enhanced the berberine bridge formation activity; for instance, the mutants Y254D and K406R of PaCYP71BG29 exhibited obviously effective conversion rates, 79.4 and 75.9%, respectively, higher than that of the wild type, 37.8%. Seven mutants of PaCYP71BG29, including S106D, R114S, S117T, E213Q, D296N, E310D, and N485I, almost lost the berberine bridge formation activity, suggesting their crucial contribution in the converting **7** to **8** (Fig. 5E). The following variants, S106D, S117T, E213Q, and E310D, located within a 5-Å range, can potentially affect substrate binding. On the other hand, variants R114S, D296N, and N485I, which are situated away from the active center, may affect the catalytic activity of PaCYP71BG29 by altering the protein conformation and surface structure (Fig. 5F). On the basis of these data, we proposed that the alternations of these residues lead to neofunctionalization of CYP71BG29.

The docking model of PaCYP71BG29/HEME/**7** showed that the carboxyl group of the crucial residue E213 was close to the *N*-methyl group of **7** (3.4 Å), and the heme iron center was near the C3′-OH group of **7** (2.6 Å) (Fig. 5F). The stepwise catalytic mechanism of PaCYP71BG29 in the berberine bridge formation was further proposed by referring to the reported BBE (*35*): Heme cofactor deprotonates the C3′-OH to increase the nucleophilic behavior, a hydride from the *N*-methyl group is transferred to the E213, and the rearomatization of intermediate leads to the cyclization reaction of **8** (Fig. 5G). Therefore, whereas all the Ranunculales plants examined to date use FAD-dependent oxidases to catalyze the berberine bridging reaction, members of the Rutaceae family (*P. amurense* and *Z. armatum*) might convergently evolve a distinct CYP450 member to mediate the identical oxidation reaction through a similar chemical mechanism.

DISCUSSION

Although the biosynthesis of berberine has been elucidated in *Coptis* spp., whether the same biosynthetic pathway occurs in other lineage plants remains unclear. The answers to this question help understand not only the berberine biosynthesis but also diverse metabolic pathways in different plants. The reported berberine biosynthesis in *Coptis* included one NCS, three OMTs (*6OMT*, *4′OMT*, and *9OMT*), one NMT (*CNMT*), two CYPs (*CYP80B* and *CYP719A*), one BBE, and one BBE like (*THBO*). The annotation of orthologous genes in the genome of *P. amurense* revealed an extremely low identity of NCS (from 33 to 37%), OMTs (from 26 to 49%), NMT (from 33 to 37%), CYPs (from 23 to 42%), and BBE genes (from 29 to 57%) compared to those found in *Coptis* (table S9). Furthermore, no CYP80 and CYP719 genes were identified in *P. amurense*. Therefore, these features suggested that the biosynthetic pathway of berberine might be different between *P. amurense* and *Coptis* spp. In addition, on the basis of these features, we hypothesized that unknown mechanisms behind methylation and oxidation reactions might be involved in the biosynthesis of berberine in *P. amurense*. This hypothesis was approved by the characterization of *N*-methylation of OMTs and neofunctinalizations of CYP enzymes discussed below.

The catalytic activities of PaNOMT9 reported here disclose a dual methylation mechanism involved in the biosynthesis of berberine alkaloids. SAM-dependent MTs play an essential role in the diversification of plant natural products. The past investigations have characterized MT members that transfer methyl groups to *S*, *N*, *O*, or *C* atoms of compounds to produce methylated natural products (*36*). On the basis of *S*, *N*, *O*, or *C* modifications, MTs have been classified to be SMT, NMT, OMT, and CMT. To date, the MT specificity in BIA biosynthesis has shown that OMT and NMT members strictly transfer methyl groups to C-OH and NH groups, respectively (*37*). In contrast with these, a few members, such as those in the class II and class III members of the OMT family, have been reported to have both *O* and *N*-methylation actives in xanthosine and anthranilate (table S10) (*38*, *39*). Here, our data added additional subclass of MTs, NOMTs. *PaNOMT2/7/8/9* from *Phellodendron*, which presented similar catalytic activities with the *Coptis* CNMT gene, are the first OMT members able to catalyze *N*-methylation in BIA biosynthesis. The berberine biosynthetic NOMT genes (*PaNOMT2/7/8/9*) from *P. amurense* are distantly sister to the class II subfamily of the OMT family (fig. S30A). The phylogenetic branch of the NOMT genes includes abundant tandem duplicated genes from *P. amurense* and *Z. armatum*, and only one homologous gene from *Citrus* (*CiOMT2*) was observed, showing 64 to 66% sequence identity compared to *PaNOMT2/7/8/9* (figs. S30B and S31). Our catalytic assays showed that *CiOMT2* could not perform the *N*-methylation activity on substrates **4** (fig. S30B), indicating that the gain of *N*-methylation function evolved after their split with *Citrus*. The catalytic mechanism of *N*-methylated PaNOMT9 differs from that of reported BIA OMT or NMT enzymes, providing valuable knowledge to enhance understanding structural methylation of BIAs and promote metabolic engineering for potent BIA drugs.

The berberine bridge formation activity of PaCYP71BG29 reported here reveals a previously unidentified case of identical catalytic activity from nonhomologous enzymes in plant specialized metabolite biosynthesis. In previous studies, BBEs from the superfamily of FAD-dependent oxidase in Ranunculales, such as *Coptis*, *Papaver*, *Thalictrum*, and *Berberis*, have a conserved contribution in berberine bridge formation in BIA biosynthesis (*21*, *40*–*42*). However, only one PaBBE gene from *P. amurense* genome clustered into the same branch with BBEs from Ranunculales plants. In addition, gene expression analysis revealed that *PaBBE* was silent in all tested tissues of *P. amurense* (fig. S32). This datum suggests that unknown mechanisms might exist in this medicinal plant. To test this, we mined both the genomic and transcriptomic sequences and

unveiled CYP71BG members that were transcriptionally associated with the accumulation patterns of berberine. A comparative genomic analysis unearthed that, on the one hand, none of the CYP71BG members was annotated in the Ranunculales genomes; on the other hand, these genes resulted from the genome expansion. These data provide evidence that the biosynthetic steps of berberine catalyzed by *CYP71BG29* and the berberine bridge formation result from an independent evolution in *P. amurense*. In addition, our biochemical assays revealed that the CYP71BG enzymes are involved in the biosynthesis of serotonin and melatonin, two plant amines associated with plant development and stress response. The phylogenetic analyses indicate that the obvious expansion and positive selection of CYP71BG members led to the neofunctionalization of *CYP71BG29* in *Phellodendron*.

9OMT (*PaOMT1*) and BBE (*CYP71BG29*) form a biosynthetic gene cluster in Chr8 of *Phellodendron*, and this gene cluster further underwent the WGD1 and WGD2 events, which expanded the *N*-methylated *PaNOMT7* and *PaNOMT9* (fig. S33), different from the dispersed distribution of BIA biosynthetic genes in *Coptis*. In addition, 4′OMT (*PaOMT3*) and 6OMT (*PaOMT4* and *PaOMT5*) cluster together in Chr20, and the WGD events also accelerate the expansion of collinearity OMTs. We also discovered that the berberine biosynthetic genes are conserved in *Phellodendron* and *Zanthoxylum*, and the duplication and neofunctionalization of CYP450 and OMT genes occurred after their split with *Citrus* (fig. S33). The berberine biosynthetic gene cluster and duplication events expanded our understanding the formation and independent evolution of metabolic gene cluster (*43*).

Plant specialized metabolites have been viewed as complex traits that can be studied to understand their evolutionary trajectories (*44*). Identical metabolic traits often emerge in distantly related species as independent evolution, including parallel evolution and convergent evolution. For instance, the parallel evolution of cannabinoid biosynthesis catalyzed by homologous enzymes in *Helichrysum* and *Cannabis* (*45*). Although multiple studies have reported the convergent evolution of caffeine (*39*), crocins (*46*), and momilactone (*47*) biosynthesis in land plants, their biosynthetic enzymes in disparate lineages are likely as homologous relationships. Referring to the distinction between parallel evolution and convergent evolution (*48*), our study presents a classic case of convergent evolution in the production of berberine in two distant lineages, *Coptis* and *Phellodendron*, deriving from completely nonhomologous enzymes and various protein folds, especially for the OMT gene with NMT catalytic function and the CYP450 with berberine bridge enzymatic activity. These data indicate that the neofunctionalization of enzymes form a drive force in promoting structural modification and diversification.

In summary, we report a high-quality genome of *Phellodendron* and deciphered an alternative berberine biosynthetic pathway, particularly identifying the *N*-methylated OMT gene and the berberine bridge activity catalyzed by CYP450. The identified biosynthetic steps of berberine in *Phellodendron* results from the WGD events and neofunctionalization. These findings reveal a convergent evolutionary model of berberine biosynthesis in distantly related *Coptis* and *Phellodendron* spp. Moreover, our study enhanced our understanding of the berberine biosynthetic enzymes, allowing us to engineer complex BIAs with diverse pharmacological activities.

MATERIALS AND METHODS

Plant materials and chemical standards

P. amurense plants grow in the research station at the Northeast Forestry University in Harbin, Heilongjiang Province, China. Fresh leaves of *P. amurense* were collected in liquid nitrogen and then stored in −80°C freezer for genome sequencing. Fresh roots of *P. amurense* were collected and then divided into two packages. One was frozen in liquid nitrogen for metabolite extraction and the other was used for MSI (MALDI-MSI) analysis. The roots and/or stems of 16 other plants and dried lotus plumule of *N. nucifera* were collected for metabolic profiling (table S1). All chemical standards listed in table S11 were purchased different companies.

LC-MS analysis of metabolites

All samples (with three replicates) were oven-dried at 50°C and then ground intro fine powder, which was filtered with a 60-mesh sieve. Fine powder (200 mg) was weighed to a 15-ml glass tube. The tube was added 10 ml of a methanol:water mixture (6:4, v/v), followed by vortexing for 1 min. It was left at room temperature overnight and then sonicated for 40 min in an ultrasonic cleaner (240 W and 40 Hz). The tube was centrifuged at 9000*g* for 10 min. The supernatant containing metabolites was collected to a clean tube for metabolite analysis. Then, the supernatant was centrifuged at 12,000*g* for 10 min. The resulting clean supernatant was transferred to a 1.5-ml tube and then filtered through a 0.22-μm membrane.

The samples were detected by SHIMADZU UPLC-QTOF-MS equipped (LCMS-9030) with a SHIMADZU GIST C18 column (2 μm, 2.1 mm by 100 mm). The mobile phase consisted of 0.1% formic acid in water (A) and acetonitrile (B), and the linear gradient elution program was as follows: 0 to 1 min, 15% B; 1 to 8 min, 15 to 45% B; 8 to 10 min, 45 to 80% B; 10 to 11 min, 80 to 15% B; and 11 to 14 min, 15% B. The flow rate was 0.2 ml/min, and the column temperature was 40°C. The mass spectrometer used was operated in data-dependent acquisition mode with the electrospray ionization (ESI) source operating in positive ion mode (*m*/*z* 100 to 700). The gas temperature was maintained at 400°C, the gas flow rate was 10 liters/min, the fragment was set to 120 V, and the collision energy was 20 to 40 eV.

MALDI imaging

Frozen roots of *P. amurense* were encased in a 5% carboxymethyl cellulose (w/v) and then solidified into blocks using dry ice. Frozen roots were sectioned to 30 μm in thickness in −20°C. The sections were immediately affixed to glass slides coated with indium tin oxide for subsequent imaging analyses. Meanwhile, 60 mg of 2,5-dihydroxybenzoic acid (DHB) was a dissolved in 2 ml of methanol and water (with 0.1% trifluoroacetic acid) in a 7:3 ratio (v/v) (*49*). The resulting DHB solutionwas used for positive mode MALDI experiments. All measurements were performed using an AP-SMALDI ion source ("AP SMALDI10," TransMIT GmbH, Giessen, Germany) coupled to an Orbitrap mass spectrometer ("Q Exactive," Thermo Fisher Scientific, Bremen, Germany). In brief, a solid-state laser ($\lambda = 343$ nm) operating at a repetition rate of 2000 Hz was used to analyze the samples in positive ion mode. A matrix of DHB was used to increase ionization efficiency. The measurement rate in full scan mode (scan range *m*/*z* 150 to 1400) was ~1.5 s/pixel at a mass resolution of 70,000 mass, and the resolution corresponds with *m*/*z* 200 within 2–part-per-million mass precision. MSI analyses were conducted with a spatial resolution of 30 μm. MirionV3 software was used to perform data analysis.

Genome sequencing, survey, and assembly

High-quality genomic DNA samples were extracted frozen leaves using the DNeasy Plant Mini Kit (QIAGEN, Germany). The quality and quantity of DNA samples were evaluated and measured on a NanoDrop spectrophotometer (Thermo Fisher Scientific, USA). DNA samples were digested into 20-kb sizes. The DNA fragments were used to construct DNA libraries using SMRTbell template Prep Kits. Briefly, according the manufacturer's protocol, the steps of the construction of DNA libraries included DNA concentration, damage repair, end repair, ligation of hairpin adapters, and template purification. The circular consensus sequencing data for genomic sequencing were produced on the PacBio Sequel II platform (Pacific Biosciences, USA) according to the manufacture's protocol. The resulting DNA sequences were used for genome assembly described below. All comparisons included three independent biological replicates.

The total RNA samples were extracted from frozen leaves, stems, and roots using the RNAprep Pure Plant Kit (TIANGEN, China). The quality of RNA samples was then evaluated with a Qubit RNA Assay Kit on an Agilent 2100 Bioanalyzer (Agilent Technologies). RNA sequencing was performed to generate 150–base pair (bp) paired-end reads on an Illumina HiSeq 2500 platform at Novogene. Hi-C libraries were constructed from the fresh leaves of *P. amurense* using the ProximoTM Hi-C Plant kit. The cDNA libraries were sequenced to generate a 150-bp-long reads on an Illumina HiSeq platform. Illumina data were used to estimate the genome size of *P. amurense* according to the *K*-mer distribution analysis ($K = 21$). PacBio Sequel reads were filtered to remove short and low-quality reads. The resulting high-quality reads were used for de novo assembly using Hifisam (v0.19.5) (*50*) with the default parameters. Purge_dups (v1.2.5) (*51*) was used to dehybridize the assembled contig sequences. ALLHiC (v0.9.8) (*52*) was used to generate an allele-aware chromosome-level genome based on Hi-C data.

Gene annotations and quality assessment

Gene annotation was accomplished with de novo prediction, homologous protein evidence, and EST evidence approaches. After the assembly, the data of RNA sequencing (RNA-seq) were mapped to the genome with HISAT2 (v2.2.1) (*53*). For ab initio prediction, the AUGUSTUS model was trained by BRAKER2 (v2.1.5) (*54*) using the RNA-seq data. Furthermore, the ab initio prediction integrating the homologous evidence was combined to annotate the genome using the MAKER pipeline to synthesize the final gene models. The integrity of the genome was evaluated using the BUSCO (v5.2.2) (*55*) with the embryophyta_odb10 database. After annotation, tandem repeats were predicted using Tandem Repeat Finder (v4.07) (*56*). LTR_retriever (v2.6) (*57*), LTR_FINDER (v1.0.6) (*57*), and RepeatModeler (v2.0) were used to construct the repeats sequence database. For de novo prediction, the LTRs and consensus repetitive libraries generated by RepeatModeler were combined and then used as the input data for RepeatMasker.

Comparative genomic analysis

The orthologous groups of *P. amurense* and 14 other sequenced plant species, including *Z. armatum* (*58*, *59*), *Murraya exotica* (*60*), *C. clementina* (*61*), *Atalantia buxfoliata* (*62*), *Melia azedarach* (*63*), *Pistacia vera* (*64*), *Acer yangbiense* (*65*), *V. vinifera* (*66*), *Macleaya cordata* (*67*), *C. chinensis* (*15*), *Annona cherimola* (*68*), and one outgroup species [*Amborella trichopoda*; (*69*)], were used for phylogenetic analysis. Single-copy families among the 15 species were obtained using OrthoFinder (v2.3.11) (*70*) with default parameters. A phylogenetic tree concatenating the 221 single-copy orthologs was constructed using the maximum-likelihood method in RAxML (v8.0.17) (*70*) with 1000 bootstrap replicates. Divergence time with approximate likelihood calculation among the 15 species was estimated using MCMCTREE in PAML (v4.9) (*71*). The calibration points for the *P. amurense*–*C. clementina* divergence (25 to 69 Ma), *M. azedarach*–*P. vera* (63 to 105 Ma), *A. yangbiense*–*V. vinifera* (110 to 124 Ma), and *A. cherimola*–*A. trichopoda* (180 to 205 Ma) were obtained from the TimeTree database (www.timetree.org).

Gene family expansion and contraction were analyzed using CAFE (v4.2.1) (*72*). On the basis of Gene Ontology and Kyoto Encyclopedia of Genes and Genomes database, the functional enrichment of expanded gene families was analyzed using R package ClusterProfiler (v4.0) (*73*).

WGD analysis

Syntenic blocks within and between genomes were established using MCScan (Python version) with default parameters. Syntenic blocks were identified on the basis of detected homologous gene pairs, and K_s between collinear genes were estimated using the Whole-Genome Duplication Identifier (WGDI) pipeline (v0.6.5) (*74*). In addition, to detect relics of possible polyploidization in the *P. amurense* genomes, the syntenic blocks of paralogous genes within each genome of *P. amurense*, *Z. armatum*, and *C. clementina* were identified using MCScan. On the basis of the *C. clementina* genome, collinear genes of each block were extracted, and a collinear gene tree was constructed using IQ-TREE2 (v2.2.2.7) (*75*). The gene trees were used as input and uploaded to ASTRAL (v5.6.1) (*76*) to infer a subgenome phylogeny. Last, the ksrates package (https://github.com/VIB-PSB/ksrates) was used to trace the WGD events based on adjusted mixed plots of paralog and ortholog K_s distributions.

Identification of candidate genes and correlation analysis

Reported BIA biosynthetic genes, including NCS, MTs, and CYP450, were chosen as the query sequences, and BLASTp was performed to identify the homologous genes from the genomes of *P. amurense*, *Z. armatum*, and *C. clementina* with an *E* value of 1×10^{-10}. Briefly, PsNCS1 (AAX56303.1) and PsNCS2 (AAX56304.1) from *Papaver somniferum*, CjNCS (BAF45338.2) from *C. japonica*, and TfNCS (ACO90248.1) from *T. flavum* were used as query sequences to identify candidate NCS encoding genes. Ps6OMT (AAQ01669.1) from *P. somniferum* and Cj6OMT (BAB08004.1), Cj4′OMT (BAB08005.1), and Cj9OMT (BAA06192.1) from *C. japonica* were used as query sequences to identify the homologs of OMTs. CjNMT (AB061863.1) from *C. japonica* was used as query sequences to identify NMTs. CjCYP719A1 (BAB68769.1), CjCYP80B2 (BAB12433.1), and CjCYP80G2 (BAF80448.1) from *C. japonica*, PsCYP80B1 (AAF61400.1) and PsCYP719B1 (ABR14720.1) from *P. somniferum*, and BsCYP80A1 (AAC48987.1) from *Berberis stolonifera* were used as query sequences to identify the candidate subfamilies of CYP450s (identity > 55%) according to the nomenclature system.

On the basis of the transcriptome (PRJNA859281) and metabolome data of *P. amurense*, the normalized expression levels of candidate genes and the content level of berberine were used to analyze Pearson's correlation coefficients (r). The candidate MTs and CYP450s with high correlation ($r > 0.8$) were selected for functional assays. The functional gene was further chosen as bait to select additional candidate genes according to the coexpression analysis.

Heterologous expression and functional validation in *E. coli* of MTs genes

The candidate MT genes were cloned or synthesized and then cloned into the pET-28a vector (table S12). The resulting positive recombinant plasmids were introduced to *Escherichia coli* strain BL21 (DE3). Positive colonies were selected for each vector to induce each recombinant protein. A common protocol was developed to express each candidate gene and to produce their recombinant protein. A positive colony was inoculated in 20 ml of liquid Luria-Bertani (LB) medium supplemented with kanamycin (50 μg/ml) contained in a 100-ml E-flask. The flask was placed on a shaker set up a speed of 250 rpm/min at 37°C. When the optical density at 600 nm (OD_{600}) value of cell suspension culture reached 0.6, isopropyl-β-D-thiogalactopyranoside (IPTG; 0.3 mM) was added to induce protein expression. After the addition of IPTG, the temperature of incubator was changed to 16°C. After an overnight culture, cells were harvested by centrifugation (8000*g*, 5 min, 4°C). The pellet was resuspended in 1 ml of tris-HCl buffer [50 mM tris-HCl (pH 7.4), 10% (v/v) glycerol, and 5 mM β-mercaptoethanol]. The cells lysed via ultrasonication on ice for 5 min. The lysate was centrifuged at 8000*g* for 10 min. The supernatant was transferred to a new tube and stored in −20° for enzymatic assays.

For purification of candidate recombinant proteins, the culture volume of *E. coli* for each recombinant vector was scaled up to 100 ml of LB medium. Other conditions were the same. The pellet was resuspended in 30 ml of lysis buffer (pH 8.0; 50 mM NaH_2PO_4, 30 mM NaCl, and 10 mM imidazole). The extraction steps were the same as described for small scales. The resulting 30 ml of supernatant was loaded onto Ni-NDA column (Ni-NDA Beads, Smart-Lifesciences, China) to purify recombinant proteins. The proteins for crystallization were further purified by gel-filtration chromatography equilibrated with 150 mM NaCl and 20 mM Hepes (pH 8.0), and the final protein was concentrated using Amicon Ultra-15 Ultracel-30 K centrifuge filters (Merck Millipore). The protein concentration was determined using the Protein Quantitative Kit (TransGen Biotech, China), and the purity was assessed by SDS–polyacrylamide gel electrophoresis.

Enzyme assays of MTs were performed overnight at 37°C with shaking at 200 rpm in a 100 μl of reaction system with Gly buffer (pH 9.0) containing 10 μg of purified proteins, 2 mM SAM, and the substrate of (*S*)-norcoclaurine, (*S*)-coclaurine, or (*S*)-scoulerine at 50 μg/ml or of 3′-hydroxy-*N*-methylcoclaurine at 20 μg/ml, and the reaction was quenched by adding methanol (50%, v/v). Cc6OMT, Cc4′OMT, Cc9OMT, and CcNMT were chosen as positive controls, whereas the empty vector was negative.

The reaction samples were detected by ultra-performance LC–tandem MS (UPLC-MS/MS). The UPLC-MS/MS system (SCIEX TripleTOF 6600+) equipped with a triple quadrupole-linear ion trap mass spectrometer. In brief, 1 μl of extract was injected and then separated in a Kinetex C18 100A analytical column (4.6 mm by 150 mm, 2.6 μm) placed in 35°C chamber. The mobile phase consisted of solution A (0.1% formic acid) and B (acetonitrile). A gradient program used was composed of 0 to 1 min, 10% B; 1 to 10 min, 10 to 95% B; 12.3 to 13 min, 95% B; and 13 to 15 min, 95 to 10% B. The flow rate was 0.5 ml/min. The mass spectrometer was operated in multiple reaction detection mode with a scan time of 35 ms per transition. The parameters for mass spectrometer were set as follows: ESI mode in positive ion mode; spray voltage, 3.5 kV; spray temperature, 550°C; curtain gas, 35 psi; GAS1, 40 psi; and GAS2, 60 psi.

Heterologous expression and functional validation in *S. cerevisiae* of CYPs genes

To validate the candidate CYP450 activities, the yeast strain *Saccharomyces cerevisiae* WAT11 engineered with *Arabidopsis thaliana* cytochrome P450 reductase (AtCPR1) was chosen as the heterologous expression host. The candidate CYP450 genes were cloned or synthesized into the pESC-His vector (primer listed in table S12) and transferred into *S. cerevisiae* WAT11 (*29*). Positive yeast colonies were screened to express the CYP450 proteins. After cultured and induced, yeast cells were harvested, crushed, and differential centrifugated to obtain microsomes as described in our previous study. For the enzyme assay, yeast expressing pESC-His empty vector was selected as negative control. The reaction system including 200 μl of microsome, 0.5 mM reduced form of nicotinamide adenine dinucleotide phosphate, and 1 μl of substrate (1 mg/ml) was incubated at 30°C for 2 hours. The reaction samples were detected by Agilent UPLC-QTOF-MS equipped with an ACQUITY UPLC HSS-T3 C18 column (1.8 μm, 2.1 mm by 100 mm) after centrifugation and filtration. The mobile phase consisted of 0.1% formic acid in water (A) and acetonitrile (B), and the linear gradient elution program was as follows: 0 to 2 min, 5% B; 2 to 10 min, 5 to 18% B; 10 to 15 min, 18 to 95% B; and 15 to 17 min, 95% B. The flow rate was set as 0.3 ml/min, and the column temperature was 40°C. The mass spectrometer used was the Agilent 6430 QTOF, which has an ESI source operating in positive ion mode (*m/z* 100 to 800). The gas temperature was maintained at 350°C, the gas flow rate was 8 liters/min, the fragment was set to 120 V, and the collision energy was 20 eV.

Transient expression of PaCYP71BG29 in *N. benthamiana*

To further validate the activity of functional PaCYP71BG29 in vivo, the gene was cloned into the pEAQ-eGFP vector and transformed into *Agrobacterium tumefaciens* (GV3101). Positive clones were cultured in 10 ml of LB medium supplemented with kanamycin and rifampicin. The cultures were centrifuged at 5000*g* for 5 min, and the pellets were resuspended in infiltration buffer (10 mM MES buffer, 10 mM $MgCl_2$, and 150 μM acetosyringone) and incubated at 28°C in the dark for 1 hour. *Agrobacterium* suspensions (OD_{600} of 0.3 to 0.6 for each strain) were infiltrated into 5- to 6-week-old *N. benthamiana* leaves. After 3 days, the substrates **5** and **7** (50 μM, in 0.1% dimethyl sulfoxide in water) were infiltrated into the abaxial side of the previously *Agrobacterium*-infiltrated leaves. After 1 day, leaves were collected for LC-MS/MS analysis. Biological replicates consisted of three or four leaves collected from three different lines. Leaves infiltrated with an empty vector and candidate genes without substrates were used as controls.

Crystallization and mutagenesis of PaNOMT9

For crystallization of PaNOMT9/**4**/SAM complex, purified PaNOMT9 protein (20 mg/ml) was incubated with 5 mM (*S*)-coclaurine (**4**) and 5 mM SAM. Then, crystals were grown by sitting-drop vapor diffusion at 4°C. The precipitant well solution consisted of 20% w/v PEG1000, 100 mM lithium sulfate monohydrate, and 100 mM sodium citrate tribasic dihydrate (pH 5.5). Crystals were gradually transferred into a precipitant well solution with 25% glycerol, followed by flash-freezing in liquid nitrogen for storage. All the datasets were collected under cryogenic conditions (100 K) at Shanghai Synchrotron Radiation Facility (SSRF) beamlines BL18U1, and structure was solved by molecular replacement with searching models 6I70. The model was then refined using Phenix (*77*), together with manual

building in Coot (*78*). The final structure has been refined to working *R*-factors (R_{work}) of 15.9% [free *R*-factors (R_{free}) = 19.3%] with good geometry (table S13). Substitutions for the selected residues in PaNOMT9 were constructed by gene splicing by overlap extension PCR, and the mutant activity was tested using **4** as the substrate (with three replicates).

Supplementary Materials

This PDF file includes:

Figs. S1 to S33
Tables S1 to S13

REFERENCES AND NOTES

1. Y. Wang, Q. Tong, S. R. Ma, Z. X. Zhao, L. B. Pan, L. Cong, P. Han, R. Peng, H. Yu, Y. Lin, T. L. Gao, J. W. Shou, X. Y. Li, X. F. Zhang, Z. W. Zhang, J. Fu, B. Y. Wen, J. B. Yu, X. Cao, J. D. Jiang, Oral berberine improves brain dopa/dopamine levels to ameliorate Parkinson's disease by regulating gut microbiota. *Signal Transduct. Target. Ther.* **6**, 77 (2021).
2. Y. Zhang, Y. Gu, H. Ren, S. Wang, H. Zhong, X. Zhao, J. Ma, X. Gu, Y. Xue, S. Huang, J. Yang, L. Chen, G. Chen, S. Qu, J. Liang, L. Qin, Q. Huang, Y. Peng, Q. Li, X. Wang, P. Kong, G. Hou, M. Gao, Z. Shi, X. Li, Y. Qiu, Y. Zou, H. Yang, J. Wang, G. Xu, S. Lai, J. Li, G. Ning, W. Wang, Gut microbiome-related effects of berberine and probiotics on type 2 diabetes (the PREMOTE study). *Nat. Commun.* **11**, 5015 (2020).
3. S. Wang, H. Ren, H. Zhong, X. Zhao, C. Li, J. Ma, X. Gu, Y. Xue, S. Huang, J. Yang, L. Chen, G. Chen, S. Qu, J. Liang, L. Qin, Q. Huang, Y. Peng, Q. Li, X. Wang, Y. Zou, Z. Shi, X. Li, T. Li, H. Yang, S. Lai, G. Xu, J. Li, Y. Zhang, Y. Gu, W. Wang, Combined berberine and probiotic treatment as an effective regimen for improving postprandial hyperlipidemia in type 2 diabetes patients: A double blinded placebo controlled randomized study. *Gut. Microbes* **14**, 2003176 (2022).
4. M. M. Zhao, J. Lu, S. Li, H. Wang, X. Cao, Q. Li, T. T. Shi, K. Matsunaga, C. Chen, H. Huang, T. Izumi, J. K. Yang, Berberine is an insulin secretagogue targeting the KCNH6 potassium channel. *Nat. Commun.* **12**, 5616 (2021).
5. M. Ekstrom, D. Ferreira, S. Chang, S. Louw, M. J. Johnson, D. J. Eckert, B. Fazekas, K. J. Clark, M. R. Agar, D. C. Currow, Australian National Palliative Care Clinical Studies Collaborative, Effect of regular, low-dose, extended-release morphine on chronic breathlessness in chronic obstructive pulmonary disease: The BEAMS Randomized Clinical Trial. *JAMA* **328**, 2022–2032 (2022).
6. Z. Wu, L. G. Spencer, W. Banya, J. Westoby, V. A. Tudor, P. Rivera-Ortega, N. Chaudhuri, I. Jakupovic, B. Patel, M. Thillai, A. West, M. Wijsenbeek, T. M. Maher, J. A. Smith, P. L. Molyneaux, Morphine for treatment of cough in idiopathic pulmonary fibrosis (PACIFY COUGH): A prospective, multicentre, randomised, double-blind, placebo-controlled, two-way crossover trial. *Lancet Respir. Med.* **12**, 273–280 (2024).
7. Q. Su, M. Fan, J. Wang, A. Ullah, M. A. Ghauri, B. Dai, Y. Zhan, D. Zhang, Y. Zhang, Sanguinarine inhibits epithelial-mesenchymal transition via targeting HIF-1alpha/TGF-beta feed-forward loop in hepatocellular carcinoma. *Cell Death Dis.* **10**, 939 (2019).
8. K. D. Wang, M. L. Zhu, C. J. Qin, R. F. Dong, C. M. Xiao, Q. Lin, R. Y. Wei, X. Y. He, X. Zang, L. Y. Kong, Y. Z. Xia, Sanguinarine induces apoptosis in osteosarcoma by attenuating the binding of STAT3 to the single-stranded DNA-binding protein 1 (SSBP1) promoter region. *Br. J. Pharmacol.* **180**, 3175–3193 (2023).
9. W. Kukula-Koch, The elevation of LC-ESI-Q-TOF-MS response in the analysis of isoquinoline alkaloids from some Papaveraceae and Berberidaceae representatives. *J. Anal. Methods Chem.* **2017**, 1–9 (2017).
10. T. Morikawa, N. Kitagawa, G. Tanabe, K. Ninomiya, S. Okugawa, C. Motai, I. Kamei, M. Yoshikawa, I. J. Lee, O. Muraoka, Quantitative determination of alkaloids in lotus flower (flower buds of *Nelumbo nucifera*) and their melanogenesis inhibitory activity. *Molecules* **21**, 930 (2016).
11. I. Ware, K. Franke, A. Frolov, K. Bureiko, E. Kysil, M. Yahayu, H. A. El Enshasy, L. A. Wessjohann, Comparative metabolite analysis of *Piper sarmentosum* organs approached by LC-MS-based metabolic profiling. *Nat. Prod. Bioprospect.* **14**, 30 (2024).
12. L. Leng, Z. Xu, B. Hong, B. Zhao, Y. Tian, C. Wang, L. Yang, Z. Zou, L. Li, K. Liu, W. Peng, J. Liu, Z. An, Y. Wang, B. Duan, Z. Hu, C. Zheng, S. Zhang, X. Li, M. Li, Z. Liu, Z. Bi, T. He, B. Liu, H. Fan, C. Song, Y. Tong, S. Chen, Cepharanthine analogs mining and genomes of *Stephania* accelerate anti-coronavirus drug discovery. *Nat. Commun.* **15**, 1537 (2024).
13. S. S. Wen, P. Li, W. Gao, Characterization and identification of alkaloids in *Phellodendri Chinensis* Cortex and *Phellodendri amurensis* Cortex based on UHPLC-IM-Q-TOF-MS. *Zhongguo Zhong Yao Za Zhi* **48**, 3294–3307 (2023).
14. D. Song, J. Hao, D. Fan, Biological properties and clinical applications of berberine. *Front. Med.* **14**, 564–582 (2020).
15. Y. Liu, B. Wang, S. Shu, Z. Li, C. Song, D. Liu, Y. Niu, J. Liu, J. Zhang, H. Liu, Z. Hu, B. Huang, X. Liu, W. Liu, L. Jiang, M. M. Alami, Y. Zhou, Y. Ma, X. He, Y. Yang, T. Zhang, H. Hu, M. S. Barker, S. Chen, X. Wang, J. Nie, Analysis of the *Coptis chinensis* genome reveals the diversification of protoberberine-type alkaloids. *Nat. Commun.* **12**, 3276 (2021).
16. N. Samanani, D. K. Liscombe, P. J. Facchini, Molecular cloning and characterization of norcoclaurine synthase, an enzyme catalyzing the first committed step in benzylisoquinoline alkaloid biosynthesis. *Plant J.* **40**, 302–313 (2004).
17. H. Minami, E. Dubouzet, K. Iwasa, F. Sato, Functional analysis of norcoclaurine synthase in *Coptis japonica*. *J. Biol. Chem.* **282**, 6274–6282 (2007).
18. F. Sato, T. Tsujita, Y. Katagiri, S. Yoshida, Y. Yamada, Purification and characterization of *S*-adenosyl-*L*-methionine: Norcoclaurine 6-*O*-methyltransferase from cultured *Coptis japonica* cells. *Eur. J. Biochem.* **225**, 125–131 (1994).
19. T. Morishige, T. Tsujita, Y. Yamada, F. Sato, Molecular characterization of the *S*-adenosyl-L-methionine:3′-hydroxy-*N*-methylcoclaurine 4′-*O*-methyltransferase involved in isoquinoline alkaloid biosynthesis in *Coptis japonica*. *J. Biol. Chem.* **275**, 23398–23405 (2000).
20. K. B. Choi, T. Morishige, N. Shitan, K. Yazaki, F. Sato, Molecular cloning and characterization of coclaurine *N*-methyltransferase from cultured cells of *Coptis japonica*. *J. Biol. Chem.* **277**, 830–835 (2002).
21. P. Steffens, N. Nagakura, M. H. Zenk, Purification and characterization of the berberine bridge enzyme from *Berberis beaniana* cell cultures. *Phytochemistry* **24**, 2577–2583 (1985).
22. S. M. He, Y. L. Liang, K. Cong, G. Chen, X. Zhao, Q. M. Zhao, J. J. Zhang, X. Wang, Y. Dong, J. L. Yang, G. H. Zhang, Z. L. Qian, W. Fan, S. C. Yang, Identification and characterization of genes involved in benzylisoquinoline alkaloid biosynthesis in *Coptis* species. *Front. Plant Sci.* **9**, 731 (2018).
23. L. Wu, B. Zhao, Z. Deng, B. Wang, Y. Yu, A biosynthetic network for protoberberine production in *Coptis chinensis*. *Hortic. Res.* **11**, uhad259 (2024).
24. N. Ikezawa, M. Tanaka, M. Nagayoshi, R. Shinkyo, T. Sakaki, K. Inouye, F. Sato, Molecular cloning and characterization of CYP719, a methylenedioxy bridge-forming enzyme that belongs to a novel P450 family, from cultured *Coptis japonica* cells. *J. Biol. Chem.* **278**, 38557–38565 (2003).
25. N. Okada, A. Shinmyo, H. Okada, Y. Yamada, Purification and characterization of (*S*)-tetrahydroberberine oxidase from cultured *Coptis japonica* cells. *Phytochemistry* **27**, 979–982 (1988).
26. J. Han, S. Li, *De novo* biosynthesis of berberine and halogenated benzylisoquinoline alkaloids in *Saccharomyces cerevisiae*. *Commun. Chem.* **6**, 27 (2023).
27. Y. Tian, L. Kong, Q. Li, Y. Wang, Y. Wang, Z. An, Y. Ma, L. Tian, B. Duan, W. Sun, R. Gao, S. Chen, Z. Xu, Structural diversity, evolutionary origin, and metabolic engineering of plant specialized benzylisoquinoline alkaloids. *Nat. Prod. Rep.* **41**, 1787–1810 (2024); https://doi.org/10.1039/D4NP00029C.
28. F. C. Huang, T. M. Kutchan, Distribution of morphinan and benzo[c]phenanthridine alkaloid gene transcript accumulation in *Papaver somninferum*. *Phytochemistry* **53**, 555–564 (2000).
29. Z. An, R. Gao, S. Chen, Y. Tian, Q. Li, L. Tian, W. Zhang, L. Kong, B. Zheng, L. Hao, T. Xin, H. Yao, Y. Wang, W. Song, X. Hua, C. Liu, J. Song, H. Fan, W. Sun, S. Chen, Z. Xu, Lineage-specific CYP80 expansion and benzylisoquinoline alkaloid diversity in early-diverging eudicots. *Adv. Sci.* **11**, e2309990 (2024).
30. C. Hao, Y. Yu, Y. Liu, A. Liu, S. Chen, The CYP80A and CYP80G are involved in the biosynthesis of benzylisoquinoline alkaloids in the sacred lotus (*Nelumbo nucifera*). *Int. J. Mol. Sci.* **25**, 702 (2024).
31. X. Liu, X. Jiao, Y. Cheng, Y. Ma, J. Bu, B. Jin, Q. Li, Z. Hu, J. Tang, C. Lai, J. Wang, G. Cui, Y. Chen, J. Guo, L. Huang, Structure-function analysis of CYP719As involved in methylenedioxy bridge-formation in the biosynthesis of benzylisoquinoline alkaloids and its *de novo* production. *Microb. Cell Fact.* **22**, 23 (2023).
32. A. Y. Robin, C. Giustini, M. Graindorge, M. Matringe, R. Dumas, Crystal structure of norcoclaurine-6-*O*-methyltransferase, a key rate-limiting step in the synthesis of benzylisoquinoline alkaloids. *Plant J.* **87**, 641–653 (2016).
33. M. R. Bennett, M. L. Thompson, S. A. Shepherd, M. S. Dunstan, A. J. Herbert, D. R. M. Smith, V. A. Cronin, B. R. K. Menon, C. Levy, J. Micklefield, Structure and biocatalytic scope of coclaurine *N*-methyltransferase. *Angew. Chem. Int. Ed. Engl.* **57**, 10600–10604 (2018).
34. T. Fujiwara, S. Maisonneuve, M. Isshiki, M. Mizutani, L. Chen, H. L. Wong, T. Kawasaki, K. Shimamoto, Sekiguchi lesion gene encodes a cytochrome P450 monooxygenase that catalyzes conversion of tryptamine to serotonin in rice. *J. Biol. Chem.* **285**, 11308–11313 (2010).
35. H. M. Gaweska, K. M. Roberts, P. F. Fitzpatrick, Isotope effects suggest a stepwise mechanism for berberine bridge enzyme. *Biochemistry* **51**, 7342–7347 (2012).
36. D. K. Liscombe, G. V. Louie, J. P. Noel, Architectures, mechanisms and molecular evolution of natural product methyltransferases. *Nat. Prod. Rep.* **29**, 1238–1250 (2012).
37. J. S. Morris, P. J. Facchini, Molecular origins of functional diversity in benzylisoquinoline alkaloid methyltransferases. *Front. Plant Sci.* **10**, 1058 (2019).

38. B. Rohde, J. Hans, S. Martens, A. Baumert, P. Hunziker, U. Matern, Anthranilate *N*-methyltransferase, a branch-point enzyme of acridone biosynthesis. *Plant J.* **53**, 541–553 (2008).
39. R. Huang, A. J. O'Donnell, J. J. Barboline, T. J. Barkman, Convergent evolution of caffeine in plants by co-option of exapted ancestral enzymes. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 10613–10618 (2016).
40. P. J. Facchini, C. Penzes, A. G. Johnson, D. Bull, Molecular characterization of berberine bridge enzyme genes from opium poppy. *Plant Physiol.* **112**, 1669–1677 (1996).
41. P. Huang, W. Liu, M. Xu, R. Jiang, L. Xia, P. Wang, H. Li, Z. Tang, Q. Zheng, J. Zeng, Modulation of benzylisoquinoline alkaloid biosynthesis by overexpression berberine bridge enzyme in Macleaya cordata. *Sci. Rep.* **8**, 17988 (2018).
42. A. Winkler, F. Hartner, T. M. Kutchan, A. Glieder, P. Macheroux, Biochemical evidence that berberine bridge enzyme belongs to a novel family of flavoproteins containing a bi-covalently attached FAD cofactor. *J. Biol. Chem.* **281**, 21276–21285 (2006).
43. Z. Xu, Z. Li, F. Ren, R. Gao, Z. Wang, J. Zhang, T. Zhao, X. Ma, X. Pu, T. Xin, S. Rombauts, W. Sun, Y. Van de Peer, S. Chen, J. Song, The genome of *Corydalis* reveals the evolution of benzylisoquinoline alkaloid biosynthesis in Ranunculales. *Plant J.* **111**, 217–230 (2022).
44. W. Sun, Z. Xu, C. Song, S. Chen, Herbgenomics: Decipher molecular genetics of medicinal plants. *Innovation Camb.* **3**, 100322 (2022).
45. P. Berman, L. A. de Haro, A. Jozwiak, S. Panda, Z. Pinkas, Y. Dong, J. Cveticanin, R. Barbole, R. Livne, T. Scherf, E. Shimoni, S. Levin-Zaidman, N. Dezorella, E. Petrovich-Kopitman, S. Meir, I. Rogachev, P. D. Sonawane, A. Aharoni, Parallel evolution of cannabinoid biosynthesis. *Nat. Plants* **9**, 817–831 (2023).
46. Z. Xu, X. Pu, R. Gao, O. C. Demurtas, S. J. Fleck, M. Richter, C. He, A. Ji, W. Sun, J. Kong, K. Hu, F. Ren, J. Song, Z. Wang, T. Gao, C. Xiong, H. Yu, T. Xin, V. A. Albert, G. Giuliano, S. Chen, J. Song, Tandem gene duplications drive divergent evolution of caffeine and crocin biosynthetic pathways in plants. *BMC Biol.* **18**, 63 (2020).
47. L. Mao, H. Kawaide, T. Higuchi, M. Chen, K. Miyamoto, Y. Hirata, H. Kimura, S. Miyazaki, M. Teruya, K. Fujiwara, K. Tomita, H. Yamane, K. I. Hayashi, H. Nojiri, L. Jia, J. Qiu, C. Ye, M. P. Timko, L. Fan, K. Okada, Genomic evidence for convergent evolution of gene clusters for momilactone biosynthesis in land plants. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 12472–12480 (2020).
48. J. K. Weng, The evolutionary paths towards complexity: A metabolic perspective. *New Phytol.* **201**, 1141–1149 (2014).
49. W. Sun, Q. Yin, H. Wan, R. Gao, C. Xiong, C. Xie, X. Meng, Y. Mi, X. Wang, C. Wang, W. Chen, Z. Xie, Z. Xue, H. Yao, P. Sun, X. Xie, Z. Hu, D. R. Nelson, Z. Xu, X. Sun, S. Chen, Characterization of the horse chestnut genome reveals the evolution of aescin and aesculin biosynthesis. *Nat. Commun.* **14**, 6470 (2023).
50. H. Cheng, G. T. Concepcion, X. Feng, H. Zhang, H. Li, Haplotype-resolved *de novo* assembly using phased assembly graphs with hifiasm. *Nat. Methods* **18**, 170–175 (2021).
51. D. Guan, S. A. McCarthy, J. Wood, K. Howe, Y. Wang, R. Durbin, Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics* **36**, 2896–2898 (2020).
52. X. Zhang, S. Zhang, Q. Zhao, R. Ming, H. Tang, Assembly of allele-aware, chromosomal-scale autopolyploid genomes based on Hi-C data. *Nat. Plants* **5**, 833–845 (2019).
53. D. Kim, J. M. Paggi, C. Park, C. Bennett, S. L. Salzberg, Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915 (2019).
54. T. Bruna, K. J. Hoff, A. Lomsadze, M. Stanke, M. Borodovsky, BRAKER2: Automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genom. Bioinform.* **3**, lqaa108 (2021).
55. M. Manni, M. R. Berkeley, M. Seppey, F. A. Simao, E. M. Zdobnov, BUSCO Update: Novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol. Biol. Evol.* **38**, 4647–4654 (2021).
56. R. Behboudi, M. Nouri-Baygi, M. Naghibzadeh, RPTRF: A rapid perfect tandem repeat finder tool for DNA sequences. *Biosystems* **226**, 104869 (2023).
57. S. Ou, N. Jiang, LTR_retriever: A highly accurate and sensitive program for identification of long terminal repeat retrotransposons. *Plant Physiol.* **176**, 1410–1422 (2018).
58. Y. Wang, J. Hao, X. Yuan, Y. Sima, B. Lu, The complete chloroplast genome sequence of *Zanthoxylum armatum*. *Mitochondrial. DNA B Resour.* **4**, 2513–2514 (2019).
59. L. Hu, Z. Xu, R. Fan, G. Wang, F. Wang, X. Qin, L. Yan, X. Ji, M. Meng, S. Sim, W. Chen, C. Hao, Q. Wang, H. Zhu, S. Zhu, P. Xu, H. Zhao, K. Lindsey, H. Daniell, J. F. Wendel, S. Jin, The complex genome and adaptive evolution of polyploid Chinese pepper (*Zanthoxylum armatum* and *Zanthoxylum bungeanum*). *Plant Biotechnol. J.* **21**, 78–96 (2023).
60. N. Zhou, A. G. Fu, Characterization of the complete chloroplast genome of *Murraya exotica* (Rutaceae) from Yunnan Province, China. *Mitochondrial. DNA B Resour.* **6**, 2288–2290 (2021).
61. C. Sun, H. Lin, The complete chloroplast genome and phylogenetic analysis of *Citrus clementina* (Rutaceae). *Mitochondrial. DNA B Resour.* **6**, 2926–2927 (2021).
62. X. Wang, Y. Xu, S. Zhang, L. Cao, Y. Huang, J. Cheng, G. Wu, S. Tian, C. Chen, Y. Liu, H. Yu, X. Yang, H. Lan, N. Wang, L. Wang, J. Xu, X. Jiang, Z. Xie, M. Tan, R. M. Larkin, L. L. Chen, B. G. Ma, Y. Ruan, X. Deng, Q. Xu, Genomic analyses of primitive, wild and cultivated citrus provide insights into asexual reproduction. *Nat. Genet.* **49**, 765–772 (2017).
63. B. Liao, J. Tan, W. Zhou, Y. Wang, Y. Li, X. Chen, The complete genome sequence of *Melia azedarach* Linn. (Meliaceae): A multi-purpose pesticide species. *Mitochondrial. DNA B Resour.* **7**, 1103–1105 (2022).
64. W. Mao, G. Yao, B. Chen, Q. Pan, H. Wang, G. Hu, G. Xin, Characterization of the complete chloroplast genome of the *Pistacia vera* L. *Mitochondrial. DNA B. Resour.* **4**, 4140–4141 (2019).
65. J. Yang, H. M. Wariss, L. Tao, R. Zhang, Q. Yun, P. Hollingsworth, Z. Dao, G. Luo, H. Guo, Y. Ma, W. Sun, *De novo* genome assembly of the endangered *Acer yangbiense*, a plant species with extremely small populations endemic to Yunnan Province, China. *Gigascience* **8**, giz085 (2019).
66. M. J. N. Ramos, J. L. Coito, D. Faisca-Silva, J. Cunha, M. M. R. Costa, S. Amancio, M. Rocheta, Portuguese wild grapevine genome re-sequencing (*Vitis vinifera* sylvestris). *Sci. Rep.* **10**, 18993 (2020).
67. X. Liu, Y. Liu, P. Huang, Y. Ma, Z. Qing, Q. Tang, H. Cao, P. Cheng, Y. Zheng, Z. Yuan, Y. Zhou, J. Liu, Z. Tang, Y. Zhuo, Y. Zhang, L. Yu, J. Huang, P. Yang, Q. Peng, J. Zhang, W. Jiang, Z. Zhang, K. Lin, D. K. Ro, X. Chen, X. Xiong, Y. Shang, S. Huang, J. Zeng, The genome of medicinal plant *Macleaya cordata* provides new insights into benzylisoquinoline alkaloids metabolism. *Mol. Plant* **10**, 975–989 (2017).
68. A. Talavera, N. Fernandez-Pozo, A. J. Matas, J. I. Hormaza, A. Bombarely, Genomics in neglected and underutilized fruit crops: A chromosome-scale genome sequence of cherimoya. *Plants People Planet* **5**, 408–423 (2023).
69. Amborella Genome Project, The Amborella genome and the evolution of flowering plants. *Science* **342**, 1241089 (2013).
70. D. M. Emms, S. Kelly, OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**, 238 (2019).
71. Z. Yang, PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).
72. T. De Bie, N. Cristianini, J. P. Demuth, M. W. Hahn, CAFE: A computational tool for the study of gene family evolution. *Bioinformatics* **22**, 1269–1271 (2006).
73. T. Wu, E. Hu, S. Xu, M. Chen, P. Guo, Z. Dai, T. Feng, L. Zhou, W. Tang, L. Zhan, X. Fu, S. Liu, X. Bo, G. Yu, clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation Camb.* **2**, 100141 (2021).
74. P. Sun, B. Jiao, Y. Yang, L. Shan, T. Li, X. Li, Z. Xi, X. Wang, J. Liu, WGDI: A user-friendly toolkit for evolutionary analyses of whole-genome duplications and ancestral karyotypes. *Mol. Plant* **15**, 1841–1851 (2022).
75. B. Q. Minh, H. A. Schmidt, O. Chernomor, D. Schrempf, M. D. Woodhams, A. von Haeseler, R. Lanfear, IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).
76. C. Zhang, C. Scornavacca, E. K. Molloy, S. Mirarab, ASTRAL-Pro: Quartet-based species-tree inference despite paralogy. *Mol. Biol. Evol.* **37**, 3292–3307 (2020).
77. P. D. Adams, P. V. Afonine, G. Bunkoczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L. W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger, P. H. Zwart, PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010).
78. P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).

Acknowledgments: We thank D. Nelson (The University of Tennessee Health Science Center) of the P450 nomenclature committee for the naming of the P450s. **Funding:** This work was supported by the National Key R&D Program of China [grant number 2023YFC3504800 (to Z. Xu)], the National Natural Science Foundation of China [grant numbers 82274037 (to Z. Xu), 82360754 (to L.T.), 82204577 (to R.G.), and U23A20511 (to H.L.)], the Fundamental Research Funds for the Central public welfare research institutes [grant numbers ZZ16-YQ-047 (to R.G.) and ZZ16-ND-10-02 (to R.G.)], the academician and expert workstation of Yunnan province [grant number 202205AF150026 (to S.C.)], and Heilongjiang Touyan Innovation Team Program (Tree Genetics and Breeding Innovation Team). **Author contributions:** Conceptualization: Z. Xu. Formal analysis: J.W. and W.Z. Visualization: Y.T., J.W. and W.Z. Validation: Y.T., Y.M., Q.L., Y. Zhou., T.L., L.K., Y.W., Z.A., H.L., and R.G. Project administration: Z. Xu, W.S., and S.C. Resources: Z. Xie, B.Z., Y. Zhang., C.C., C.L., L.T., C.F., J.L., Y.H., J.S., and B.D. Writing—original draft: Y.T. and J.W. Writing—review and editing: Z. Xu, Y.T., H.L., R.G., W.S., and S.C. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. The raw data of genome sequencing reported in this paper have been deposited in the Genome Sequence Archive in BIG Data Center, Beijing Institute of Genomics (BIG), Chinese Academy of Sciences, under accession number CRA016709 (PRJCA026511) that are publicly accessible at http://bigd.big.ac.cn/gsa. The genome assembly and gene annotation datasets have been also deposited in figshare (https://doi.org/10.6084/m9.figshare.27310872.v1).

Submitted 9 August 2024
Accepted 28 October 2024
Published 29 November 2024
10.1126/sciadv.ads3596