The biosynthetic origin of psychoactive kavalactones in kava

Tomáš Pluskal[1], Michael P. Torrens-Spence[1], Timothy R. Fallon[1,2], Andrea De Abreu[1,2], Cindy H. Shi[1,2], and Jing-Ke Weng[1,2,*]

[1]Whitehead Institute for Biomedical Research, 455 Main Street, Cambridge, MA 02142 USA.

[2]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139 USA.

[*]Corresponding author. Email: wengj@wi.mit.edu

Kava (*Piper methysticum*) is an ethnomedicinal shrub native to the Polynesian islands with well-established anxiolytic and analgesic properties. Its main psychoactive principles, kavalactones, form a unique class of polyketides that interact with the human central nervous system through mechanisms distinct from those of conventional psychiatric drugs. However, an unknown biosynthetic machinery and difficulty in chemical synthesis hinder the therapeutic use of kavalactones. In addition, kava contains several chalconoids with reported anti-cancer properties, known as flavokavains. Here, we report *de novo* elucidation of key components of the kavalactone and flavokavain biosynthetic network, encompassing eight specialized metabolic enzymes. We present the structural basis for the evolutionary development of a pair of paralogous styropyrone synthases that establish the kavalactone scaffold as well as the catalytic mechanism of a regio- and stereo-specific kavalactone reductase that produces a subset of chiral kavalactones. We further demonstrate the feasibility of engineering kavalactone production in heterologous hosts, thus opening an avenue towards development of novel non-addictive psychiatric therapeutics through the means of synthetic biology.

For millennia, humans have used plants for medicinal purposes. However, our limited understanding of plant biochemistry hinders the translation of such ancient wisdom into modern pharmaceuticals[1]. The *Piper* (pepper) genus, consisting of about 2,000 pantropically distributed species, has been long recognized as a remarkable source of bioactive specialized metabolites[2]. The dried fruits of black pepper (*Piper nigrum*) are used worldwide as a spice, and contain high content of alkaloids piperine and chavicine responsible for their pungent taste[3]. The Indian long pepper (*P. longum*), one of the oldest herbs described in Ayurvedic medicine, biosynthesizes the phenolic amide piperlongumine, which was recently shown to exhibit selective anticancer activities[4,5]. Leaves of the betel plant (*P. betle*) are commonly used as a wrapper for the chewing of psychoactive areca nut in South Asia. Betel leaves accumulate several specialized allylphenols, including chavibetol and hydroxychavicol, which may serve as antimutagens that counteract the carcinogenic effect of areca nut, while enhancing its psychoactive effect[6,7].

The kava plant (*P. methysticum;* Fig. 1a) is native to the Polynesian islands, together with species under the subgenus *Macropiper*, forming a South Pacific clade of *Piper* plants phylogenetically distant from *Piper* plants from Asia[8]. Throughout the Polynesian indigenous cultures, kava root is prepared into a beverage and consumed for its unique sedative and anesthetic effects in religious and cultural rituals. Its anxiolytic and analgesic properties are supported by over 3,000 years of traditional use as well as numerous recent clinical trials[9–12]. The bioactive ingredients of kava are phenolic polyketides collectively known as kavalactones (Supplementary Fig. 1a). Kava roots can contain up to 15% kavalactones by dry weight[13], whereby a person can ingest grams of kavalactones during one traditional kava drinking session. Although some concerns were raised in the past regarding potential hepatotoxicity of kava, it is

now generally accepted that the several reported toxicity cases arose from additional causative factors rather than kavalactones[14]. Due to its unique anxiolytic effect and generally safe nature, kava drinking has gained considerable popularity worldwide in recent years. Over 100 kava bars are now operating throughout the United States[15], and some athletes have embraced kava as a safe and non-addictive remedy for pain management[16]. Kavalactones interact with the human central nervous system through mechanisms distinct from those of common prescription psychiatric drugs such as benzodiazepines or opioids, and therefore show a great promise to be developed into new drugs for treating anxiety, insomnia and pain[17,18].

Despite the worldwide consumption of kava as a sedative beverage and the potential of kavalactones to be developed as a new class of anxiolytic medicines, the biosynthesis pathway of kavalactones is unknown. Here, we unravel key specialized metabolic enzymes involved in the biosynthesis of structurally diverse kavalactones in kava, and trace their evolutionary origins in other ancestral metabolic pathways. We also show it is feasible to utilize these newly discovered catalysts to engineer microbes to produce kavalactones as an alternative means to access these high-value natural products.

Results

The biosynthetic proposal for kavalactones. To investigate kavalactone biosynthesis, we first surveyed the taxonomic distribution of kavalactones within the *Piper* genus by untargeted metabolomics. Total methanolic extracts from kava and four other *Piper* species were analyzed by liquid chromatography high-resolution accurate-mass mass-spectrometry (LC-HRAM-MS). While kavalactones are among the most abundant metabolites present in kava

extract, they are absent in the *P. nigrum*, *P. betle*, *P. auritum*, and *P. sarmentosum* (Fig. 1b), suggesting the machinery underlying kavalactone biosynthesis likely emerged recently after kava diverged from other *Piper* species. In addition to kavalactones, we also detected flavokavain A, B, and C in kava, three previously characterized chalconoids with reported anti-cancer properties[19,20] (Fig. 1b).

Considering the structural relationship between kavalactones and flavokavains, which feature the styrylpyrone and chalcone backbones, respectively, we hypothesized that kavalactone biosynthesis likely involves styrylpyrone synthase (SPS), a polyketide synthase (PKS) related to chalcone synthase (CHS). CHS catalyzes the first committed enzymatic step in flavonoid biosynthesis and is ubiquitously present in all land plants[21] (Fig. 1c). Like CHS, SPS would accept a 4-coumarate-CoA ligase (4CL)-generated hydroxycinnamoyl-CoA thioester (e.g., *p*-coumaroyl-CoA) as a starter substrate, and catalyzes iterative decarboxylative condensations of two-carbon ketone units derived from its co-substrate, malonyl-CoA. While CHS produces a tetraketide intermediate that further cyclizes via Claisen condensation to yield the chalcone backbone, SPS would yield a triketide intermediate that lactonizes to form the styrylpyrone backbone (Fig. 1c). Additional tailoring enzymes, such as *O*-methyltransferases (OMTs) and oxidoreductases, are necessary to further decorate the initial scaffolds to give rise to the full repertoire of kavalactones and flavokavains present in kava plants (Supplementary Fig. 1).

Kava contains two SPSs derived from ancestral CHS. Kava is a decaploid (2n = 10x = 130 chromosomes)[22], rendering biosynthetic inquiry via conventional genetic approaches infeasible. We therefore turned to a multi-omics-guided candidate gene approach[23–25]. We first identified three *CHS*-like genes in a kava transcriptome assembled *de novo* from leaf and root

tissues (Supplementary Table 1). Phylogenetic analysis of these genes in the context of other land plant *CHS*s suggests that two of the three genes were likely derived from recent gene duplication events specific to kava (Fig. 2a). *In vitro* enzyme assays using heterologously expressed and purified enzymes established that these two *CHS*-like genes indeed encode functional SPSs. Both enzymes exclusively produce the triketide lactone bisnoryangonin from *p*-coumaroyl-CoA (Fig. 2b), and are therefore named *Pm*SPS1 and *Pm*SPS2 hereafter. The assay also suggested that the third *CHS*-like gene encodes a bona fide CHS, as it was the only enzyme capable of producing the expected tetraketide-derived naringenin chalcone, and we refer to it as *Pm*CHS hereafter (Fig. 2b). It is noted that in comparison to naringenin chalcone, *Pm*CHS produced significantly larger amounts of *p*-coumaroyltriacetic acid lactone (CTAL) and bisnoryangonin *in vitro*. Although such minor derailment products have been reported for a number of previously characterized CHSs[26,27], *Pm*CHS seems to be atypical in the sense that CTAL and bisnoryangonin are its major products *in vitro*, while naringenin chalcone is a minor product.

To assess the *in vivo* function of the three newly identified kava PKSs, we expressed each of them as stable transgenes in the *Arabidopsis thaliana CHS*-null mutant *tt4-2* background[28]. Whereas *PmCHS* restored flavonoid biosynthesis in *tt4-2*, neither *PmSPS1* nor *PmSPS2* did, consistent with their respective *in vitro* biochemical activities (Supplementary Fig. 2). However, we could not detect bisnoryangonin or other styrylpyrones in transgenic *A. thaliana* expressing *PmSPS1* or *PmSPS2*. Similarly, Agrobacterium-infiltration-mediated transient expression of *PmSPS1* or *PmSPS2* in the leaves of *Nicotiana benthamiana*[29] also failed to yield any styrylpyrone compounds. Upon careful examination, we instead identified a series of

benzalacetone-type metabolites that accumulated in transgenic *A. thaliana* or *N. benthamiana* plants expressing *PmSPS1* (Supplementary Note 2). We conclude that styrylpyrones generated by heterologously expressed *PmSPS1* are rapidly turned over by unknown enzymes present in *A. thaliana* or *N. benthamiana* to yield benzalacetones as breakdown products.

To probe the mechanistic basis for SPS neofunctionalization from the ancestral CHS, we solved the apo structure of *Pm*CHS and the holo structure of *Pm*SPS1 in complex with *p*-coumaroyl-CoA by X-ray crystallography (Supplementary Table 2). Both proteins are homodimers and share the canonical αβαβα thiolase fold typical for plant type III PKSs[21] (Fig. 2c). An evolutionary analysis using the Mixed Effects Model of Evolution (MEME)[30] on the ancestral branch of kava SPSs (Fig. 2a) detected six amino acid residues under episodic diversifying selection ($p<0.05$) (Supplementary Fig. 3). Three of these substitutions, S133C, T198N, and Q213L, are mapped to the enzyme active site (Fig. 2d). Notably, the T198N substitution was previously reported as one of the three mutations sufficient to convert CHS into a 2-pyrone synthase (2PS)[31]. We also observed an unusual insertion of threonine (T194) in a loop region abutting the active site, together with the S133C substitution, causing a reduction of the *Pm*SPS1 active site volume compared to *Pm*CHS in the direction along the polyketide chain elongation (Fig. 2e). In comparison, the previously reported *Rheum palmatum* benzalacetone synthase (*Rp*BAS) , which catalyzes only a single polyketide elongation[32], features a further shortened active site compared to *Pm*SPS1 (Fig. 2e). These observations therefore support the notion that subtle changes of active site volume and shape of type III PKSs dictate the iterative cycles of polyketide elongation and alternative cyclization mechanisms[33]. In addition, we observed two different conformations of T194 among the six protomers present in the

asymmetric unit, suggesting the dynamic nature of the underlying loop and its potential role in the catalytic cycle of *Pm*SPS1 (Fig. 2f). One of the protomers in our *Pm*SPS1-*p*-coumaroyl-CoA complex structure also captures a *p*-coumaroyl-monoketide intermediate covalently bound to the catalytic cysteine (Fig. 2f), providing a rare glimpse of the active site configuration after the starter substrate loading step of the PKS catalytic cycle[33]. Homology modeling and sequence analysis of *Pm*SPS2 suggest that its active site contains a similar set of amino acid substitutions as observed in *Pm*SPS1, except for the T194 insertion (Supplementary Fig. 3 and 4).

Although the *Pm*SPS1 and *Pm*SPS2 enzymes seem redundant in their *in vitro* activity, we hypothesized that they may exhibit differences in substrate preference. Using another *in vitro* assay with an equimolar mixture of six different CoA thioesters as the starter substrate input, we indeed observed substantial differences in the substrate preference of the three kava PKSs (Supplementary Note 3). Whereas *Pm*SPS1 preferentially produces 11-methoxy-bisnoryangonin and 4-hydroxy-6-styryl-2-pyrone derived from feruloyl-CoA and cinnamoyl-CoA, respectively, *Pm*SPS2 produces mostly dihydrobisnoryangonin and bisnoryangonin derived from *p*-dihydrocoumaroyl-CoA and *p*-coumaroyl-CoA, respectively. No tetraketide products were observed in the *Pm*SPS1 and *Pm*SPS2 assays. In contrast, *Pm*CHS produces both triketide and tetraketide products, preferentially derived from feruloyl-CoA and cinnamoyl-CoA. This is unusual as most of the previously characterized CHS orthologs accept *p*-coumaroyl-CoA as its preferred starter substrate[21].

Two kava OMTs catalyze regio-specific methylations of the styrylpyrone backbone. Structures of naturally occuring kavalactones implicate rich *O*-methylation reactions during their biosynthesis (Supplementary Fig. 1a). We hypothesized that kavalactone biosynthetic OMTs

might have been recruited from more conserved OMTs involved in plant phenylpropanoid metabolism, e.g., the caffeic acid *O*-methyltransferase (COMT) in lignin biosynthesis[34]. Combining phylogenomics and expression analysis, we identified six putative *OMT* candidate genes from the kava transcriptomes (Supplementary Table 1). Since kavalactones contain four possible *O*-methylation sites (−OH groups at the C_4, C_{10}, C_{11}, or C_{12} positions), we established a coupled enzymatic system using a 4CL enzyme cloned from kava (designated *Pm*4CL1) and *Pm*SPS1 to produce five styrylpyrones featuring different aromatic ring modification patterns derived from five hydroxycinnamic acids, namely cinnamic, *p*-coumaric, caffeic, ferulic, and umbellic acid (Supplementary Note 4). Subsequent OMT activity assays using recombinant kava OMTs revealed that two of the six kava OMT candidates showed activity against various styrylpyrones (Fig. 3a), and were thus named kava *O*-methyltransferases 1 and 2 (*Pm*KOMT1 and *Pm*KOMT2), respectively. *Pm*KOMT1 is one of the most highly expressed enzymes in the kava transcriptomes (Supplementary Table 1), and indeed is capable of methylating the −OH groups at the C_4, C_{11}, or C_{12} positions of the styrylpyrone backbone. On the other hand, *Pm*KOMT2 activity is specific to the −OH group attached at the C_{10} position (Fig. 3a). Transient co-expression of *PmKOMT1* together with *PmSPS1* or *PmSPS2* in *N. benthamiana* resulted in the production of yangonin, a kavalactone with two methoxy groups at the C_4 and C_{12} positions (Fig. 3b), supporting its enzymatic activities observed *in vitro*. This result also suggests that *O*-methylation protects styrylpyrones produced by kava SPSs from being catabolized in *N. benthamiana*.

Biosynthesis of 7,8-saturated kavalactones. We hypothesized that the 7,8-saturated kavalactones could be derived from dihydro-hydroxycinnamoyl-CoA precursors already

containing a reduced α−β bond. This hypothesis was provoked by previous reports that two type III PKSs in the genera *Pinus* and *Malus* can accept dihydro-*p*-coumaroyl-CoA as a starter substrate to produce 5-hydroxylunularic acid and phloretin, respectively[35,36]. To test this, we performed coupled *Pm*4CL1−*Pm*SPS1 *in vitro* enzyme assays using phloretic acid or phenylpropanoic acid as starter substrates, both of which contain a reduced α−β bond. These assays resulted in the production of the corresponding 7,8-saturated kavalactones, respectively, suggesting that *Pm*4CL1 and *Pm*SPS1 can utilize substrates derived from dihydro-hydroxycinnamic acids with a reduced α−β bond (Supplementary Fig. 5). Although dihydro-hydroxycinnamic acids are widespread in plants[37,38], the mechanism underlying the α−β olefin reduction of the hydroxycinnamic acid precursors is yet to be characterized.

A regio- and stereo-specific kavalactone reductase is required to produce kavain-like chiral kavalactones. The styrylpyrone backbones produced by SPSs from hydroxycinnamic acids contain two olefins (C_5=C_6 and C_7=C_8), whereas a number of naturally occuring kavalactones possess alkane at either position (e.g., kavain with C_5–C_6 or 7,8-dihydroyangonin with C_7–C_8) or both positions (e.g., 7,8-dihydrokavain) (Supplementary Fig. 1a). Whereas 7,8-dihydro-kavalactones can be derived from dihydro-hydroxycinnamic acids, reduction of the 5,6-olefin has to occur after the styrylpyrone backbone is formed by SPS. Moreover, 5,6-olefin reduction is stereo-specific, suggesting that this reaction is likely enzymatic. Using similar approaches as described above, we identified and characterized twelve kava reductases that are homologous to various known reductases in plant specialized metabolism (Supplementary Table 1). By transgenic co-expression of each of these reductase candidates together with *PmSPS1* and *PmKOMT1* in *N. benthamiana*, we identified a single

reductase candidate that produces metabolites with mass spectra corresponding to styrylpyrones with the 5,6-olefin reduced to alkane (Supplementary Note 5). We thus named this enzyme kavalactone reductase 1 (*Pm*KLR1). A combined enzyme assay using purified *Pm*4CL1, *Pm*SPS1, *Pm*KOMT1 and *Pm*KLR1 enzymes starting from cinnamic acid successfully produced kavain *in vitro* (Fig. 4a). Furthermore, we determined using chiral chromatography that *Pm*KLR1 produces only (*R*)-(+)-kavain, the same stereoisomer as present in kava root extract (Fig. 4b).

Since *Pm*KLR and *Pm*KOMT1 likely compete for common intermediates, we sought to determine the catalytic order of these two enzymes. We first conducted an isotopic tracing experiment by incubating a piece of live kava root with deuterium-labeled styrylpyrone precursors with or without 4-*O*-methylation (Supplementary Fig. 6a). This experiment shows that *Pm*KLR1 can only accept the unmethylated styrylpyrone precursor *in vivo*. Subsequent *in vitro* enzyme assays further corroborated this observation (Supplementary Fig. 6b). On the other hand, *Pm*KOMT1 is permissive towards styrylpyrone substrates regardless of whether the 5,6-olefin is reduced or not (Supplementary Fig. 6b). We therefore conclude that *Pm*KLR1 competes with *Pm*KOMT1 for the same styrylpyrone substrates, but is unable to reduce the 5,6-olefin once the 4-*O*-methylation has occurred, providing an explanation to the fact that only a fraction of the major kavalactones contain the reduced C_5−C_6 bond (Supplementary Fig. 1a).

Phylogenetic analysis of *Pm*KLR1 together with several related reductases from *A. thaliana* suggests that *Pm*KLR1 likely originated from an ancestral cinnamoyl-CoA reductase (Fig. 4c). To further dissect the structural basis for the reaction mechanism and regio- and stereo-selectivity of *Pm*KLR1, we solved its holo structure in complex with $NADP^+$ by X-ray

crystallography (Supplementary Table 2), and computationally docked a styrylpyrone substrate, 4-hydroxy-6-styryl-2-pyrone, into its active site (Fig. 4d). The structural analysis reveals the substrate-binding pocket geometry and several active-site residues key to the regio- and stereo-selective reduction of the styrylpyrone 5,6-olefin by *Pm*KLR1 (Fig. 4e). The docked styrylpyrone substrate is highly coordinated by several active-site-lining residues with a specific binding pose indicating that the initial hydride transfer from the nicotinamide group of NADPH is targeted to C_5. Lys169, which helps to coordinate the binding of NADPH by hydrogen bonding with its ribosyl −OH groups, likely also serves as a proton donor that relays its side-chain amine proton through Ser128 to C_6 of the styrylpyrone substrate exclusively from one side of the lactone ring (Fig. 4e). As the 5,6-olefin is reduced to alkane, a new chiral center at C_6 is formed. The intricate hydrogen bonds between the styrylpyrone 4-OH, Ser128 and Thr129 further corroborates our previous observations that 4-*O*-methylated styrylpyrone substrates would not bind properly and productively to the active site. Moreover, using custom-synthesized NADPH isotopic analogs with deuterium labeling at either *pro-S* or *pro-R* position of the nicotinamide group (Supplementary Fig. 7), we show unequivocally that the *pro-S*-hydride of NADPH is transferred to the styrylpyrone substrate by *Pm*KLR1 (Fig. 4f), which is in agreement with the relative configuration of the docked substrate and $NADP^+$ in *Pm*KLR1 active site (Fig. 4e). Based on these results, we outline the catalytic mechanism of *Pm*KLR1 (Fig. 4g).

A methylenedioxy-bridge-forming CYP719A26 enzyme produces methysticin-like kavalactones. Methysticin-like kavalactones, which carry a methylenedioxy bridge at the C_{11}−C_{12} position (Supplementary Fig. 1a), exhibit strong modulatory effects on human liver cytochromes P450, and thus contribute to the complex pharmacology of kava[39]. The CYP719

family of plant P450 enzymes is known to catalyze the formation of the methylenedioxy bridge moiety from vicinal methoxyl and hydroxyl groups, e.g., in the biosynthesis of berberine and etoposide[40,41]. We identified a single gene of the CYP719 family in kava, *PmCYP719A26* (Fig. 4c), and hypothesized that it may catalyze the formation of methylenedioxy bridge using ferulic acid-derived styrylpyrone precursors carrying vicinal 11-methoxyl and 12-hydroxyl groups (Fig. 4d). Indeed, co-expression of *PmSPS1*, *PmKOMT1* and *PmCYP719A26* in *N. benthamiana* resulted in the transgenic production of a compound with the molecular mass and retention time identical to that of methysticin (Fig. 4e). Based on an MS/MS fragmentation pattern comparison with a methysticin standard, we tentatively assigned this compound as 5,6-dehydro-7,8-dihydromethysticin (Supplementary Note 6). We therefore designated *Pm*CYP719A26 as methysticin synthase 1 (*Pm*MTS1).

Flavokavain biosynthesis is facilitated by the same KOMT enzymes that produce kavalactones. Through OMT activity screen, we found that *Pm*KOMT1 and *Pm*KOMT2 also exhibit *O*-methylation activities against naringenin chalcone produced by CHS (Supplementary Note 7). Using enzyme assays and heterologous expression in *N. benthamiana*, we confirmed that *Pm*KOMT1 methylates the hydroxy groups at the C_4 and $C_{4'}$ positions on the chalcone scaffold, whereas *Pm*KOMT2 methylates the hydroxy group at the $C_{2'}$ position (Supplementary Fig. 8). We thus established that the three known flavokavains (Supplementary Fig. 1b) can be produced by the combinatorial activities of *Pm*CHS, *Pm*KOMT1 and *Pm*KOMT2.

Feasibility of heterologous production of kavalactones in microorganisms. As a proof of concept, we examined the feasibility of kavalactone production in microbial hosts through the means of metabolic engineering. Using an operon-like construct co-expressing *Pm4CL1* together

with *PmSPS1* or *PmSPS2* under the constitutive *pGAP* promoter, we could reconstitute the biosynthesis of the kavalactone precursor, bisnoryangonin, from supplemented *p*-coumaric acid in *E.coli* under shake flask condition (Supplementary Fig. 9a). Similarly, plasmid-based expression of the same combination of enzymes under the constitutive *pTEF* promoter in the yeast *Saccharomyces cerevisiae* also resulted in bisnoryangonin production from supplemented *p*-coumaric acid under shake flask condition (Supplementary Fig. 9b). Future efforts in combining kavalactone biosynthetic genes with previously developed background bacterial or yeast strains overproducing *p*-coumaroyl-CoA can potentially further increase the titer of kavalactone production in these hosts[42].

Discussion

Plant type III PKSs were descended from the ancestral 3-ketoacyl-acyl carrier protein synthase III (KAS III) enzymes involved in fatty acid metabolism, and have radiated during land plant evolution to give rise to a diverse class of specialized metabolic enzymes responsible for producing a battery of important polyketide scaffolds, including chalcones, stilbenes (e.g., resveratrol), benzalacetones, and others[21]. The remarkably diverse polyketide natural products subsequently stemmed from these scaffolds contribute to the fitness of their plant hosts in challenging and ever-changing terrestrial environments. Although plant natural products that carry the styrylpyrone structural motif have been known for decades[43], our study presents the first molecular cloning and functional characterization of two dedicated SPSs from a styrylpyrone-producing plant. It is worth noting that significant SPS activity was observed in *Pm*CHS, which represents the extant version of the evolutionary progenitor that gave rise to

*Pm*SPS1 and *Pm*SPS2. The latent SPS activity of *Pm*CHS likely caused the initial serendipitous accumulation of styrylpyrones at appreciable amount in the ancestors of kava, which kick-started the natural selection process that further shaped the evolutionary trajectory of these newly emerged metabolic traits[44,45]. Mainly through the mechanism of gene duplication followed by neofunctionalization, two dedicated SPSs as well as additional kavalactone-biosynthetic tailoring enzymes were derived from existing progenitor enzymes of various families and assembled into a new metabolic network to produce the full repertoire of structurally diverse kavalactones as we observe in the extant kava cultivars today[46]. Although the biosynthetic origin of kavalactones as delineated in this study is unique to kava, it epitomizes a common underlying evolutionary process that drives the rapid expansion of chemodiversity in the plant kingdom[46].

Given the psychoactive properties of kavalactones, the accumulation of specific kavalactone structural variants and/or combination of them likely confer protection to kava against herbivorous animals. Consistent with this notion, folk stories about small rodents paralyzed by chewing on the roots of kava plants are widespread in the Pacific islands. It is also likely that the thousands of years of consumption and cultivation of kava plants by Polynesian islanders have led to human artificial selection of kava plants that produce higher content and specific composition of kavalactones.

In animals, kavalactones display complex pharmacology, including modulation of $GABA_A$ and cannabinoid (CB_1) receptors, blockade of voltage-gated sodium ion channels, reduced release and reuptake of neurotransmitters, and interactions with monoamine oxidase B[11,12]. In particular, kavain was shown to act as a positive allosteric modulator of the $GABA_A$ receptor through a binding site different from that of benzodiazepine psychiatric drugs[17].

The capability to produce structurally diverse kavalactones and their derivatives by metabolic engineering of the biosynthetic enzymes described herein will help delineate the structure-function relationships of kavalactones and potentially lead to novel biomedical applications. In light of several major challenges facing modern society, such as anxiety, opioid crisis, and the scarcity of effective treatments for psychiatric disorders, kavalactones and their derivatives present a promising class of psychoactive molecules to potentially answer these unmet needs.

Methods

Chemicals and reagents. Kavalactone standards (yangonin, methysticin, desmethoxyyangonin, 7,8-dihydrokavain, and 7,8-dihydromethysticin) were obtained from AvaChem Scientific. Cinnamoyl-CoA, *p*-coumaroyl-CoA, *p*-dihydrocoumaroyl-CoA, caffeoyl-CoA, feruloyl-CoA, and sinapoyl-CoA were obtained from MicroCombiChem. Naringenin was obtained from AK Scientific. Isopropanol-d_8 was obtained from Cambridge Isotope Laboratories. (±)-Kavain, malonyl-CoA, *trans*-cinnamic acid-d_6, *D*-glucose-1-d_1, and other cofactors and reagents were obtained from Sigma-Aldrich.

RNA extraction and cDNA template preparation. Total RNA was extracted separately from kava root and leaf tissue using the RNeasy Plant Mini Kit (QIAGEN). First-strand cDNAs were synthesized by RT–PCR from the total RNA samples as templates using the SuperScript III First-Strand Synthesis System with the oligo$(dT)_{20}$ primer (Thermo Fisher Scientific).

Transcriptome sequencing and assembly. The transcriptome library preparation and sequencing were performed at the Beijing Genomics Institute using the standard BGISEQ-500 RNA sample preparation protocol. The libraries were sequenced as 50×50 bp (PE-50) reads on the BGISEQ-500 platform. Sequence reads (FASTQ files) from leaf and root samples were merged into a single set of reads, trimmed for sequencing adaptors using Trimmomatic[47], and assembled into a *de novo* transcriptome using Trinity[48], resulting in 285,753 transcript sequences. Within those, we identified 302,366 putative open reading frames using Transdecoder[49]. We evaluated the transcriptome as 78.6% complete by the metric of Benchmarking Universal Single-Copy Orthologs (BUSCO)[50]. Gene expression statistics were determined using RSEM[51]. Transcripts and predicted protein sequences were annotated with transcript-per-million (TPM) values and closest BLAST hits from the UniProtKB/Swiss-Prot database using in-house scripts. Transcriptome mining was performed on a local BLAST server[52]. In addition, two kava root RNAseq datasets previously published by the PhytoMetaSyn Project[53] (NCBI SRA accessions SRX202785 and SRX202184) were assembled and used as complementary sources of candidate enzyme sequences. Existing RNAseq datasets of *P. auritum, P. betle*, and *P. nigrum* (NCBI SRA accessions ERX2099199, SRX691517, and SRX890122) were also assembled into *de novo* transcriptomes to obtain their corresponding CHS sequences.

Sequence alignment and phylogenetic analyses. Sequence alignments were performed using the MUSCLE[54] algorithm in MEGA7[55]. Evolutionary histories were inferred by using the

Maximum Likelihood method based on the JTT matrix-based model[56]. Bootstrap values were calculated using 1,000 replicates. All phylogenetic analyses were conducted in MEGA7[55].

Evolutionary analysis. The Mixed Effects Model of Evolution (MEME) analysis[30] was performed using the Hypothesis Testing using Phylogenies (HyPhy) package[57], starting from codon-aligned CHS nucleotide sequences and the corresponding phylogenetic tree. The p-value threshold was set to 0.05.

Cloning of candidate genes from cDNA. Phusion High-Fidelity DNA Polymerase (Thermo Fisher Scientific) was used for PCR amplifications from kava cDNA. Gibson assembly was used to clone the amplified genes into target vectors[58]. Restriction enzymes and Gibson assembly reagents were purchased from New England Biolabs. Oligonucleotide primers were purchased from Integrated DNA Technologies. All primers used for cloning are listed in Supplementary Table 4.

Protein expression and purification. Candidate genes were cloned into pHis8-4, a bacterial expression vector containing an N-terminal 8×His tag followed by a tobacco etch virus (TEV) cleavage site for recombinant protein production in *E. coli*. Proteins were expressed in the BL21(DE3) *E. coli* strain cultivated in terrific broth (TB) and induced with 0.1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) overnight at 18 °C. For the OMT enzyme screen, crude protein extracts were prepared from *E. coli* cultures using Bacterial Protein Extraction Reagent (B-PER, Thermo Fisher Scientific). For protein purification, *E. coli* cells were harvested by

centrifugation, resuspended in 150 mL lysis buffer (50 mM Tris pH 8.0, 500 mM NaCl, 30 mM imidazole, 5 mM DTT), and lysed with five passes through an M-110L microfluidizer (Microfluidics). The resulting crude protein lysate was clarified by centrifugation (19,000 g, 1 h) prior to QIAGEN nickel–nitrilotriacetic acid (Ni–NTA) gravity flow chromatographic purification. After loading the clarified lysate, the Ni–NTA resin was washed with 20 column volumes of lysis buffer and eluted with 1 column volume of elution buffer (50 mM Tris pH 8.0, 500 mM NaCl, 300 mM imidazole, 5 mM DTT). 1 mg of His-tagged TEV protease[59] was added to the eluted protein, followed by dialysis at 4 °C for 16 h in dialysis buffer (50 mM Tris pH 8.0, 500 mM NaCl, 5 mM DTT). After dialysis, protein solution was passed through Ni–NTA resin to remove uncleaved protein and His-tagged TEV. The recombinant proteins were further purified by gel filtration on a fast protein liquid chromatography (FPLC) system (GE Healthcare Life Sciences). The principal peaks were collected, verified by SDS–PAGE, and dialyzed into a storage buffer (12.5 mM Tris pH 8.0, 50 mM NaCl, 5 mM DTT). Finally, proteins were concentrated to >10 mg/mL using Amicon Ultra-15 Centrifugal Filters (Millipore). The *Pm*SPS1, *Pm*SPS2 and *Pm*CHS proteins purified as homodimers, while *Pm*4CL1, *Pm*KOMT1, *Pm*KOMT2 and *Pm*KLR1 purified as monomers.

***In vitro* enzyme assays.** Enzyme assays were performed in 50 mM potassium phosphate buffer, pH 7.6 containing 5 mM $MgCl_2$, 3 mM adenosine triphosphate (ATP), 1 mM coenzyme A, 3 mM malonyl-CoA, and 0.25 mM initial substrate (typically, cinnamic acid or *p*-coumaric acid). Recombinant enzymes were added to final concentration of 10 μg/mL. For OMT assays, additional 3 mM S-adenosyl-methionine was added. For reductase assays, additional 5 mM

NADPH was added. Reactions were incubated at 30 °C until completion (typically overnight), followed by the addition of methanol to 50% final concentration. Samples were centrifuged (13,000 g, 20 min) and supernatants were collected for LC-MS analyses.

Stereo-specific deuterium labeling using NADPH isotopic analogs. Enzyme assays using site specifically deuterium-labeled NADPH isotopic analogs were performed as coupled reactions where additional components were added to produce the desired NADPH isotopic analog. (*R*)-NADPD was produced by adding 1 mM $NADP^+$, 10 mM isopropanol-d_8, 1% (v/v) alcohol dehydrogenase (Sigma-Aldrich cat. no. 49641), and 1 mM semicarbazide. (*S*)-NADPD was produced by adding 1 mM $NADP^+$, 2 mM *D*-glucose-1-d_1, 5 mM $MgCl_2$, 4 mM ATP, 1 unit hexokinase and 1 unit glucose-6-phosphate dehydrogenase.

Transgenic Arabidopsis. Kava *PKS* genes were cloned into the pEarleyGate 100 vector[60] and transformed into *Agrobacterium tumefaciens* strain GV3130. The *A. thaliana tt4-2* mutant[28] was then transformed using the Agrobacterium-mediated floral dip method[61]. Transformants were selected by spraying with Finale (contains 11.33% glufosinate ammonium; Bayer) diluted 1:500 in water. Selection was repeated in each subsequent generation and experiments were performed in the T2 or T3 generations, as indicated in Supplementary Fig. 2.

qRT-PCR analysis. qRT-PCR reactions were performed on a QuantStudio 6 Flex system (Thermo Fisher Scientific) using SYBR Green Master Mix (Thermo Fisher Scientific) and

primers listed in Supplementary Table 4. Gene expression Ct values were normalized using the reference gene *At1g13320*[62].

Plant metabolite extraction. Approximately 100 mg of plant leaf tissue was dissected, transferred into grinding tubes containing ~15 zirconia/silica disruption beads (2 mm diameter; Research Products International), and snap-frozen in liquid nitrogen. The frozen samples were homogenized twice on a TissueLyser II (QIAGEN). Metabolites were extracted using 5–10 volumes (w/v) of 50% methanol at 55 °C for 1 h. Extracts were centrifuged twice (13,000 g, 20 min) and supernatants were collected for LC-MS analysis.

LC-MS analysis. LC was conducted on a Dionex UltiMate 3000 UHPLC system (Thermo Fisher Scientific), using water with 0.1% formic acid as solvent A and acetonitrile with 0.1% formic acid as solvent B. Reverse phase separation of analytes was performed on a Kinetex C18 column, 150 × 3 mm, 2.6 μm particle size (Phenomenex). Chiral chromatography was performed on a Lux i-Amylose-1 column, 250 × 4.6 mm, 5 μm particle size (Phenomenex). The column oven was held at 30 °C. Most injections were eluted with a gradient of 5–60% B for 9 min, 95% B for 3 min, and 5% B for 3 min, with a flow rate of 0.7 mL/min. The initial analysis of *Piper* species leaf extracts was performed on a gradient of 5–80% B for 40 min, 95% B for 4 min, and 5% B for 5 min, with a flow rate of 0.8 mL/min. Chiral chromatography was performed on a gradient of 5–95% B for 30 min, 95% B for 5 min, and 5% B for 10 min, with a flow rate of 1 mL/min. MS analyses for the OMT enzyme screen and for chiral chromatography were performed on a TSQ Quantum Access Max mass spectrometer (Thermo Fisher Scientific)

operated in positive ionization mode with full scan range of 100–400 m/z. Other MS analyses and plant metabolic profiling were performed on a high-resolution Q-Exactive benchtop Orbitrap mass spectrometer (Thermo Fisher Scientific) operated in positive ionization mode with full scan range of 100–1000 m/z and top 5 data-dependent MS/MS scans. Raw LC-MS data were analyzed using XCalibur (Thermo Fisher Scientific), MZmine 2[63], and MetaboAnalyst[64]. The SIRIUS tool was used to interpret MS/MS spectra[65,66]. Identified metabolites are listed in Supplementary Table 3.

X-ray crystallography. The purified *Pm*SPS1 protein (17.1 mg/mL) was incubated with 4 mM *p*-coumaroyl-CoA for 1 h prior to setting crystal trays. Crystals of *Pm*SPS1 were obtained after 4 days at 21 °C in hanging drops containing 0.8 μL of protein solution and 0.8 μL of reservoir solution (50 mM HEPES pH 7.5, 10% w/v PEG 8000, 4% v/v ethylene glycol). Crystals were frozen in reservoir solution with ethylene glycol concentration increased to 15% (v/v).
SPS1 X-ray diffraction data were collected on the 24-ID-C beam line equipped with a Pilatus-6MF pixel array detector (Advanced Photon Source, Argonne National Laboratory). Crystals of the purified *Pm*CHS protein (17.6 mg/mL) were obtained from a screening plate after 4 days at 4°C in sitting drops containing 150 nL protein solution and 150 nL of reservoir solution (100 mM HEPES pH 7.0, 10% w/v PEG 6000). Crystals of the purified *Pm*KLR1 protein (14 mg/mL) were obtained after 4 days at 21 °C in hanging drops containing 0.8 μL of protein solution (including 1 mM $NADP^+$) and 0.8 μL of reservoir solution (100 mM HEPES pH 7.5, 25% w/v PEG 3350, 200 mM NH_4Ac). Crystals were frozen in reservoir solution with PEH 3350 concentration increased to 34% (w/v). *Pm*CHS and *Pm*KLR1 X-ray diffraction data were

collected on the 24-ID-E beam line equipped with an Eiger-16M pixel array detector (Advanced Photon Source, Argonne National Laboratory). Diffraction intensities were indexed and integrated with iMosflm[67] and scaled with SCALA[68]. Initial phases were determined by molecular replacement using Phaser[69] in PHENIX[70], using search models generated from protein sequences on the Phyre[2] server[71]. Subsequent structural building and refinements were conducted in PHENIX. Coot was used for graphical map inspection and manual tuning of atomic models[72]. Root mean square deviation (RMSD) of *Pm*SPS1 and *Pm*CHS structures was calculated using the Structure comparison function in PHENIX. Crystallography statistics are listed in Supplementary Table 2.

Protein structure modeling, docking, and rendering. The *Pm*SPS2 structure model was generated on the Phyre[2] server using the *Pm*SPS1 crystal structure as a modeling template[71]. The active-site pocket in the *Pm*SPS1 structure was determined using KVFinder[73]. Ligand docking calculations were performed with AutoDock Vina 1.1.2[74]. Molecular graphics were rendered with PyMOL 2.0.7 (Schrödinger).

Transient expression in *N. benthamiana*. Target kava genes were cloned into the pEAQ-HT vector[75,76] and transformed into the *A. tumefaciens* strain LBA4404[77]. Bacteria were cultivated at 30 °C to OD_{600} of 1.5 in 50 mL of YM medium (0.4 g/L yeast extract, 10 g/l mannitol, 0.1 g/L NaCl, 0.2 g/L $MgSO_4 \cdot 7H_2O$, 0.5 g/L $K_2HPO_4 \cdot 3H_2O$), washed with 0.5×PBS buffer (68 mM NaCl, 1.4 mM KCl, 5 mM Na_2HPO_4, 0.9 mM KH_2PO_4), and resuspended in 0.5×PBS buffer to OD_{600} of 0.8. For co-expressing multiple genes, individual *A. tumefaciens* cultures containing the

transgene constructs were grown, pelleted, and washed separately, and finally mixed to reach the final OD_{600} of 0.8 for each culture. 1 mL of the final culture was used to infiltrate the underside of 5-6 weeks old *N. benthamiana* leaves. Leaves were harvested 5 days post infiltration for metabolite extraction.

Reconstitution of kavalaction biosynthesis in *E. coli.* *Pm4CL1*, *PmSPS1*, and *PmSPS2* were cloned into a custom-made, operon-like expression vector pJKW1565. The genes were preceded by a constitutively expressed *pGAP* promoter of the *E. coli* glyceraldehyde-3-phosphate dehydrogenase (GAPDH)[78], and each gene was preceded by a 23-bp ribosome binding site (RBS) sequence[79]. *tSPY* was used as a terminator[80]. The constructed plasmids (pJKW1574 containing the *Pm4CL1* + *PmSPS1* operon and pJKW1658 containing the *Pm4CL1* + *PmSPS2* operon) were transformed into the BW27784 *E. coli* strain[81] and cultivated at 30 °C overnight in TB medium supplied with 1 mM *p*-coumaric acid, followed by metabolite extraction from the culture fluid. An aliquot of the bacterial culture was mixed with an equal volume of methanol (50% final methanol concentration), centrifuged twice (13,000 g, 20 min) and supernatants were collected for LC-MS analysis. Since bisnoryangonin standard was not available, its concentrations were estimated using naringenin as the calibration-curve standard.

Reconstitution of kavalaction biosynthesis in *S. cerevisiae*. The p42xTEF 2μ plasmids with various auxotrophic growth markers for constitutive expression in yeast were used[82]. The *Pm4CL1* gene was cloned into p425TEF (plasmid pJKW1413), while *PmSPS1* and *PmSPS2* were cloned into p426TEF (plasmids pJKW1538 and pJKW1547, respectively). The constructed

plasmids were transformed into the BY4743 yeast strain[83]. Yeast cells were cultivated at 30 °C for 2 days in Yeast Nitrogen Base (YNB) + histidine + dextrose medium supplied with 2 mM *p*-coumaric acid, followed by metabolite extraction from the culture fluid. An aliquot of the yeast culture was mixed with an equal volume of methanol (50% final methanol concentration), centrifuged twice (13,000 g, 20 min) and supernatants were collected for LC-MS analysis. Since bisnoryangonin standard was not available, its concentrations were estimated using naringenin as the calibration-curve standard.

References

1. Li, F.-S. & Weng, J.-K. Demystifying traditional herbal medicine with modern approach. *Nature Plants* **3**, 17109 (2017).
2. Parmar, V. S. *et al.* Phytochemistry of the genus Piper. *Phytochemistry* **46**, 597–673 (1997).
3. Mehmood, M. H. & Gilani, A. H. Pharmacological basis for the medicinal use of black pepper and piperine in gastrointestinal disorders. *J. Med. Food* **13**, 1086–1096 (2010).
4. Bharadwaj, U. *et al.* Drug-repositioning screening identified piperlongumine as a direct STAT3 inhibitor with potent activity against breast cancer. *Oncogene* **34**, 1341–1353 (2015).
5. Raj, L. *et al.* Selective killing of cancer cells by a small molecule targeting the stress response to ROS. *Nature* **475**, 231–234 (2011).
6. Chu, N. S. Effects of Betel chewing on the central and autonomic nervous systems. *J. Biomed. Sci.* **8**, 229–236 (2001).
7. Amonkar, A. J., Nagabhushan, M., D'Souza, A. V. & Bhide, S. V. Hydroxychavicol: a new phenolic antimutagen from betel leaf. *Food Chem. Toxicol.* **24**, 1321–1324 (1986).
8. Jaramillo, M. A. & Manos, P. S. Phylogeny and patterns of floral diversity in the genus Piper (Piperaceae). *Am. J. Bot.* **88**, 706–716 (2001).
9. Lebot, V. & Lèvesque, J. The origin and distribution of kava (Piper methysticum Forst. f., Piperaceae): a phytochemical approach. *Allertonia* **5**, 223–281 (1989).
10. Singh, Y. N. Kava: an overview. *J. Ethnopharmacol.* **37**, 13–45 (1992).
11. Cairney, S., Maruff, P. & Clough, A. R. The neurobehavioural effects of kava. *Aust. N. Z. J. Psychiatry* **36**, 657–662 (2002).
12. Sarris, J., LaPorte, E. & Schweitzer, I. Kava: a comprehensive review of efficacy, safety, and psychopharmacology. *Aust. N. Z. J. Psychiatry* **45**, 27–35 (2011).

13. Lasme, P., Davrieux, F., Montet, D. & Lebot, V. Quantification of kavalactones and determination of kava (Piper methysticum) chemotypes using near-infrared reflectance spectroscopy for quality control in vanuatu. *J. Agric. Food Chem.* **56**, 4976–4981 (2008).
14. World Health Organization. *Kava: a review of the safety of traditional and recreational beverage consumption*. (World Health Organization, 2016).
15. Stressed New Yorkers Take to Kava, 'Nature's Xanax'. (2017). Available at: https://www.nytimes.com/2017/12/21/nyregion/stressed-new-yorkers-take-to-kava-natures-xanax.html. (Accessed: 21st July 2018)
16. Kava: The NFL's newest and safest painkiller. *ESPN.com* (2017). Available at: http://www.espn.com/blog/nflnation/post/_/id/256632. (Accessed: 21st July 2018)
17. Chua, H. C. *et al.* Kavain, the Major Constituent of the Anxiolytic Kava Extract, Potentiates GABAA Receptors: Functional Characteristics and Molecular Mechanism. *PLOS ONE* **11**, e0157700 (2016).
18. Jamieson, D. D. & Duffield, P. H. The antinociceptive actions of kava components in mice. *Clin. Exp. Pharmacol. Physiol.* **17**, 495–507 (1990).
19. Zi, X. & Simoneau, A. R. Flavokawain A, a novel chalcone from kava extract, induces apoptosis in bladder cancer cells by involvement of Bax protein-dependent and mitochondria-dependent apoptotic pathway and suppresses tumor growth in mice. *Cancer Res.* **65**, 3479–3486 (2005).
20. Zhou, P. *et al.* Flavokawain B, the hepatotoxic constituent from kava root, induces GSH-sensitive oxidative stress through modulation of IKK/NF-kappaB and MAPK signaling pathways. *FASEB J.* **24**, 4722–4732 (2010).
21. Abe, I. & Morita, H. Structure and function of the chalcone synthase superfamily of plant type III polyketide synthases. *Nat. Prod. Rep.* **27**, 809–838 (2010).
22. Lebot, V., Aradhya, M. K. & Manshardt, R. M. Geographic survey of genetic variation in kava

(Piper methysticum Forst. f. and P. wichmannii C. DC.). *Pacific Science* **45**, 169–185 (1991).

23. Torrens-Spence, M. P., Fallon, T. R. & Weng, J. K. A Workflow for Studying Specialized Metabolism in Nonmodel Eukaryotic Organisms. in *Methods in Enzymology* 69–97 (2016).
24. Owen, C., Patron, N. J., Huang, A. & Osbourn, A. Harnessing plant metabolic diversity. *Curr. Opin. Chem. Biol.* **40**, 24–30 (2017).
25. Tatsis, E. C. & O'Connor, S. E. New developments in engineering plant metabolic pathways. *Curr. Opin. Biotechnol.* **42**, 126–132 (2016).
26. Akiyama, T., Shibuya, M., Liu, H. M. & Ebizuka, Y. p‐Coumaroyltriacetic acid synthase, a new homologue of chalcone synthase, from Hydrangea macrophylla var. thunbergii. *FEBS J.* **263**, 834–839 (1999).
27. Abe, I., Sano, Y., Takahashi, Y. & Noguchi, H. Site-directed Mutagenesis of Benzalacetone Synthase: THE ROLE OF PHE215 IN PLANT TYPE III POLYKETIDE SYNTHASES. *J. Biol. Chem.* **278**, 25218–25226 (2003).
28. Burbulis, I. E., Iacobucci, M. & Shirley, B. W. A null mutation in the first enzyme of flavonoid biosynthesis does not affect male fertility in Arabidopsis. *Plant Cell* **8**, 1013–1025 (1996).
29. Sainsbury, F. & Lomonossoff, G. P. Transient expressions of synthetic biology in plants. *Curr. Opin. Plant Biol.* **19**, 1–7 (2014).
30. Murrell, B. *et al.* Detecting individual sites subject to episodic diversifying selection. *PLoS Genet.* **8**, e1002764 (2012).
31. Jez, J. M. *et al.* Structural control of polyketide formation in plant-specific polyketide synthases. *Chem. Biol.* **7**, 919–930 (2000).
32. Morita, H. *et al.* A structure-based mechanism for benzalacetone synthase from Rheum palmatum. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 669–673 (2010).
33. Austin, M. B. & Noel, J. P. The chalcone synthase superfamily of type III polyketide synthases. *Nat.*

Prod. Rep. **20**, 79–110 (2003).

34. Moinuddin, S. G. A. *et al.* Insights into lignin primary structure and deconstruction from Arabidopsis thaliana COMT (caffeic acid O-methyl transferase) mutant Atomt1. *Org. Biomol. Chem.* **8**, 3928–3946 (2010).
35. Eckermann, C. *et al.* Stilbenecarboxylate biosynthesis: a new function in the family of chalcone synthase-related proteins. *Phytochemistry* **62**, 271–286 (2003).
36. Gosch, C., Halbwirth, H. & Stich, K. Phloridzin: biosynthesis, distribution and physiological relevance in plants. *Phytochemistry* **71**, 838–843 (2010).
37. Narváez-Cuenca, C.-E., Vincken, J.-P. & Gruppen, H. Identification and quantification of (dihydro) hydroxycinnamic acids and their conjugates in potato by UHPLC–DAD–ESI-MSn. *Food Chem.* **130**, 730–738 (2012).
38. Harbaum, B. *et al.* Identification of Flavonoids and Hydroxycinnamic Acids in Pak Choi Varieties (Brassica campestrisL. ssp.chinensisvar.communis) by HPLC–ESI-MSnand NMR and Their Quantification by HPLC–DAD. *J. Agric. Food Chem.* **55**, 8251–8260 (2007).
39. Li, Y. *et al.* Methysticin and 7,8-dihydromethysticin are two major kavalactones in kava extract to induce CYP1A1. *Toxicol. Sci.* **124**, 388–399 (2011).
40. Ikezawa, N. *et al.* 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).
41. Lau, W. & Sattely, E. S. Six enzymes from mayapple that complete the biosynthetic pathway to the etoposide aglycone. *Science* **349**, 1224–1228 (2015).
42. Chemler, J. A. & Koffas, M. A. G. Metabolic engineering for plant natural product biosynthesis in microbes. *Curr. Opin. Biotechnol.* **19**, 597–605 (2008).
43. Beckert, C. *et al.* Styrylpyrone biosynthesis in Equisetum arvense. *Phytochemistry* **44**, 275–283

(1997).

44. Anarat-Cappillino, G. & Sattely, E. S. The chemical logic of plant natural product biosynthesis. *Curr. Opin. Plant Biol.* **19**, 51–58 (2014).

45. Weng, J.-K. & Noel, J. P. The remarkable pliability and promiscuity of specialized metabolism. *Cold Spring Harb. Symp. Quant. Biol.* **77**, 309–320 (2012).

46. Weng, J.-K., Philippe, R. N. & Noel, J. P. The rise of chemodiversity in plants. *Science* **336**, 1667–1670 (2012).

47. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).

48. Grabherr, M. G. *et al.* Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–652 (2011).

49. Haas, B. J. *et al.* De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* **8**, 1494–1512 (2013).

50. Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015).

51. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).

52. Priyam, A. *et al.* Sequenceserver: a modern graphical user interface for custom BLAST databases. *bioRxiv* 033142 (2015). doi:10.1101/033142

53. Xiao, M. *et al.* Transcriptome analysis based on next-generation sequencing of non-model plants producing specialized metabolites of biotechnological interest. *J. Biotechnol.* **166**, 122–134 (2013).

54. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).

55. Kumar, S., Stecher, G. & Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.* **33**, 1870–1874 (2016).

56. Jones, D. T., Taylor, W. R. & Thornton, J. M. The rapid generation of mutation data matrices from protein sequences. *Comput. Appl. Biosci.* **8**, 275–282 (1992).

57. Pond, S. L. K., Frost, S. D. W. & Muse, S. V. HyPhy: hypothesis testing using phylogenies. *Bioinformatics* **21**, 676–679 (2005).

58. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345 (2009).

59. Tropea, J. E., Cherry, S. & Waugh, D. S. Expression and purification of soluble His(6)-tagged TEV protease. *Methods Mol. Biol.* **498**, 297–307 (2009).

60. Earley, K. W. *et al.* Gateway-compatible vectors for plant functional genomics and proteomics. *Plant J.* **45**, 616–629 (2006).

61. Clough, S. J. & Bent, A. F. Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. *Plant J.* **16**, 735–743 (1998).

62. Czechowski, T., Stitt, M., Altmann, T., Udvardi, M. K. & Scheible, W.-R. Genome-wide identification and testing of superior reference genes for transcript normalization in Arabidopsis. *Plant Physiol.* **139**, 5–17 (2005).

63. Pluskal, T., Castillo, S., Villar-Briones, A. & Oresic, M. MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics* **11**, 395 (2010).

64. Xia, J., Sinelnikov, I. V., Han, B. & Wishart, D. S. MetaboAnalyst 3.0—making metabolomics more meaningful. *Nucleic Acids Res.* **43**, W251–W257 (2015).

65. Böcker, S., Letzel, M. C., Lipták, Z. & Pervukhin, A. SIRIUS: decomposing isotope patterns for metabolite identification. *Bioinformatics* **25**, 218–224 (2009).

66. Dührkop, K., Shen, H., Meusel, M., Rousu, J. & Böcker, S. Searching molecular structure databases with tandem mass spectra using CSI:FingerID. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 12580–12585 (2015).

67. Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. W. \it iMOSFLM: a new graphical interface for diffraction-image processing with \it MOSFLM. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 271–281 (2011).

68. Evans, P. Scaling and assessment of data quality. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 72–82 (2006).

69. McCoy, A. J. Solving structures of protein complexes by molecular replacement with Phaser. *Acta Crystallogr. D Biol. Crystallogr.* **63**, 32–41 (2007).

70. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010).

71. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845–858 (2015).

72. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).

73. Oliveira, S. H. P. *et al.* KVFinder: steered identification of protein cavities as a PyMOL plugin. *BMC Bioinformatics* **15**, 197 (2014).

74. Trott, O. & Olson, A. J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **31**, 455–461 (2010).

75. Sainsbury, F., Thuenemann, E. C. & Lomonossoff, G. P. pEAQ: versatile expression vectors for easy and quick transient expression of heterologous proteins in plants. *Plant Biotechnol. J.* **7**, 682–693 (2009).

76. Peyret, H. & Lomonossoff, G. P. The pEAQ vector series: the easy and quick way to produce

recombinant proteins in plants. *Plant Mol. Biol.* **83**, 51–58 (2013).

77. Hoekema, A., Hirsch, P. R., Hooykaas, P. J. J. & Schilperoort, R. A. A binary plant vector strategy based on separation of vir- and T-region of the Agrobacterium tumefaciens Ti-plasmid. *Nature* **303**, 179 (1983).
78. Charpentier, B., Bardey, V., Robas, N. & Branlant, C. The EIIGlc protein is involved in glucose-mediated activation of Escherichia coli gapA and gapB-pgk transcription. *J. Bacteriol.* **180**, 6476–6483 (1998).
79. Olins, P. O. & Rangwala, S. H. A novel sequence element derived from bacteriophage T7 mRNA acts as an enhancer of translation of the lacZ gene in Escherichia coli. *J. Biol. Chem.* **264**, 16973–16976 (1989).
80. Chen, Y.-J. *et al.* Characterization of 582 natural and synthetic terminators and quantification of their design constraints. *Nat. Methods* **10**, 659–664 (2013).
81. Khlebnikov, A., Datsenko, K. A., Skaug, T., Wanner, B. L. & Keasling, J. D. Homogeneous expression of the P(BAD) promoter in Escherichia coli by constitutive expression of the low-affinity high-capacity AraE transporter. *Microbiology* **147**, 3241–3247 (2001).
82. Mumberg, D., Müller, R. & Funk, M. Yeast vectors for the controlled expression of heterologous proteins in different genetic backgrounds. *Gene* **156**, 119–122 (1995).
83. Brachmann, C. B. *et al.* Designer deletion strains derived from Saccharomyces cerevisiae S288C: a useful set of strains and plasmids for PCR-mediated gene disruption and other applications. *Yeast* **14**, 115–132 (1998).
84. Ferrer, J. L., Jez, J. M., Bowman, M. E., Dixon, R. A. & Noel, J. P. Structure of chalcone synthase and the molecular basis of plant polyketide biosynthesis. *Nat. Struct. Biol.* **6**, 775–784 (1999).

Acknowledgments

This work was supported by grants from the Smith Family Foundation, Edward N. and Della L. Thome Memorial Foundation, the Family Larsson-Rosenquist Foundation, and the National Science Foundation (CHE-1709616). T.P. is a Simons Foundation Fellow of the Helen Hay Whitney Foundation. J.K.W is supported by the Beckman Young Investigator Program, Pew Scholars Program in the Biomedical Sciences (grant number 27345), and the Searle Scholars Program (grant number 15-SSP-162). This work is based on research conducted at the Northeastern Collaborative Access Team (NE-CAT) beamlines, which are funded by the National Institute of General Medical Sciences from the National Institutes of Health (P41 GM103403). The Pilatus 6M detector on NE-CAT 24-ID-C beam line is funded by a NIH-ORIP HEI grant (S10 RR029205). This research used resources of the Advanced Photon Source, a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under Contract No. DE-AC02-06CH11357. RNAseq service was provided free of charge by the Beijing Genome Institute in exchange for an evaluation of their BGISEQ-500 sequencing platform. We thank Bryan Marotta for an introduction to kava, Chi Nguyen and Fadel A. Samatey for advice regarding crystallography, Gerald Fink for providing yeast strains and expression vectors, and Weng lab members for constructive comments. Kava plant photo courtesy of Randy Travis.

Data Availability

The sequences of the genes reported in this article have been deposited in NCBI GenBank (accessions MK058492–MK058514). Protein expression plasmids are available from Addgene (see Supplementary Table 5). The raw sequencing reads have been submitted to NCBI SRA (accession PRJNA494686) and the *de novo* assembled transcriptome to NCBI TSA (accession GHAC00000000). Raw metabolomic LC-MS datasets have been uploaded to the GNPS-MassIVE database (accessions MSV000083272 and MSV000083274–MSV000083277). Protein structures have been deposited in Protein Data Bank (accessions 6CO0, 6CQB, and 6NBR).

Author Contributions

T.P. and J.K.W. designed experiments. T.P. performed most of the experiments. M.P.T.S. assisted with cloning and crystallography. T.R.F. assisted with transcriptome assembly and LC-MS analyses. A.D.A. cloned genes and purified proteins. C.H.S. constructed expression vectors. T.P. analysed data. T.P. and J.K.W. wrote the paper.

Competing Interests

T.P. and J.K.W. have filed a patent application on metabolic engineering of kavalactones and flavokavains using the enzymes discovered in this study. J.K.W. is a co-founder, a member of the Scientific Advisory Board, and a shareholder of DoubleRainbow Biosciences, which develops biotechnologies related to natural products.

Figures

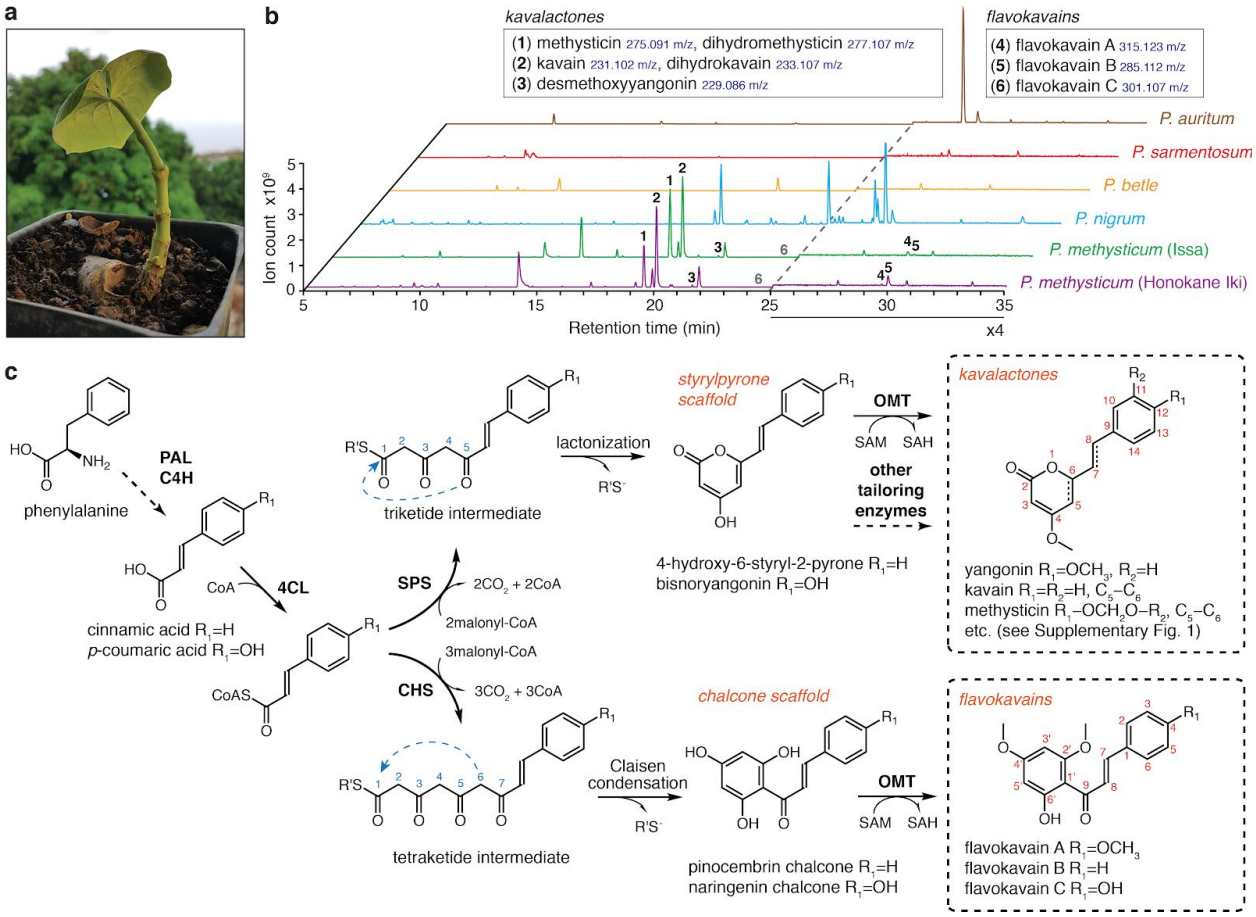


Figure 1 | Chemotypes of select *Piper* species and the bifurcation of kavalactone and flavokavain biosynthesis from common (hydroxy)cinnamoyl-CoA precursors in kava. a, A young root cutting of the Hawaiian kava cultivar *Honokane Iki*. **b,** LC-MS base peak plots of leaf tissue extracts from two different cultivars of kava (*Piper methysticum*) and four other *Piper* species showing that kavalactones and flavokavains are major secondary metabolites in kava absent in the other examined *Piper* species. The intensity in the 25-35 min region was increased 4-fold to help visualize the flavokavain peaks. The flavokavain C (**6**) peak is present but below the plot baseline. **c,** Hypothesized mechanisms for kavalactone and flavokavain biosynthesis. The polyketide cyclization mechanisms are denoted by blue arrows. PAL, phenylalanine ammonia-lyase; C4H, *trans*-cinnamate 4-monooxygenase; 4CL, 4-coumarate-CoA ligase; SPS, styrylpyrone synthase; CHS, chalcone synthase; OMT, *O*-methyltransferase; CoA, coenzyme A; SAM, *S*-adenosyl-L-methionine; SAH, *S*-adenosyl-L-homocysteine.

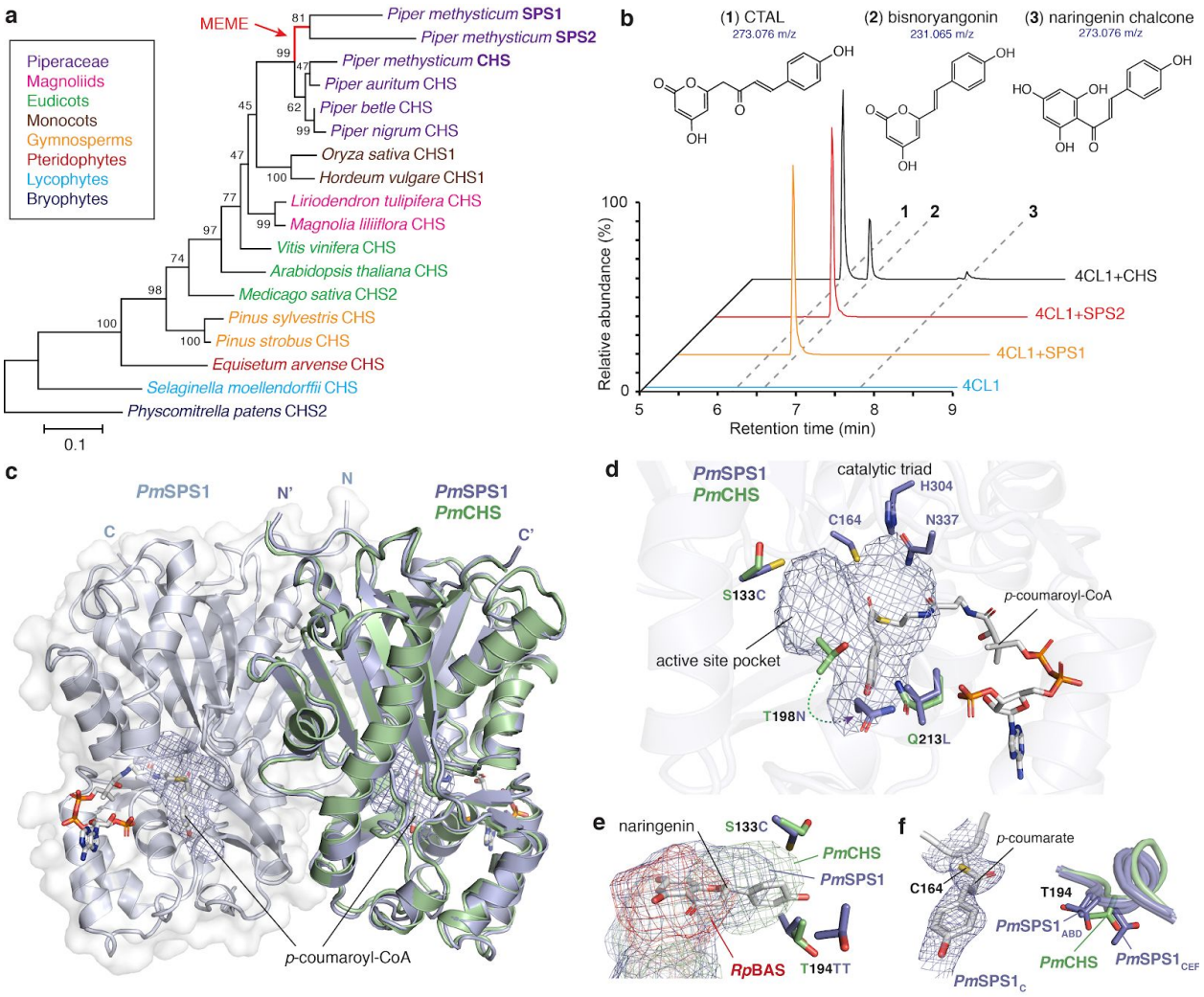


Figure 2 | Mechanistic basis for the neofunctionalization of SPSs from ancestral CHS in kava. a, Maximum-likelihood phylogenetic analysis of *Pm*SPS1, *Pm*SPS2 and *Pm*CHS from kava together with other CHS othologs from select land plant lineages. Bootstrap values (based on 1000 replicates) are indicated at the tree nodes. The branch undergone episodic positive selection as detected by the Mixed Effects Model of Evolution analysis[30] is highlighted in red and denoted with MEME. **b,** Combined LC-MS extracted ion chromatograms (XICs) of 273.076 m/z (naringenin chalcone and CTAL) and 231.065 m/z (bisnoryangonin) showing the *in vitro* activities of *Pm*SPS1, *Pm*SPS2 and *Pm*CHS from a 4CL-coupled enzyme assay using *p*-coumaric acid as the starter substrate. **c,** An overlay of the *Pm*CHS apo structure and the *Pm*SPS1-*p*-coumaroyl-CoA holo structure (RMSD=1.019 Å). Both proteins form homodimers (only a single protomer of *Pm*CHS is shown). Active-site pocket is indicated with a mesh. **d,** Structural features of the *Pm*SPS1 active site. The three active-site amino acid substitutions between *Pm*SPS1 and *Pm*CHS are highlighted. **e,** The C133 and T194 residues of *Pm*SPS1 dictate the size and

shape of the *Pm*SPS1 active site pocket in comparison to *Pm*CHS and *Rheum palmatum* BAS (*Rp*BAS). The naringenin coordinates are from the *Medicago sativa* CHS structure (PDB 1CGK[84]) based on structural alignment. **f,** *p*-coumaroyl-monoketide intermediate covalently bound to the catalytic cysteine as captured in chain C of the *Pm*SPS1-*p*-coumaroyl-CoA structure. The $|2F_o - F_c|$ electron density map is contoured at 1.2 σ. T194 residue of a nearby loop captured in closed (chains A, B, and D) and open (chains C, E, and F) conformations, in comparison with the corresponding residue in the *Pm*CHS structure.

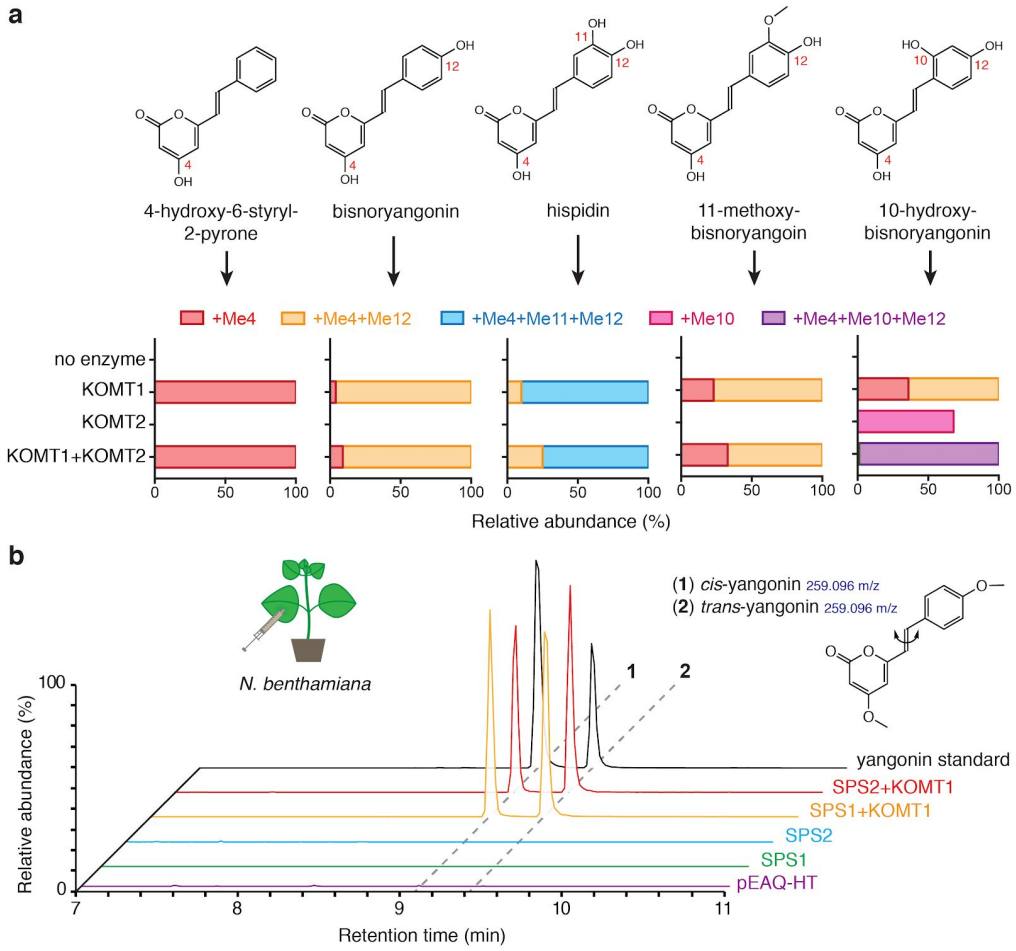


Figure 3 | Functional characterization of kava OMTs. a, *In vitro* enzyme activities of *Pm*KOMT1 and *Pm*KOMT2 against five different styrylpyrone substrates. The bar charts indicate proportions of differentially methylated products relative to the unmethylated precursor (e.g., +Me4+Me12 indicates methyl group addition to both C_4 and C_{12} sites). Detailed results are shown in Supplementary Note 4. **b,** LC-MS XICs of 259.096 m/z (yangonin) in leaf extracts from *N. benthamiana* transiently expressing indicated enzymes. Yangonin readily isomerizes between *cis* and *trans* configurations in solution.

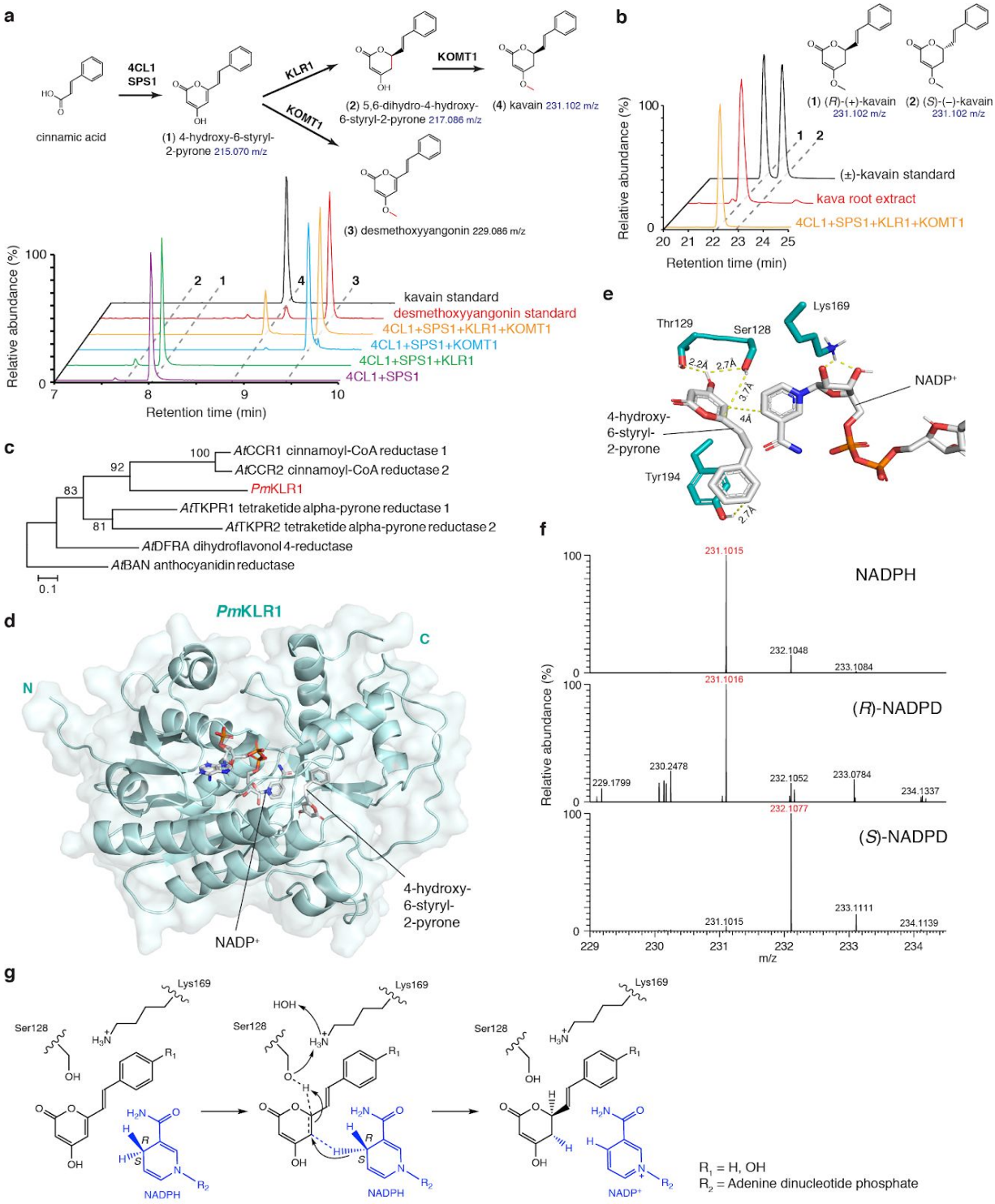


Figure 4 | Functional characterization of *Pm*KLR1. a, Combined LC-MS XICs of 215.070 m/z (**1**), 217.086 m/z (**2**), 229.086 m/z (**3**), and 231.102 m/z (**4**) from *in vitro* enzyme assays using the indicated combinations of enzymes and cinnamic acid as the initial substrate to show the specific activity of

*Pm*KLR1. **b,** Chiral chromatography-MS traces of 231 m/z (kavain) demonstrating the stereo-selectivity of *Pm*KLR1. **c,** Maximum likelihood phylogenetic analysis of *Pm*KLR1 together with several characterized reductases from *A. thaliana*. Bootstrap values based on 1,000 replicates are indicated at the tree nodes. The scale measures evolutionary distances in substitutions per amino acid. **d,** The holo structure of *Pm*KLR1 in complex with $NADP^+$. The substrate 4-hydroxy-6-styryl-2-pyrone was computationally docked into the active site, where Tyr194 was set as flexible. **e,** A detailed view of *Pm*KLR1 active site, highlighting several key residues involved in substrate binding and catalysis. **f,** Mass spectra of kavain produced by an *in vitro* enzyme assay using *Pm*4CL1, *Pm*SPS1, *Pm*KLR1 and *Pm*KOMT1 enzymes starting from cinnamic acid. NADPH, deuterium-labeled NADPH isotopic analog (*R*)-NADPD or (*S*)-NADPD was used as a co-substrate for *Pm*KLR1. **g,** The proposed catalytic mechanism of *Pm*KLR1.

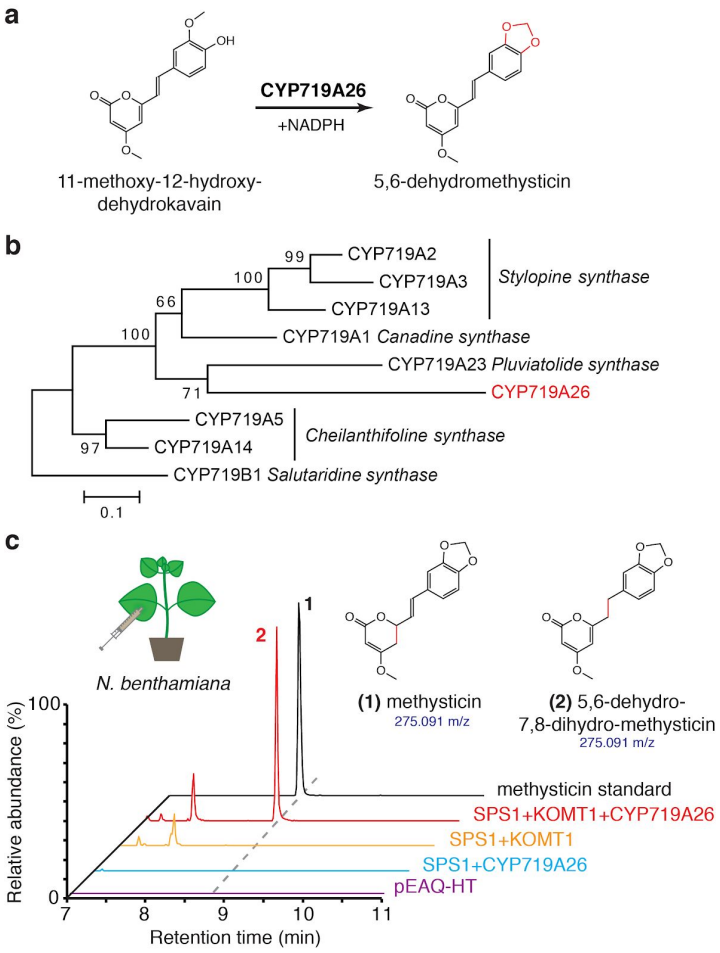


Figure 5 | Functional characterization of the methylenedioxy-bridge-forming enzyme *Pm*CYP719A26 (*Pm*MTS1). a, Proposed reaction catalyzed by *Pm*CYP719A26 (*Pm*MTS1). **b,** Maximum likelihood phylogenetic analysis of *Pm*CYP719A26 together with several characterized members of the CYP719 family P450s. Bootstrap values (1,000 replicates) are indicated at the tree nodes. The scale measures evolutionary distances in substitutions per amino acid. **c,** LC-MS XICs of 275.091 m/z (methysticin) in *N. benthamiana* leaves infiltrated with the indicated combinations of enzymes to show the specific activity of *Pm*MTS1. The two structural isomers shown were distinguished using MS/MS analysis (Supplementary Note 6).

Supplementary Information for

The biosynthetic origin of psychoactive kavalactones in kava

Tomáš Pluskal[1], Michael P. Torrens-Spence[1], Timothy R. Fallon[1,2], Andrea De Abreu[1,2], Cindy H. Shi[1,2], and Jing-Ke Weng[1,2]*

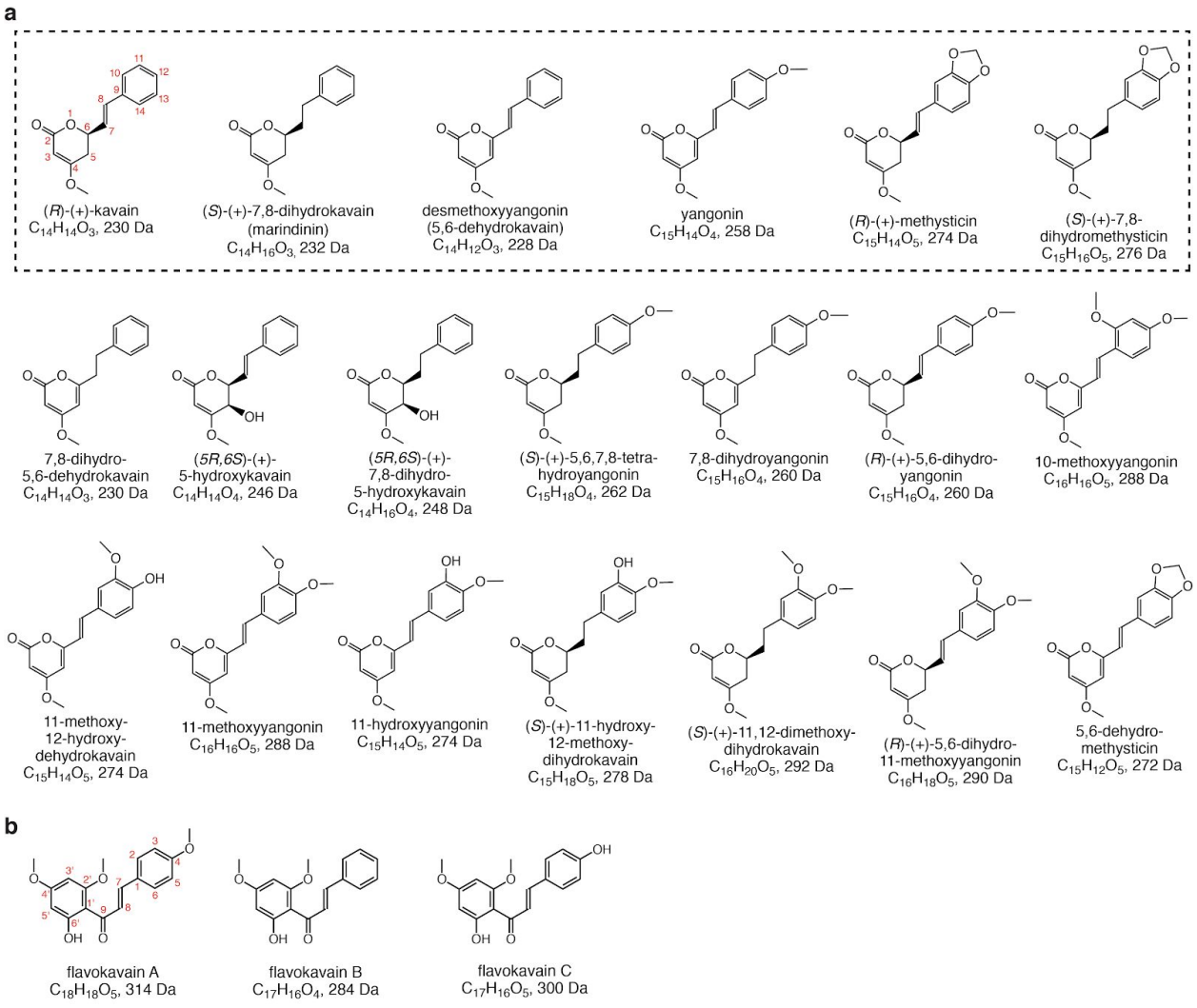
[1]Whitehead Institute for Biomedical Research, 455 Main Street, Cambridge, MA 02142 USA.

[2]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139 USA.

*Correspondence to: wengj@wi.mit.edu

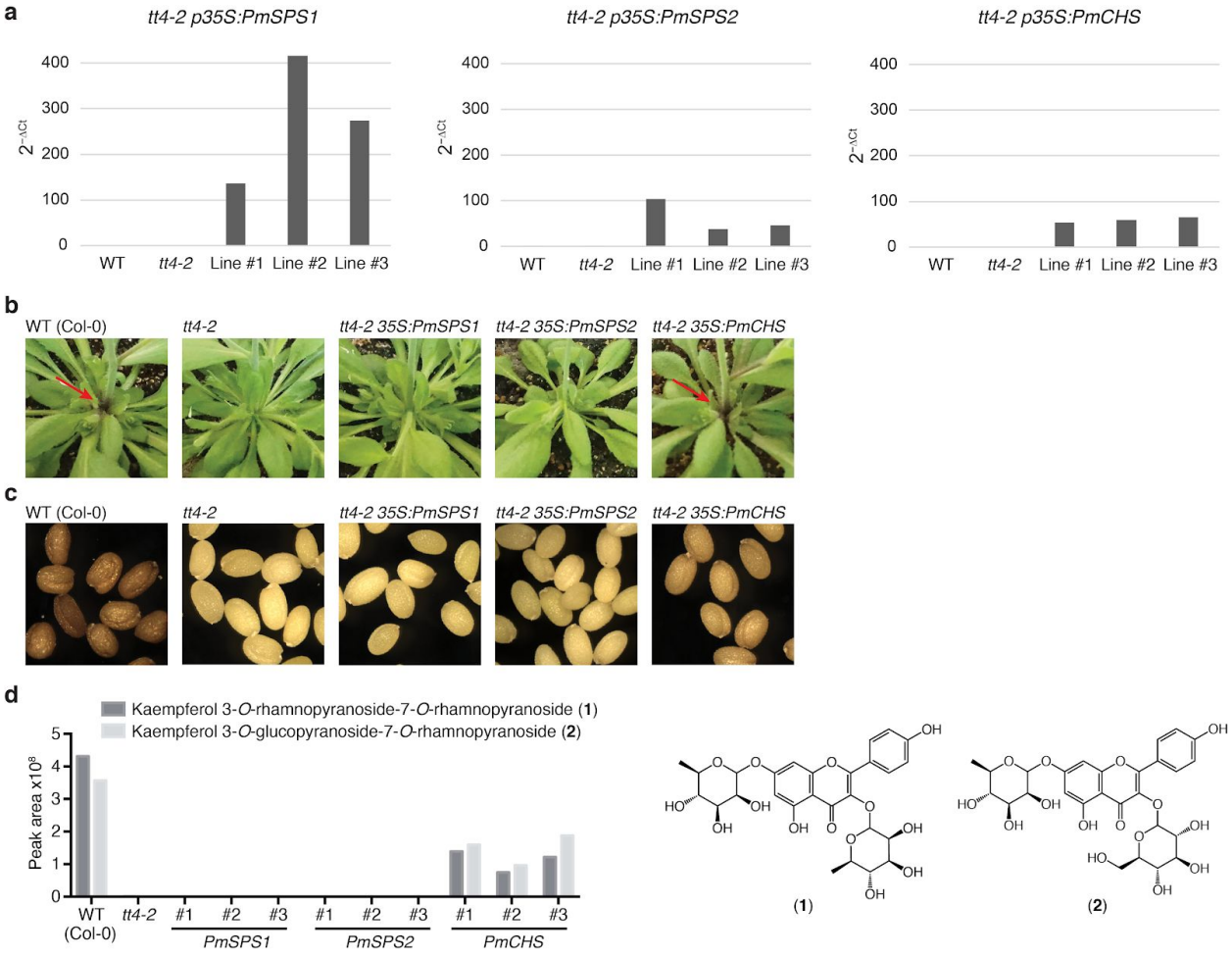
Table of contents

Supplementary Figures



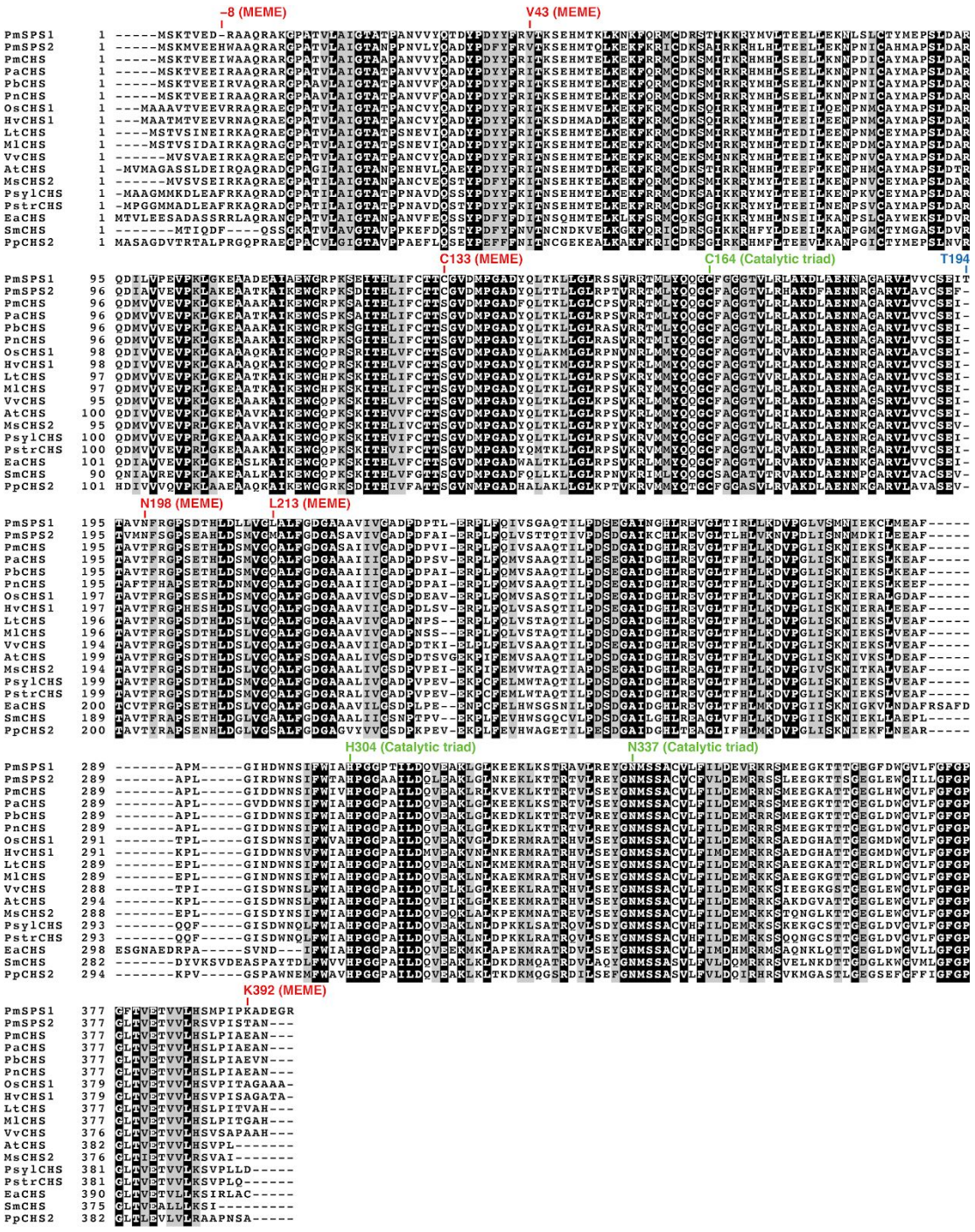
Supplementary Figure 1 | The chemical diversity of kavalactones and flavokavains.

a, Chemical structures, formulas, and molecular masses of twenty known kavalactones. The six major kavalactones, which constitute over 96% kavalactone content in the plant rhizome, are highlighted within a dashed rectangle. **b,** Chemical structures, formulas, and molecular masses of three known flavokavains.



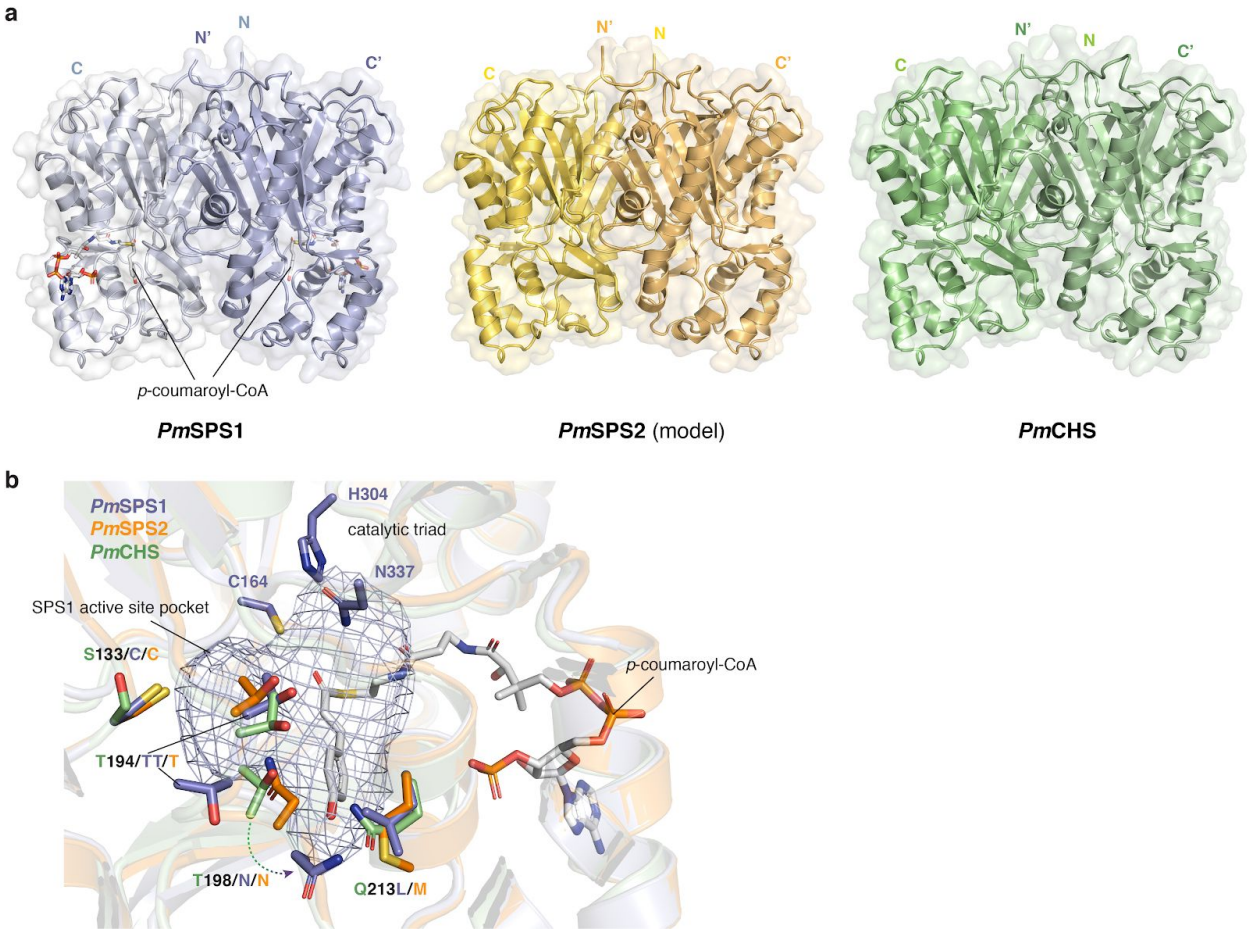
Supplementary Figure 2 | Stable transgenic expression of *PmSPS1*, *PmSPS2*, and *PmCHS* in *A. thaliana tt4-2* background.

a, Expression levels of *PmSPS1*, *PmSPS2*, and *PmCHS* relative to a reference gene, *At1g13320*, in three different *A. thaliana* lines measured by qRT-PCR in the T2 generation. **b,** Transgenic expression of *PmCHS*, but not *PmSPS1* and *PmSPS2* rescues the anthocyanin phenotype observed at the stem base in the *A. thaliana tt4-2* mutant background. **c,** Transgenic expression of *PmCHS* partially rescues the transparent testa phenotype of the seed coat in the *A. thaliana tt4-2* mutant background. The difference in coloring intensity in WT and *tt4-2*-35S:*PmCHS* seed coat an be attributed to the strength of the 35S promoter in seeds. **d,** LC-MS quantification of two major flavonoids[1] in the leaves of three independent transgenic lines of *tt4-2*-35S:*PmCHS* plants.



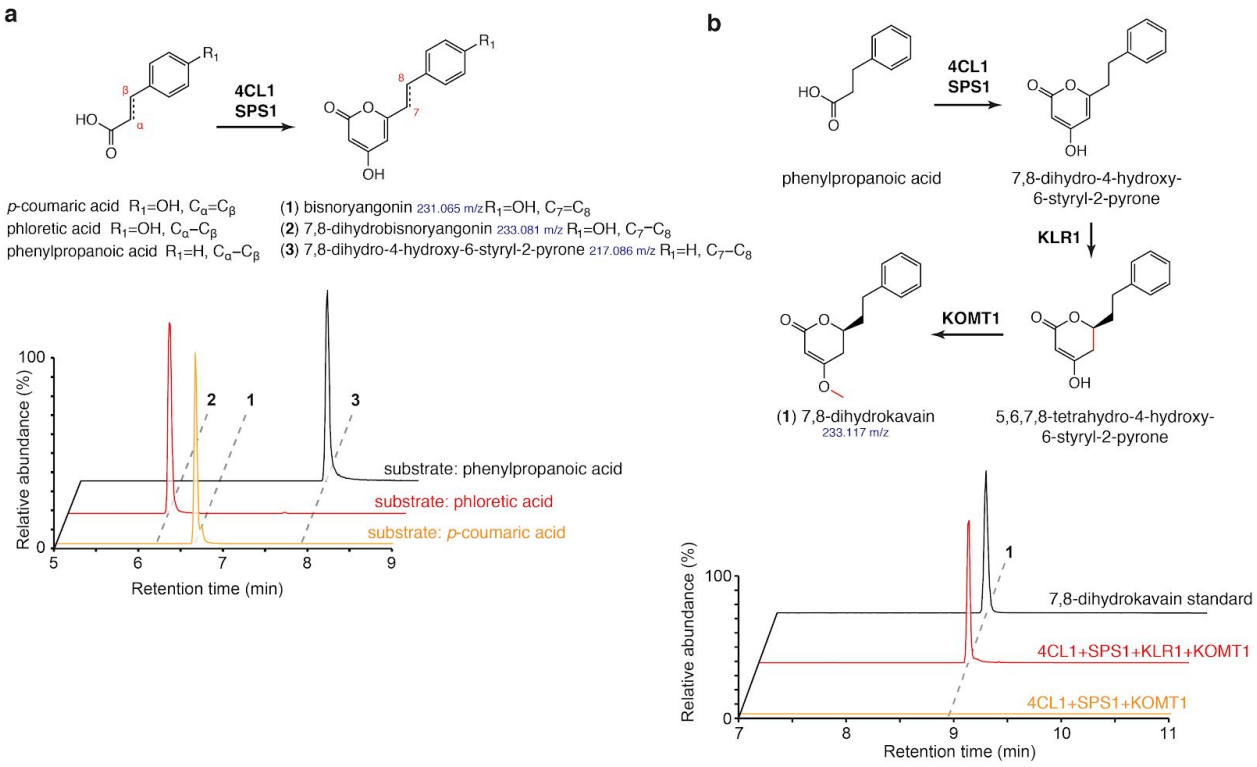
Supplementary Figure 3 | Multiple sequence alignment of CHS homologs as in Fig. 2A.

The residues undergone episodic positive selection were detected based on the Mixed Effects Model of Evolution[2] along the ancestral branch of kava SPSs, and are annotated with MEME. Residue numbering is according to *Pm*SPS1 sequence.



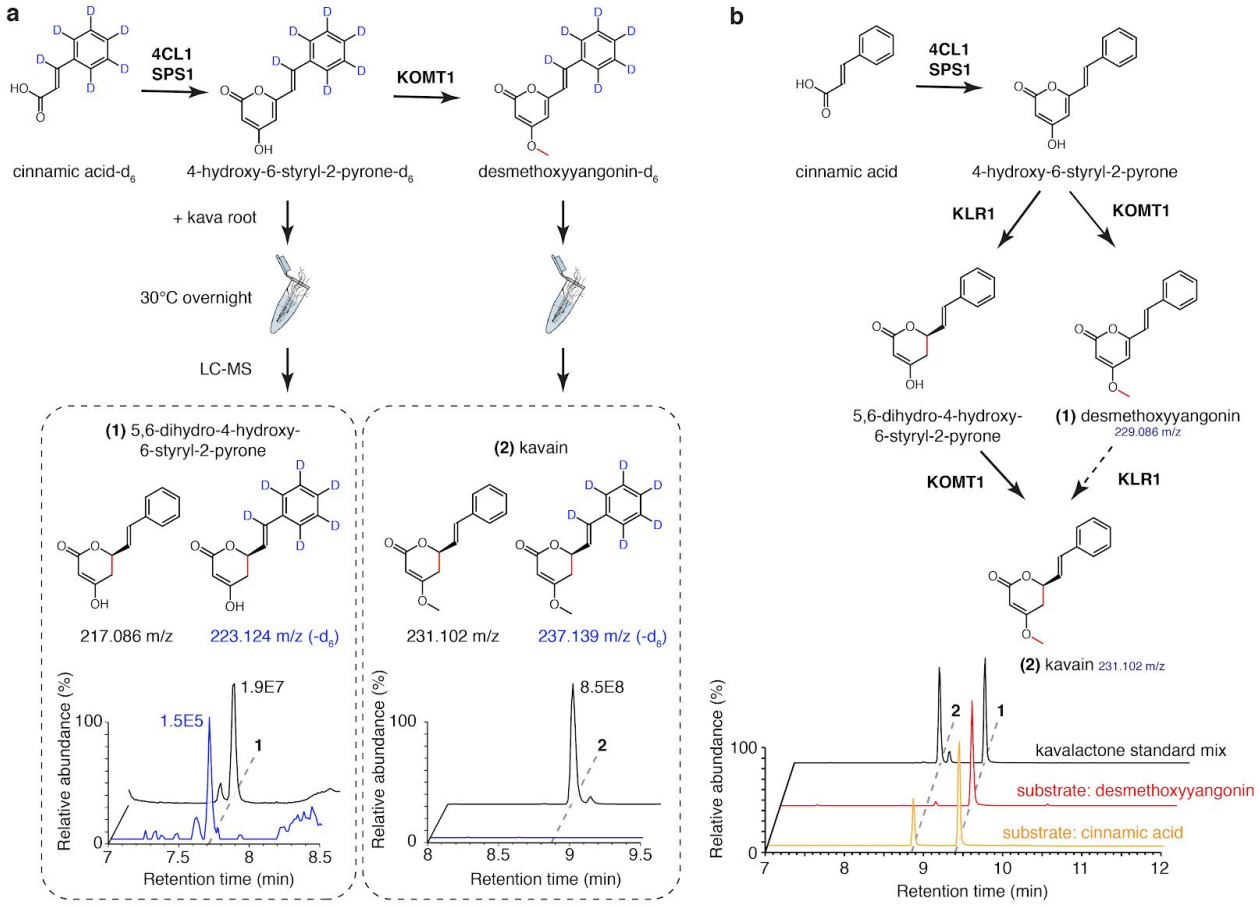
Supplementary Figure 4 | Structural comparison of *Pm*SPS1, *Pm*SPS2, and *Pm*CHS.

a, The overall fold of experimentally determined *Pm*SPS1-*p*-coumaroyl-CoA and *Pm*CHS structures in comparison with the computationally modeled *Pm*SPS2 structure. **b,** *Pm*SPS1 and *Pm*SPS2 contain a similar set of amino acid substitutions in the active site relative to *Pm*CHS. Although both *Pm*SPS1 and *Pm*SPS2 contain the T198N substitution, the actual position of this residue is different due to the T194 insertion in *Pm*SPS1.



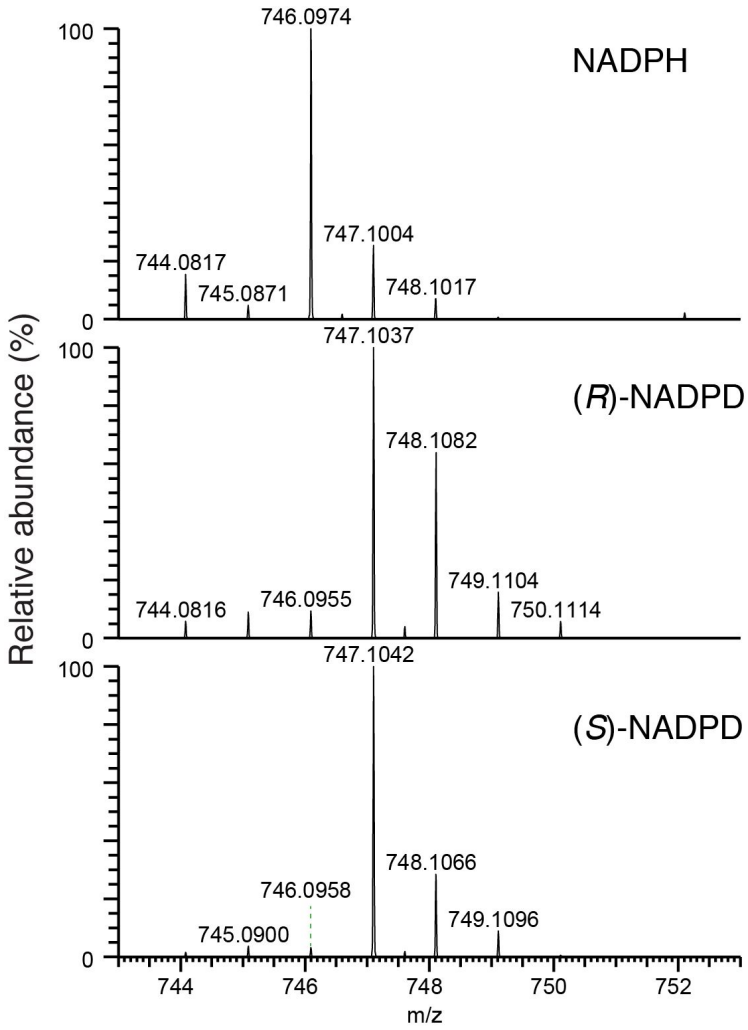
Supplementary Figure 5 | The biosynthesis of 7,8-saturated kavalactones.

a, Combined LC-MS XICs of 231 m/z (**1**), 233 m/z (**2**), and 217 m/z (**3**) in coupled (*Pm*4CL1 + *Pm*SPS1) *in vitro* enzyme assays to show that kavalactone biosynthesis is operational starting with dihydro-hydroxycinnamic acids. The use of α,β-reduced substrates (phloretic acid and phenylpropanoic acid) resulted in the production of corresponding 7,8-saturated styrylpyrones. **b,** LC-MS XICs of 233.117 m/z (7,8-dihydrokavain) in combinatorial *in vitro* enzyme assays using the indicated enzymes and phenylpropanoic acid as a starter substrate. The addition of *Pm*KLR1 to the enzyme mix resulted in the production of 7,8-dihydrokavain, thus further supporting the conclusion that *Pm*KLR1 acts as a NADPH-dependent kavalactone 5,6-olefin reductase.



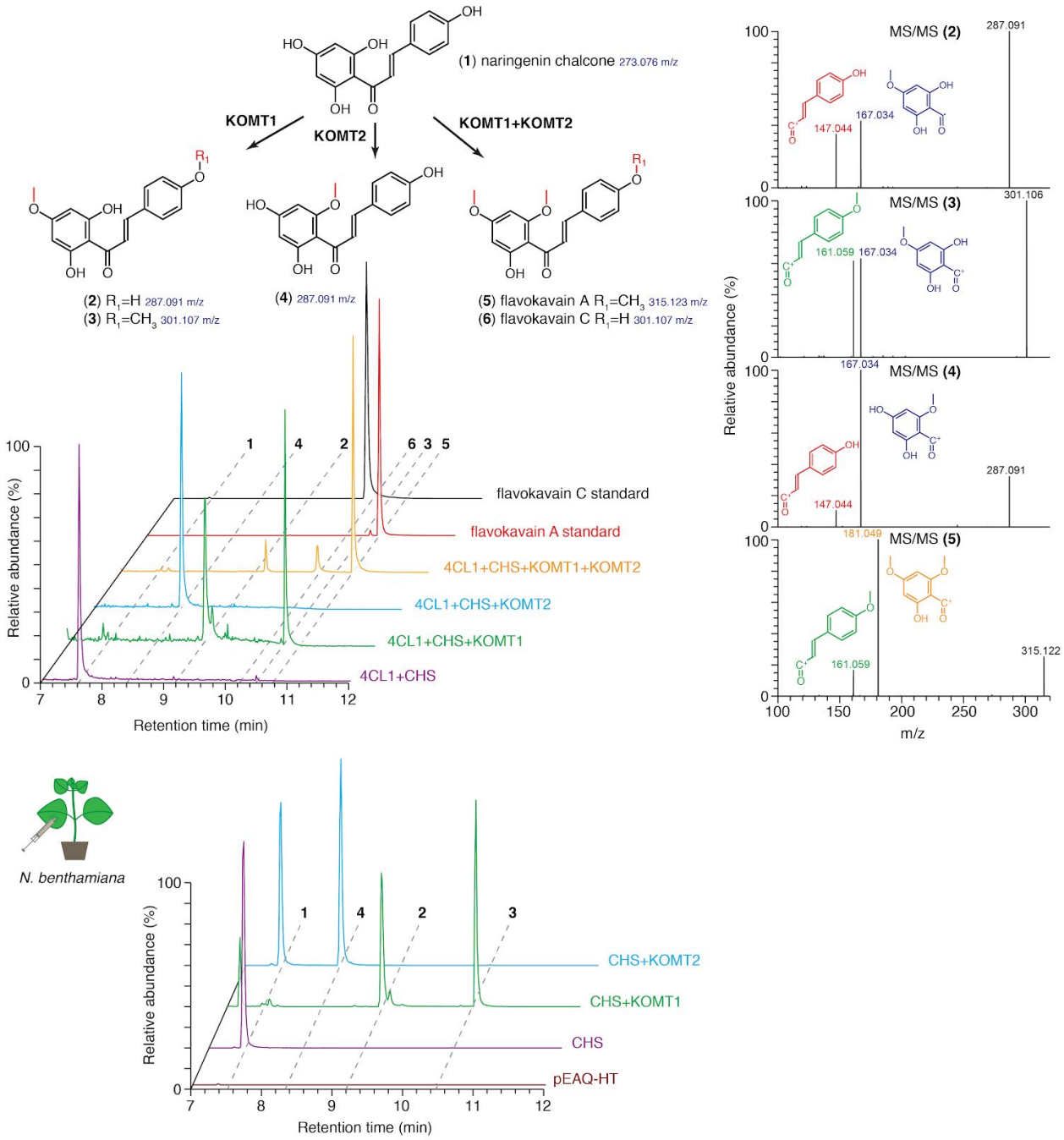
Supplementary Figure 6 | Resolving the order of *Pm*KOMT1 and *Pm*KLR1 reactions *in vivo* and *in vitro*.

a, Isotope tracing using labeled styrylpyrone or desmethoxyyangonin precursors in live kava root suggests that only the unmethylated precursor serve as a substrate for the reduction of C_5=C_6 olefin. **b,** Combined LC-MS XICs of 229.086 m/z (**1**) and 231.102 m/z (**2**) from *in vitro* enzyme assays using cinnamic acid and desmethoxyyangonin as starter substrates to show that *Pm*KLR1 precedes *Pm*KOMT1. The reaction from cinnamic acid successfully produced kavain, while the same assay starting from desmethoxyyangonin failed to show any activity. The determined order of the pathway is shown with solid arrows, while the rejected order is shown with a dashed arrow.



Supplementary Figure 7 | Mass spectra of $[M+H]^+$ ions of NADPH, (*R*)-NADPD, and (*S*)-NADPD.

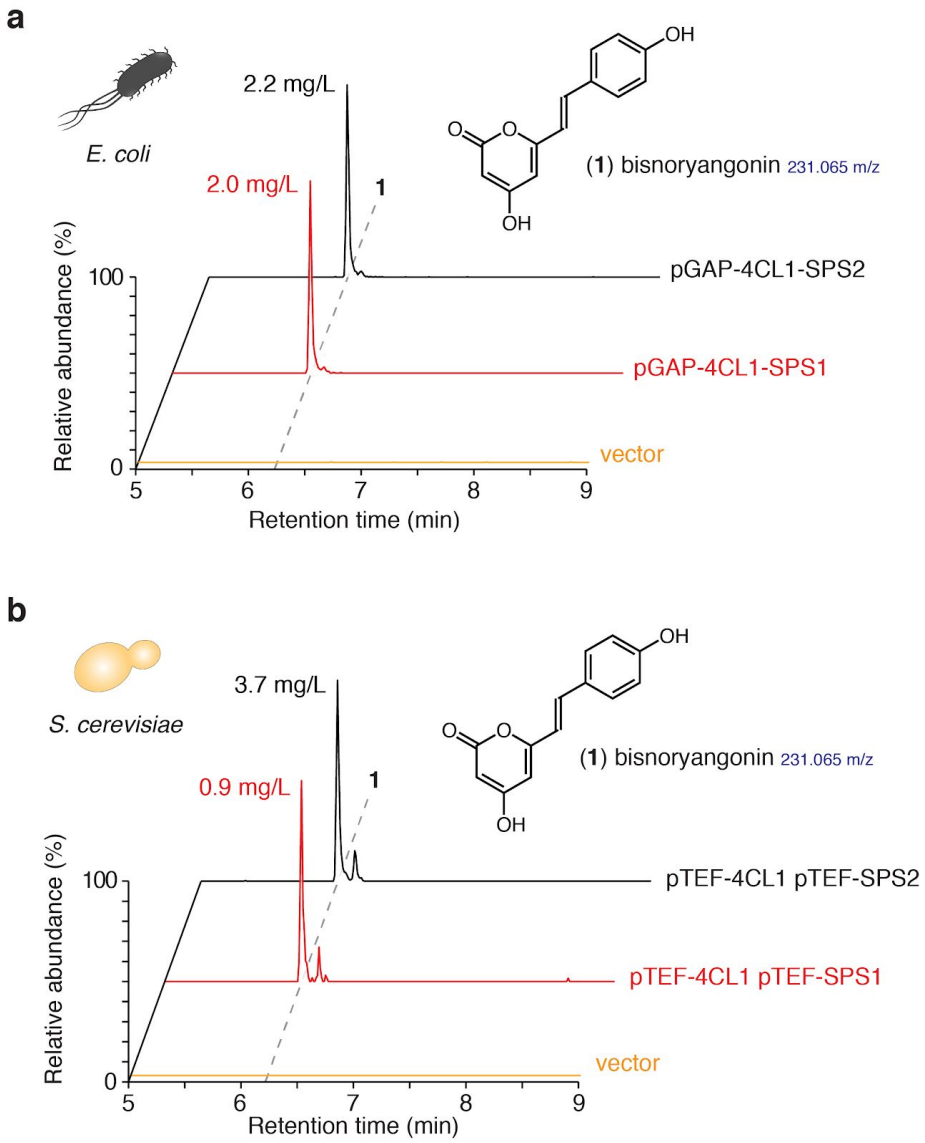
(*R*)-NADPD was produced by alcohol dehydrogenase from isopropanol-d8 and (*S*)-NADPD was produced by glucose-6-phosphate dehydrogenase and hexokinase from glucose-1-d.



Supplementary Figure 8 | The biosynthesis of flavokavains using *Pm*CHS, *Pm*KOMT1, and *Pm*KOMT2.

a, Combined LC-MS XICs of 273.076 m/z (**1**), 287.091 m/z (**2, 4**), 301.107 m/z (**3, 6**), and 315.123 m/z (**5**) from *in vitro* enzyme assays using the indicated combinations of enzymes and *p*-coumaric acid as the starter substrate. MS/MS analysis interpretation is shown on the right to support the identification of

individual products**. b,** Combined LC-MS XICs of 273.076 m/z (**1**), 287.091 m/z (**2, 4**), 301.107 m/z (**3, 6**) in leaf extracts from *N. benthamiana* transiently expressing indicated enzymes. Expression in *N. benthamiana* resulted in equivalent compounds to those produced by *in vitro* enzyme assays.

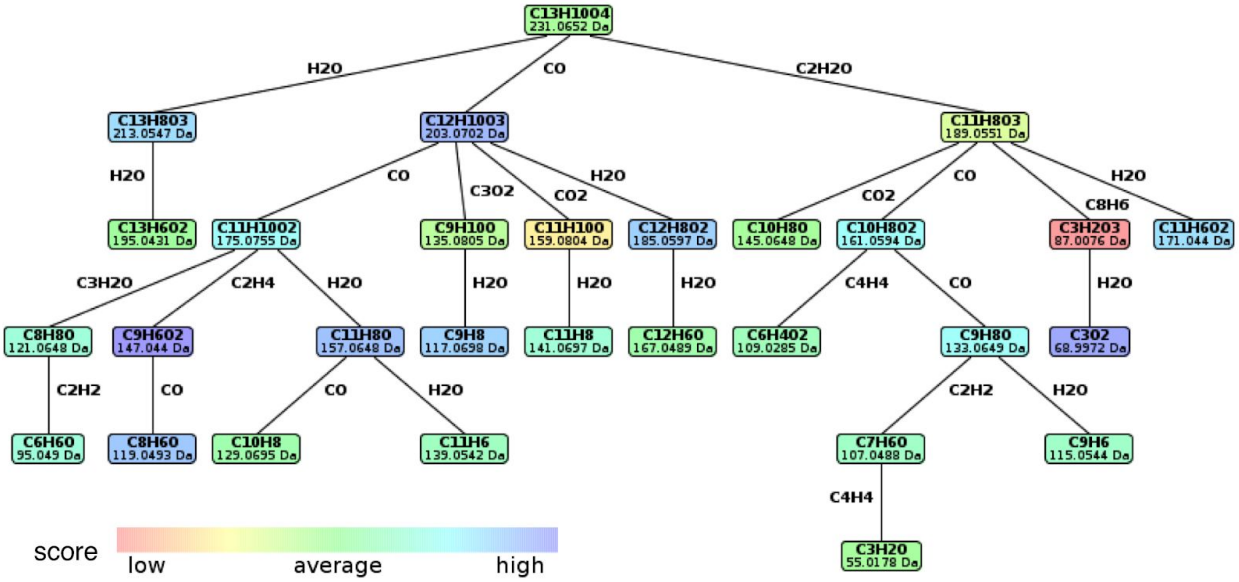


Supplementary Figure 9 | Heterologous production of the kavalactone precursor, bisnoryangonin, in *E. coli* and yeast.

a, LC-MS XIC of 231.065 m/z (bisnoryangonin) from the media supernatants of *E. coli* cultures incubated overnight at 30°C with 1 mM *p*-coumaric acid. Cells were expressing the indicated enzymes from a single plasmid under the control of a constitutive *pGAP* promoter. **b,** LC-MS XIC of 231.065 m/z (bisnoryangonin) from the media supernatants of *S. cerevisiae* cultures incubated for 2 days at 30°C with 2 mM *p*-coumaric acid. Cells were expressing the indicated enzymes from two plasmids under the control of a constitutive *pTEF* promoter.

Supplementary Note 1. Confirmation of bisnoryangonin by LC-MS.

The product of kava SPSs, bisnoryangonin, is a key precursor for kavalactone biosynthesis (Fig. 1C). Since its standard is not commercially available, we used MS/MS spectral analysis to confirm its identity. First, we interpreted the MS/MS spectrum of the suspected biosnoryangonin LC-MS peak (231.065 m/z @6.28 min) from *in vitro* enzyme assays using SIRIUS[3] and generated a fragmentation tree of this peak:

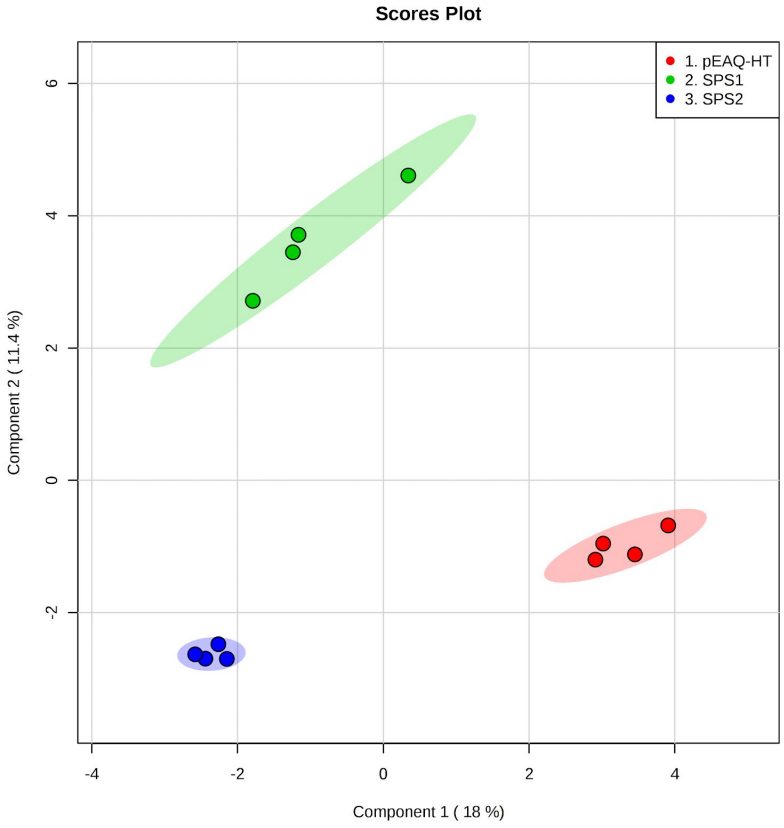


Next, we matched the fragmentation tree against all compounds in the PubChem database[4] using the machine learning-based CSI-FingerID[5] algorithm in SIRIUS. Bisnoryangonin structure was determined to be the best-fitting candidate for this fragmentation tree among ~94 million molecules in PubChem:

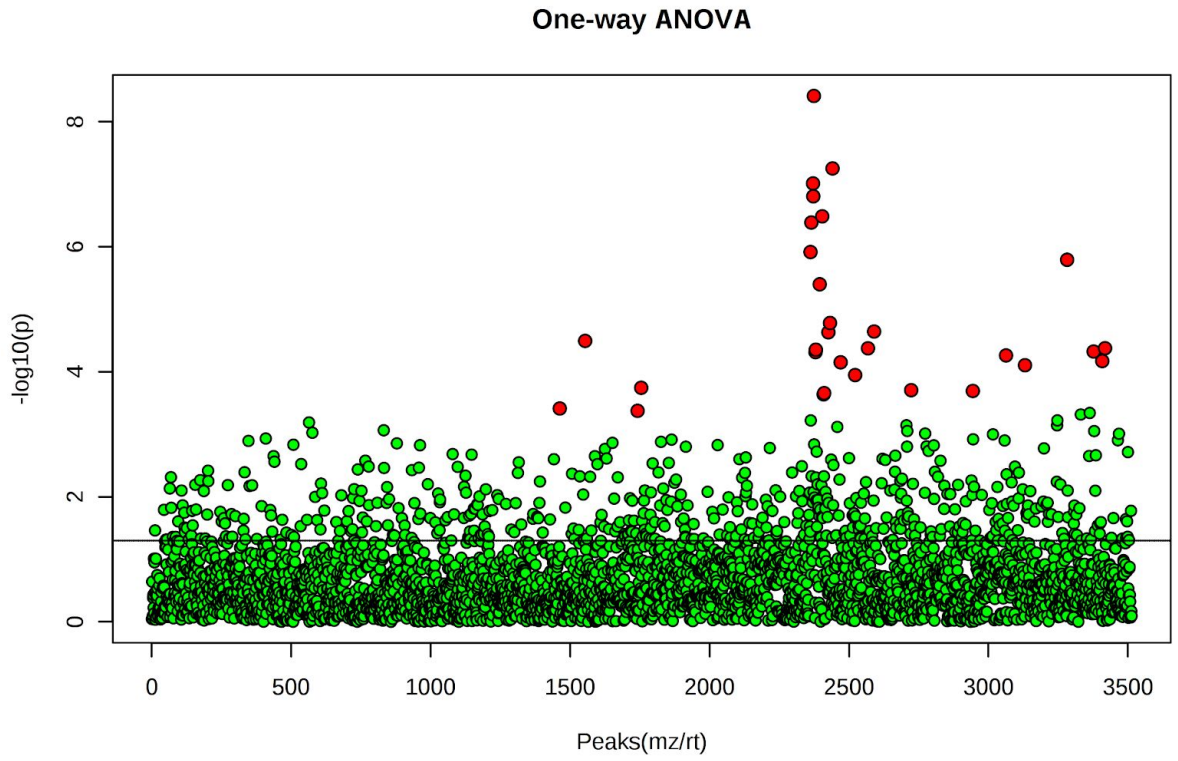
Rank	Molecular Formula	Similarity	Name	CSI:FingerID Score	XLogP	InChIKey
1	C13H10O4	63.71%	Bisnoryangonin	-110.53	2.16	ORVQWHLMVLOZPX
2	C13H10O4	58.36%		-126.23	2.87	YOCUVYCKSGIQBX
3	C13H10O4	56.41%		-133.65	2.87	QDGJMIGWGXAORV
4	C13H10O4	57.36%		-133.73	2.87	RPDNNTACNWVSBD
5	C13H10O4	56.46%		-135.86	3.75	FAMUOIFJNGMNSN
6	C13H10O4	55.76%		-136.25	3.75	LQPQIGOXKGBSJB
7	C13H10O4	58.49%	Oprea1_062728	-136.81	2.79	OKJFKPFBSPZTAH

Supplementary Note 2. Benzalacetone-like compounds detected in *A. thaliana* and *N. benthamiana* expressing *PmSPS1*.

Since we detected no bisnoryangonin or other styrylpyrone compounds in *A. thaliana* lines expressing *PmSPS1* or *PmSPS2*, we proceeded to perform *Agrobacterium* infiltration-mediated transient expression of these kava genes in the leaves of *N. benthamiana*. We infiltrated four leaves each with agrobacteria carrying an empty vector (pEAQ-HT), a plasmid expressing *PmSPS1*, and a plasmid expressing *PmSPS2*. Each construct was infiltrated in four replicates (four leaves on different plants). Five days after infiltration, metabolites were extracted from the leaves and extracts analyzed by LC-MS. We detected 3,709 features in the datasets using MZmine 2 [6]. We applied sparse partial least squares discriminant analysis (sPLS-DA) algorithm [7] with 5 components and 10 features per component using MetaboAnalyst [8], resulting in clear separation of the samples according to expressed enzymes:



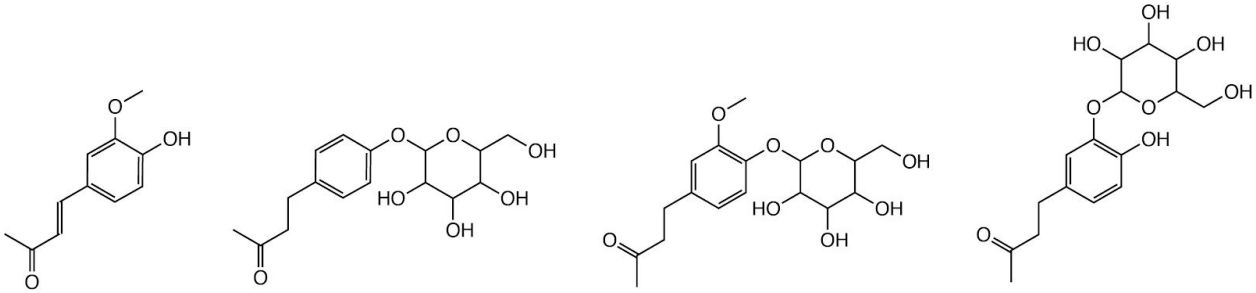
We further identified 30 significantly dysregulated peaks using one-way analysis of variance (ANOVA) with false discovery (FDR) p-value cutoff 0.05:



In a separate experiment, we co-infiltrated the *PmSPS1* and *PmSPS2* enzymes with an isotopically labeled substrate, cinnamic acid-d_6, with the intention to trace compounds derived from the phenylpropanoid pathway. This allowed us to monitor the labeling state each peak of interest: if the peak was derived from cinnamic acid-d_6, a labeled mass of that peak should be observed at the same retention time, carrying multiple deuterium atoms (typically d_6, d_5, or d_4, depending on the number of metabolic conversions from cinnamic acid-d_6).

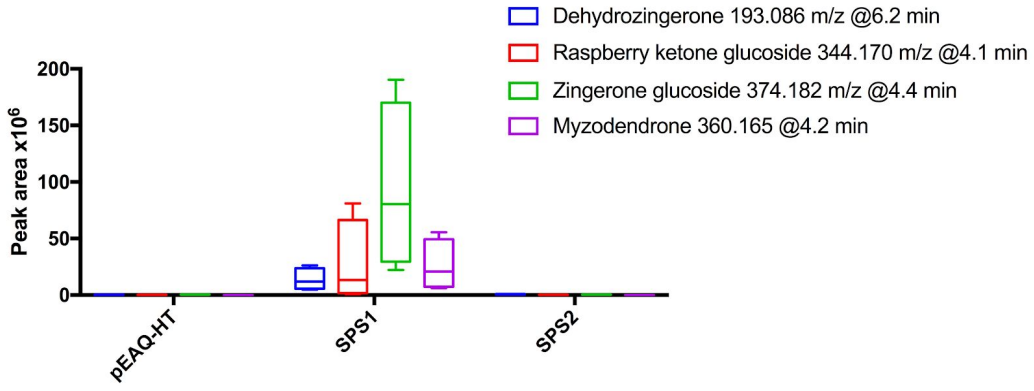
By manually examining the peaks revealed by ANOVA, as well as peaks with the highest peak area fold-change increase in *N. benthamiana* expressing *PmSPS1* and *PmSPS2*, we identified a series of metabolites with masses that matched polyketide structures derived from a benzalacetone backbone, listed below. The isotope labeling data further supported that these compounds were derived from cinnamic acid. We verified two of these metabolites (dehydrozingerone and raspberry ketone glucoside) using pure standards.

m/z	RT (min)	Type of ion	Formula	Isotope labeling	Identified metabolite
193.086	6.2	$[M+H]^+$	$C_{11}H_{12}O_3$	d_4	Dehydrozingerone (**1**; *standard verified*)
344.170	4.1	$[M+NH_4]^+$	$C_{16}H_{22}O_7$	d_5	Raspberry ketone glucoside (**2**; *standard verified*)
349.125	4.1	$[M+Na]^+$	$C_{16}H_{22}O_7$	–	Raspberry ketone glucoside (**2**; *standard verified*)
374.182	4.4	$[M+NH_4]^+$	$C_{17}H_{24}O_8$	d_4	Zingerone glucoside (**3**)
379.136	4.4	$[M+Na]^+$	$C_{17}H_{24}O_8$	d_4	Zingerone glucoside (**3**)
360.165	4.2	$[M+NH_4]^+$	$C_{16}H_{22}O_8$	d_4	Myzodendrone (**4**)
365.122	4.2	$[M+Na]^+$	$C_{16}H_{22}O_8$	–	Myzodendrone (**4**)

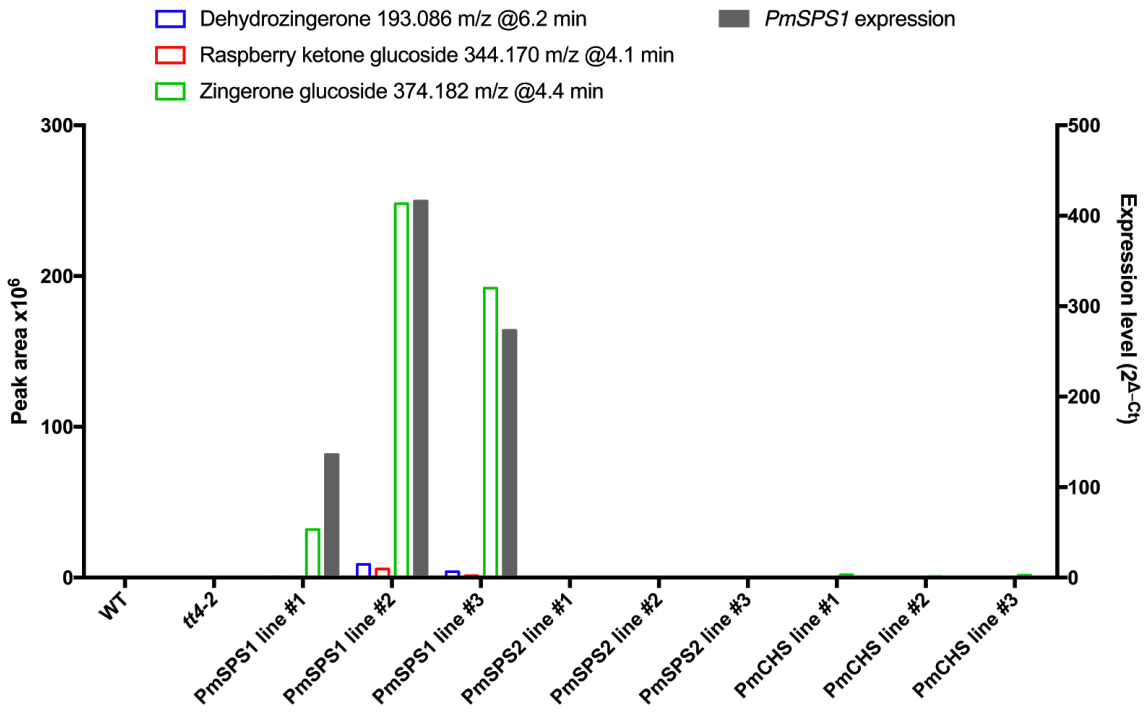


(**1**) (**2**) (**3**) (**4**)

These benzalacetone-like compounds were only produced in *PmSPS1*-expressing *N. benthamiana* leaves. We did not observe any compounds derived from cinnamic acid-d_6 in *PmSPS2*-expressing plants:



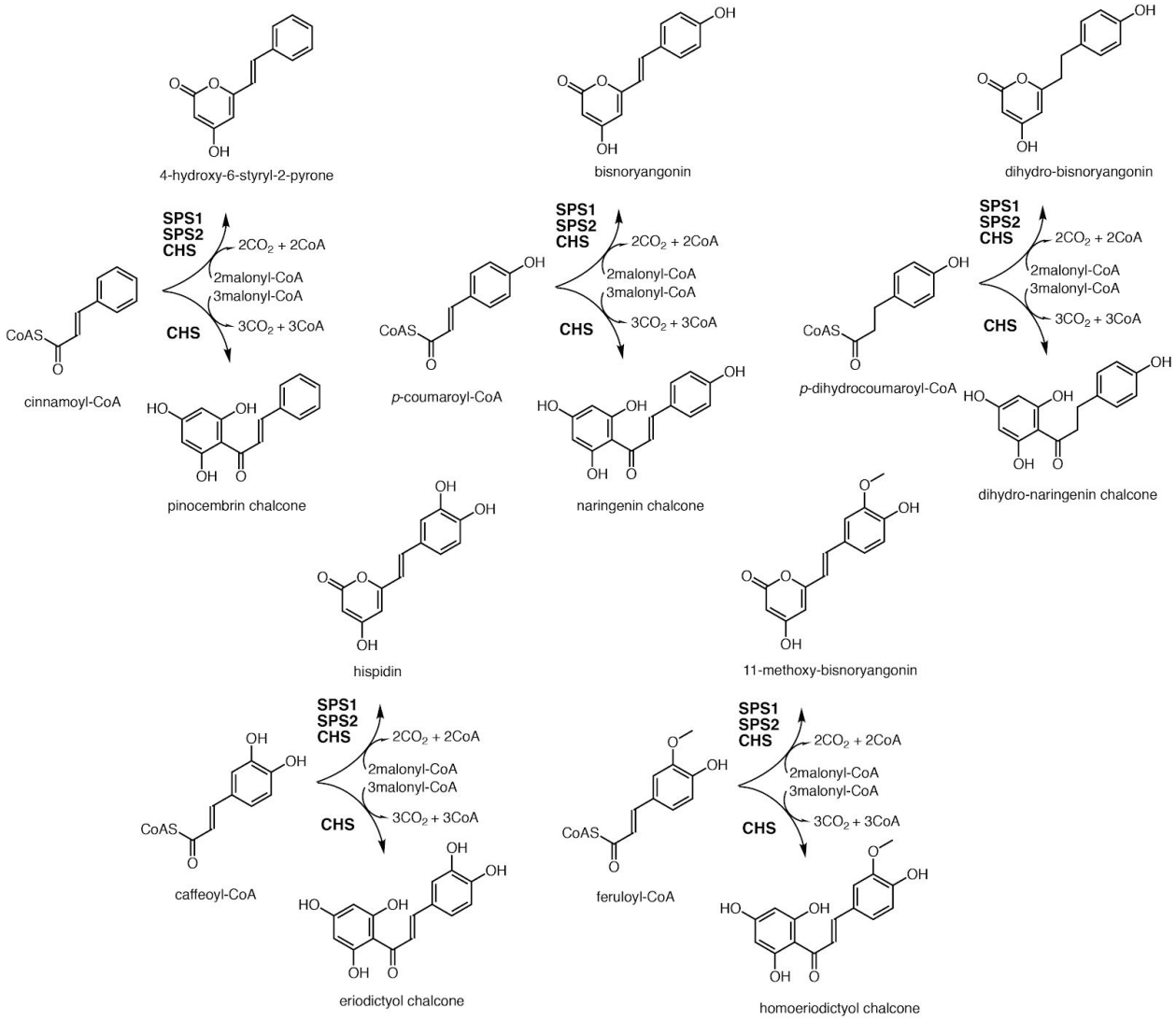
Three of the four identified benzalacetones were also detected in *A. thaliana* lines expressing *PmSPS1* (but not *PmSPS2*). In the *PmSPS1*-expressing lines, the peak areas of the benzalacetones were correlated with the *PmSPS1* expression levels measured by qRT-PCR (Fig. S2A):



We thus concluded that the benzalacetones were likely products of catabolism of styrylpyrones produced by *Pm*SPS1. This was later confirmed by co-infiltration with *Pm*KOMT1, which methylated the styrylpyrones, thus protecting them from degradation in both *PmSPS1*- and *PmSPS2*-expressing *N. benthamiana* plants (Fig. 3B).

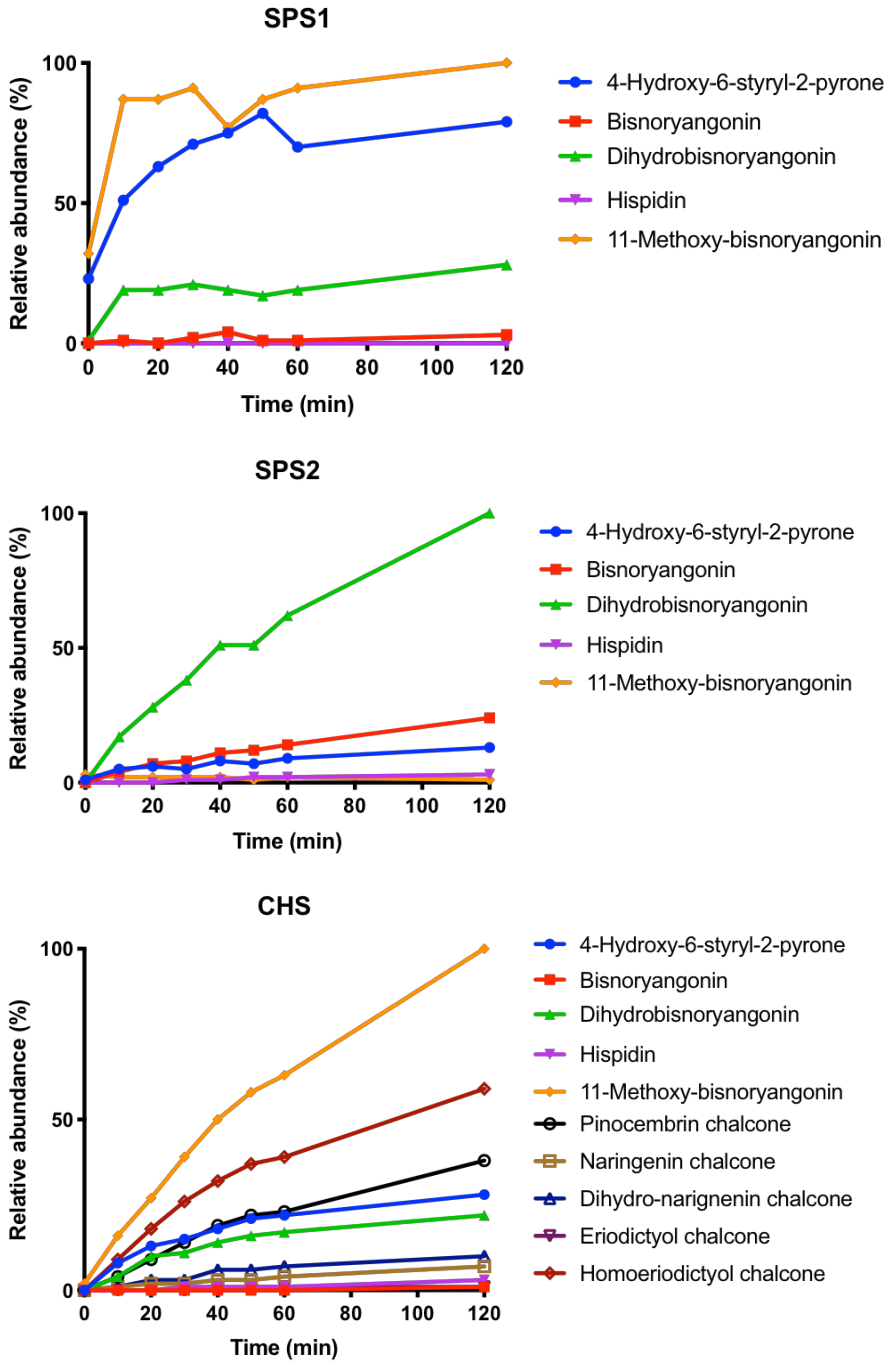
Supplementary Note 3. Substrate preference of *Pm*SPS1, *Pm*SPS2, and *Pm*CHS.

To test the substrate preference of the three kava PKSs, we used an *in vitro* enzyme assay with an equimolar (0.5 mM) mixture of substrates: cinnamoyl-CoA, *p*-coumaroyl-CoA, *p*-dihydrocoumaroyl-CoA, *p*-caffeoyl-CoA, feruloyl-CoA, and sinapoyl-CoA. No activity was detected on sinapoyl-CoA for any of the enzymes, and it was therefore excluded from further analysis. We considered all triketide and tetraketide products as shown in the following schema.



Note that the *Pm*CHS enzyme produces additional triacetic acid lactone peaks (e.g., CTAL), but we excluded those from the analysis to simplify the interpretation of the results.

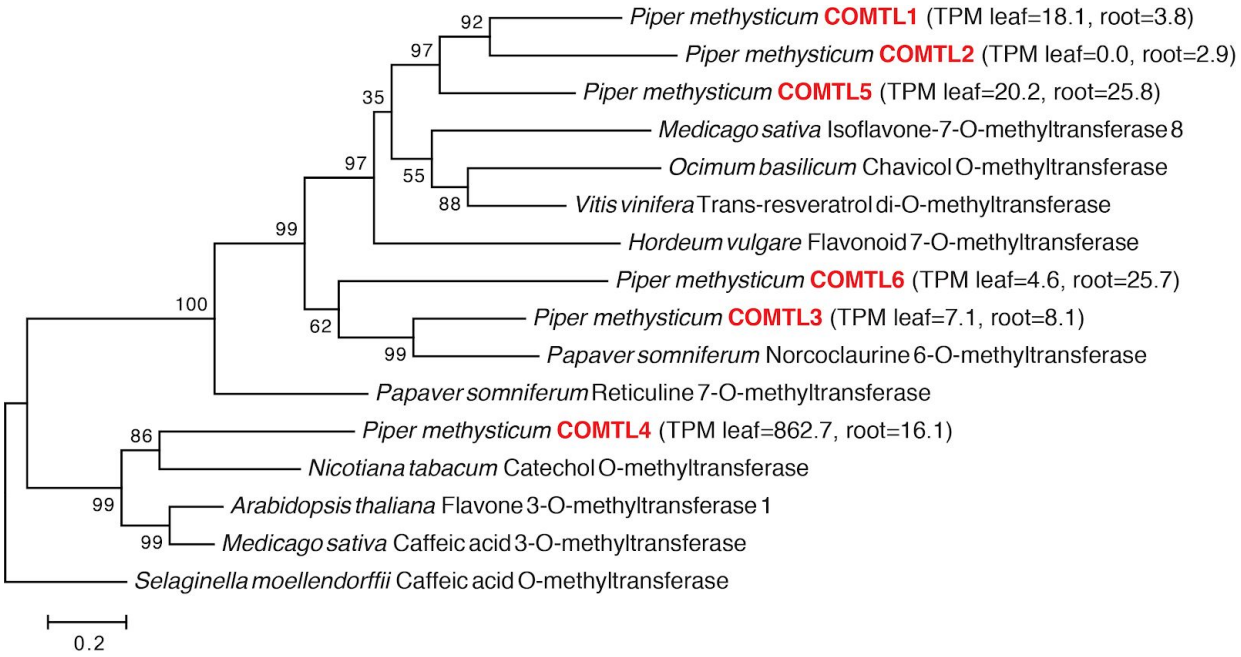
After adding a single enzyme (*Pm*SPS1, *Pm*SPS2, and *Pm*CHS) into the mixture of substrates, we measured in a time course experiment the peak areas of all the product peaks. The peak areas are plotted below, normalized to the highest peak produced in each experiment.



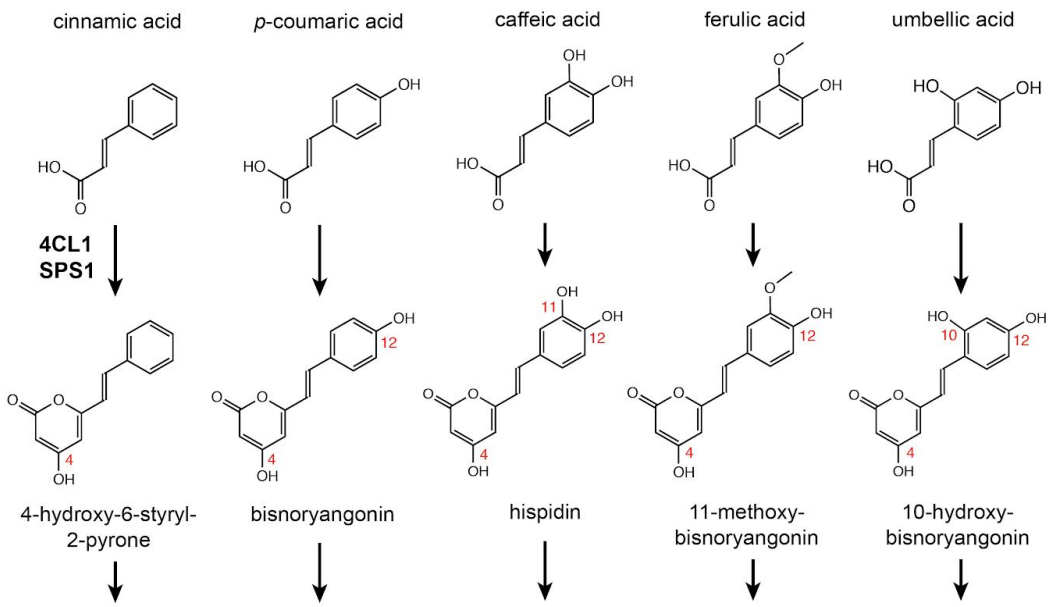
Since the substrates for each enzyme were presents in equimolar amounts, the difference in product abundance reflects the substrate preference of each enzyme.

Supplementary Note 4. Kavalactone OMT candidate screening.

We mined the assembled kava transcriptome using protein sequence BLAST (blastp) with *A. thaliana* flavone 3-*O*-methyltransferase (*At*COMT1) amino acid sequence as a bait. We found six putative OMT genes (designated *COMTL1-6*), which showed the following phylogeny in the context of other plant specialized metabolism OMTs:



We expressed the six *COMTL* genes in an *E. coli* expression vector and extracted the expressed protein using the B-PER detergent. The crude protein lysate was then used for *in vitro* enzyme assays in combination with purified *Pm*4CL1 and *Pm*SPS1 proteins. The methylated products were monitored by LC-MS, since each added methyl group adds 14 Da to the molecular mass. Below is a summary of all detected LC-MS peaks. Peak m/z values denoted in red color indicate a mass increase in comparison to the 'No enzyme' condition. In cases where multiple products were detected, the percentage of the peak area of each product is shown.

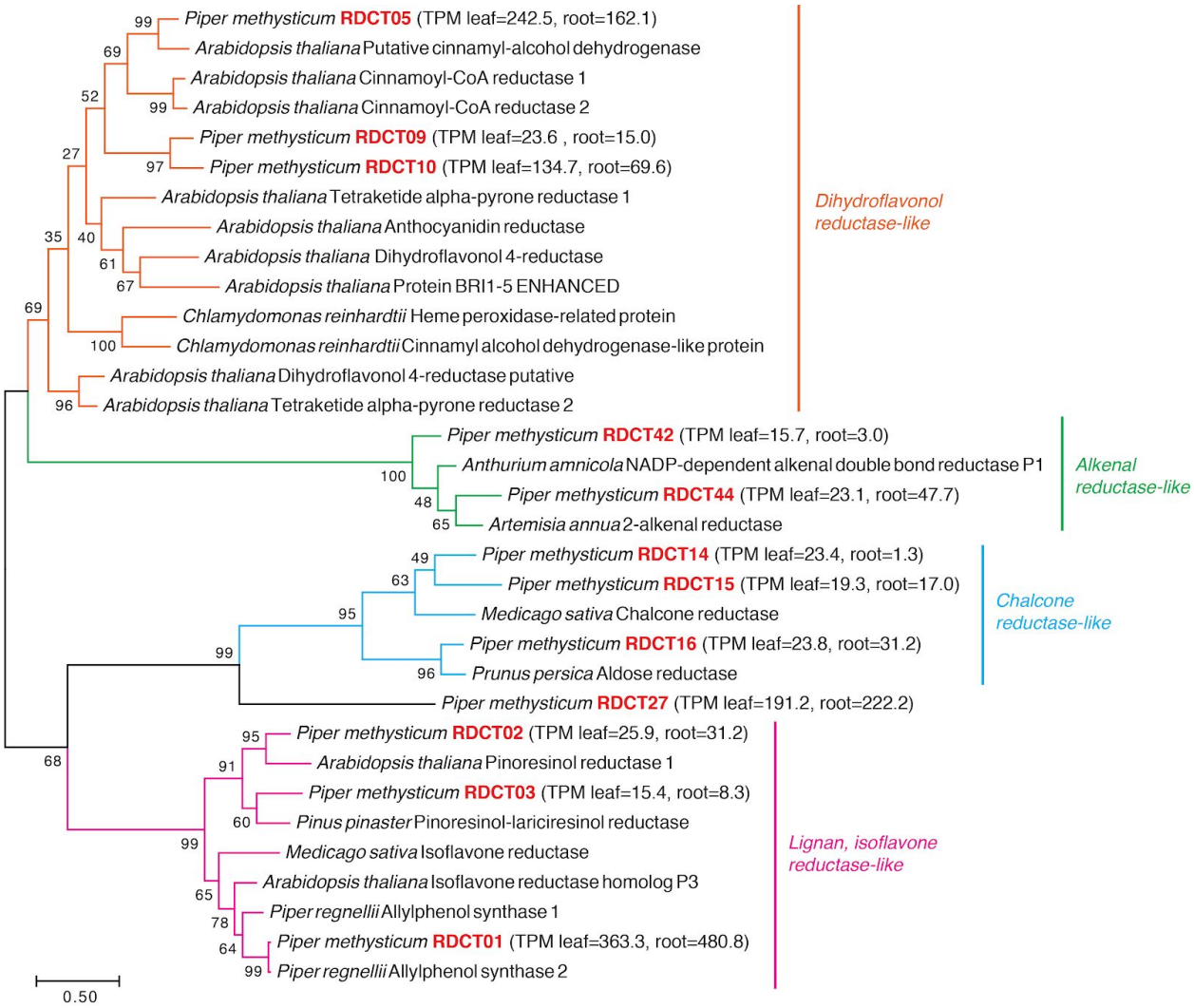


No enzyme	215 m/z @8.0 min	231 m/z @6.3 min	247 m/z @5.6 min	261 m/z @6.5 min	247 m/z @5.8 min
COMTL1 **(KOMT2)**	215 m/z @8.0 min	231 m/z @6.3 min	247 m/z @5.6 min	261 m/z @6.5 min	247 m/z @5.8 min 32% 261 m/z @6.7 min 68%
COMTL2	215 m/z @8.0 min	231 m/z @6.3 min	247 m/z @5.6 min	261 m/z @6.5 min	247 m/z @5.8 min
COMTL3	215 m/z @8.0 min	231 m/z @6.3 min	247 m/z @5.6 min	261 m/z @6.5 min	247 m/z @5.8 min
COMTL4 **(KOMT1)**	229 m/z @9.4 min	245 m/z @7.5 min 1% 245 m/z @7.6 min 3% 259 m/z @9.1 min 27% 259 m/z @9.5 min 69%	275 m/z @7.8 min 10% 289 m/z @8.5 min 32% 289 m/z @8.7 min 58%	275 m/z @7.7 min 23% 289 m/z @8.5 min 44% 289 m/z @8.7 min 33%	261 m/z @6.5 min 36% 275 m/z @7.7 min 64%
COMTL5	215 m/z @8.0 min	231 m/z @6.3 min	247 m/z @5.6 min	261 m/z @6.5 min	247 m/z @5.8 min
COMTL6	215 m/z @8.0 min	231 m/z @6.3 min	247 m/z @5.6 min	261 m/z @6.5 min	247 m/z @5.8 min
COMTL4+COMTL1 **(KOMT1+KOMT2)**	229 m/z @9.4 min	245 m/z @7.5 min 3% 245 m/z @7.6 min 6% 259 m/z @9.1 min 25% 259 m/z @9.5 min 66%	275 m/z @7.8 min 25% 289 m/z @8.5 min 21% 289 m/z @8.7 min 53%	275 m/z @7.7 min 33% 289 m/z @8.5 min 38% 289 m/z @8.7 min 29%	275 m/z @7.7 min 1% 289 m/z @9.3 min 69% 289 m/z @9.9 min 30%

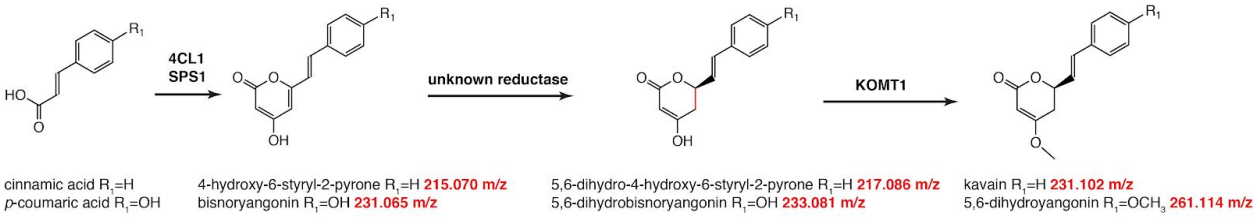
While all the starting substrates were *trans* isomers, yangonin-like compounds spontaneously isomerize in solution in the presence of light[9]. This explains the detection of two different chromatographic peaks with the same mass in several reactions (e.g., 259 m/z @9.1 min and @9.5 min, corresponding to *cis*- and *trans*-yangonin, respectively; Supplementary Table 3). Note that a combination of both COMTL4 (KOMT1) and COMTL1 (KOMT2) enzymes was required to fully methylate the three −OH groups of 10-hydroxybisnoryangonin.

Supplementary Note 5. Kavalactone reductase candidate screening.

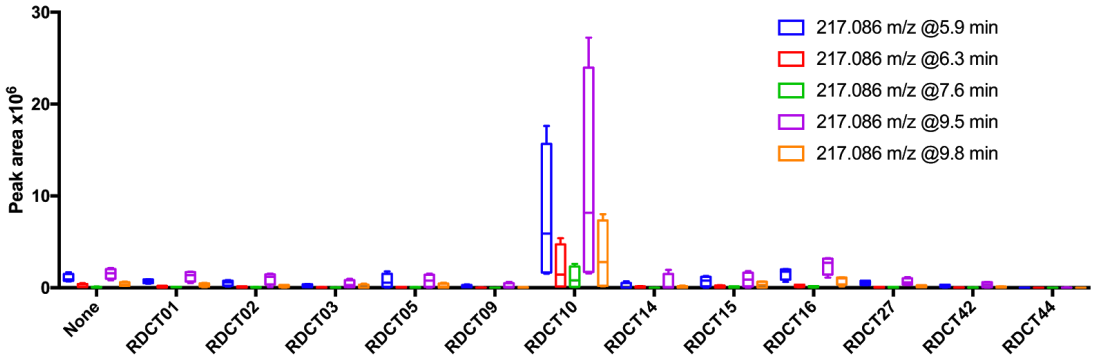
We mined the assembled kava transcriptome for sequences with the closest blastp hit in the UniprotKB/Swiss-Prot database[10] annotated as "reductase", starting from the transcripts with the highest expression (TPM) levels. We further mined the transcriptome for sequences homologous to *A. thaliana* dihydroflavonol 4-reductase (*At*DFRA), tetraketide alpha-pyrone reductase (*At*TKPR1), cinnamoyl-CoA reductase (*At*CCR1), and other characterized reductases involved in flavonoid and lignin metabolism[11,12]. The resulting list of ~50 reductase sequences was reduced by removing obvious homologs of primary metabolism enzymes and twelve candidate genes (designated *Pm*RDCT*xx*) were successfully cloned into the *N. benthamiana* expression vector pEAQ-HT:



We hypothesized that the kavalactone pathway including the reductase enzyme would progress as follows (note that the order of the reductase and OMT reactions was unclear at this point, but it would not change the identity of the expected products):



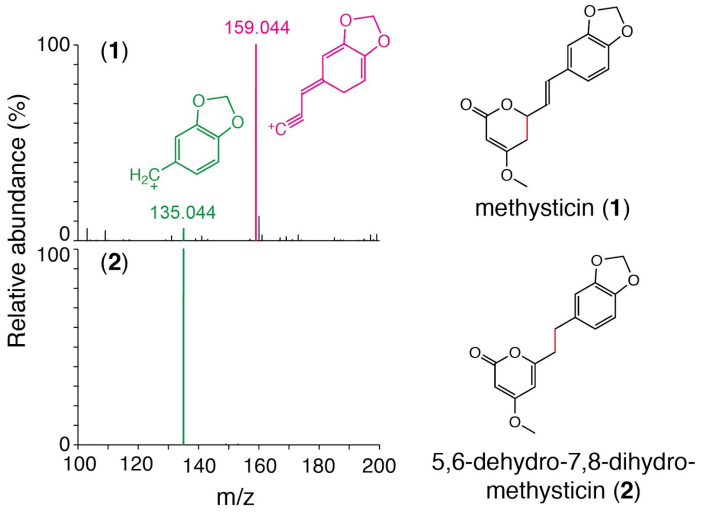
We expressed the twelve candidate genes in *N. benthamiana*, in combination with *Pm*SPS1 and *Pm*KOMT1, in four replicates (four leaves on different plants). Five days after infiltration, metabolites were extracted from the leaves and extracts analyzed by LC-MS. We detected 4,812 features in the datasets using MZmine 2[6]. Within these, we searched for features consistent with the target masses indicated in the pathway above (217.086 m/z, 233.081 m/z, 231.102 m/z, and 261.114 m/z). Using this approach, we found several features of 217.096 m/z elevated in plants expressing *Pm*RDCT10, as shown in box plots below.



We thus selected *Pm*RDCT10 for purification and activity verification using *in vitro* assays (Fig. 4A). Note that the *in vitro* assays established that the 217.086 m/z @7.6 min signal was the expected product (5,6-dihydro-4-hydroxy-6-styryl-2-pyrone). We assume the other detected features of the same mass correspond to in-source fragments of glycosylated or otherwise modified versions of this molecule.

Supplementary Note 6. MS/MS analysis of the main *Pm*CYP719A26 product.

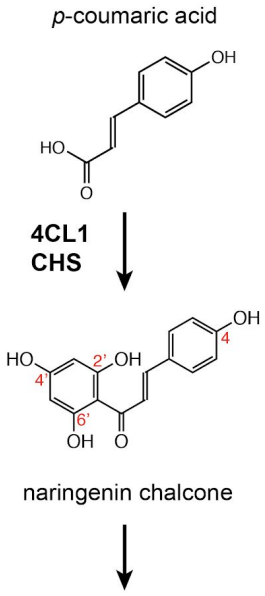
Expression of *PmCYP719A26* in N. benthamiana resulted in the production of a compound mass (275.091 m/z) and retention time (8.65 min) identical to that of methysticin (Fig. 4E). However, the MS/MS spectra of the two compounds were different. Based on the interpretation of the masses of the MS/MS fragments, we concluded that the produced compound was most likely 5,6-dehydro-7,8-dihydromethysticin:



Notably, the dominant 135.044 m/z fragment in the MS/MS pattern of the produced compound (**2**) cannot be rationalized without the presence of the methylenedioxy bridge, thus confirming the methylenedioxy bridge-forming activity of *Pm*CYP719A26. We assume that the C_7–C_8 single bond in the produced compound was likely reduced by an unknown reductase in *N. benthamiana*.

Supplementary Note 7. Flavokavain OMT candidate screening.

The biosynthesis of flavokavains involves OMT enzymes that act on the naringenin chalcone backbone produced by CHS (Fig. 1C). We screened the six previously identified kava OMTs (Note S3) in combination with purified *Pm*4CL1 and *Pm*CHS enzymes to test if any of the enzymes could methylate naringenin chalcone:



No enzyme	273 m/z @7.5 min
COMTL1 **(KOMT2)**	287 m/z @8.3 min
COMTL2	273 m/z @7.5 min
COMTL3	273 m/z @7.5 min
COMTL4 **(KOMT1)**	287 m/z @9.0 min
COMTL5	273 m/z @7.5 min
COMTL6	273 m/z @7.5 min

The enzymes KOMT1 and KOMT2 added a methyl group, producing LC-MS peaks of 287 m/z at two different retention times. This assay thus established that KOMT1 and KOMT2 can methylate different −OH groups of naringenin chalcone.

Supplementary Table 1. List of kava enzymes cloned from kava cDNA.

For each enzyme, its expression levels (TPM values)[13] in kava leaf and root are shown, as well as its rank among the expression levels of 285,753 total transcripts. The rank was calculated by estimating the TPM values for all transcripts using merged sequencing reads from both leaf and root, and sorting the transcripts by these TPM values. For each enzyme, the closest blastp match in the UniprotKB/Swiss-Prot database[10] is shown along with % identity of the amino acid sequence.

Enzyme	Expression levels			Closest blastp match in UniProtKB/Swiss-Prot
	Leaf TPM	Root TPM	Rank / 285,753	
*Pm*4CL1	7.2	37.5	31890	4-coumarate-CoA ligase 1 (*Nicotiana tabacum*), 74% id
*Pm*CHSL1 (*Pm*SPS1)	44.7	2.0	8918	Chalcone synthase 2 (*Citrus sinensis*), 80% id
*Pm*CHSL2 (*Pm*SPS2)	309.6	13.7	550	Chalcone synthase 2 (*Citrus sinensis*), 76% id
*Pm*CHSL3 (*Pm*CHS)	4.7	1.9	78535	Chalcone synthase 3 (*Ruta graveolens*), 87% id
*Pm*COMTL1 (*Pm*KOMT2)	18.1	3.8	80564	Trans-resveratrol di-O-methyltransferase (*Vitis vinifera*), 48% id
*Pm*COMTL2	0.0	2.9	73113	Trans-resveratrol di-O-methyltransferase (*Vitis vinifera*), 45% id
*Pm*COMTL3	7.1	8.1	56758	(RS)-norcoclaurine 6-O-methyltransferase (*Coptis japonica*), 65% id
*Pm*COMTL4 (*Pm*KOMT1)	862.7	16.1	**218**	Caffeic acid 3-O-methyltransferase 1 (*Populus tremuloides*), 52% id
*Pm*COMTL5	20.2	25.8	7941	Trans-resveratrol di-O-methyltransferase (*Vitis vinifera*), 49% id
*Pm*COMTL6	4.6	25.7	13609	(RS)-norcoclaurine 6-O-methyltransferase (*Coptis japonica*), 41% id
*Pm*RDCT01	363.3	480.8	582	Isoflavone reductase-like protein (*Olea europaea*), 71% id
*Pm*RDCT02	25.9	31.2	28332	Pinoresinol-lariciresinol reductase 2 (*Linum usitatissimum*), 76% id
*Pm*RDCT03	15.4	8.3	25238	Pinoresinol-lariciresinol reductase (*Thuja plicata*), 58% id
*Pm*RDCT05	242.9	162.1	593	Cinnamoyl-CoA reductase 1 (*Arabidopsis thaliana*), 53% id
*Pm*RDCT09	23.6	15.0	10813	Cinnamoyl-CoA reductase 1 (*Oryza sativa*), 43% id
*Pm*RDCT10 (*Pm*KLR1)	134.7	69.6	4212	Cinnamoyl-CoA reductase 1 (*Oryza sativa*), 42% id
*Pm*RDCT14	23.4	1.3	19569	Codeinone reductase 2 (*Papaver somniferum*), 59% id
*Pm*RDCT15	19.3	17.0	24773	Codeinone reductase 2 (*Papaver somniferum*), 56% id
*Pm*RDCT16	23.8	31.2	25553	Aldo-keto reductase fam. 4 mem. C10 (*Arabidopsis thaliana*), 42% id
*Pm*RDCT27	191.2	222.2	741	Probable aldo-keto reductase 2 (*Oryza sativa*), 81% id
*Pm*RDCT42	15.7	3.0	27563	2-alkenal reductase (*Nicotiana tabacum*), 63% id
*Pm*RDCT44	23.1	47.7	6785	2-alkenal reductase (*Nicotiana tabacum*), 68% id
*Pm*CYP719A26 (*Pm*MTS1)	217.9	2.0	2393	(S)-canadine synthase (*Coptis japonica*), 49% id

Supplementary Table 2. Crystallography data collection and refinement statistics.

Statistics for the highest-resolution shell are shown in parentheses.

	*Pm*SPS1	*Pm*CHS	*Pm*KLR1
Wavelength	0.9793 Å	0.9793 Å	0.9793 Å
Resolution range	82.2 - 2.7 (2.797 - 2.7)	62.7 - 2.91 (3.014 - 2.91)	62.13 - 2.8 (2.9 - 2.8)
Space group	P 31	P 21 21 21	P 21 21 21
Unit cell	164.4 164.4 87.1 90 90 120	50.9 120.6 125.4 90 90 90	63.2 63.2 333.7 90 90 90
Total reflections	3027278 (290476)	1083656 (107234)	2489595 (244476)
Unique reflections	72283 (7187)	17602 (1726)	34057 (3212)
Multiplicity	41.9 (40.2)	61.6 (62.1)	73.1 (73.2)
Completeness (%)	99.59 (99.31)	99.95 (100.00)	96.46 (94.76)
Mean I/sigma(I)	6.90 (1.30)	12.54 (2.85)	25.67 (2.62)
Wilson B-factor		44.73	
R-merge	1.406 (4.383)	0.926 (2.025)	0.7669 (2.471)
R-meas	1.424 (4.445)	0.9341 (2.043)	0.7725 (2.489)
R-pim	0.2252 (0.7278)	0.1204 (0.2639)	0.09055 (0.2895)
CC1/2	0.775 (0.157)	0.949 (0.614)	0.908 (0.701)
CC*	0.935 (0.521)	0.987 (0.872)	0.975 (0.908)
Reflections used in refinement	72027 (7187)	17602 (1726)	32903 (3183)
Reflections used for R-free	1987 (199)	1771 (179)	3323 (331)
R-work	0.2111 (0.2718)	0.2615 (0.3382)	0.2381 (0.3370)
R-free	0.2606 (0.3123)	0.3156 (0.3688)	0.2790 (0.3878)
CC(work)	0.668 (0.224)	0.921 (0.694)	0.825 (0.417)
CC(free)	0.403 (0.278)	0.881 (0.694)	0.816 (0.193)
Number of non-hydrogen atoms	18141	5880	10204
macromolecules	17603	5846	9963
ligands	135		192
solvent	403	34	49
Protein residues	2293	762	1270
RMS(bonds)	0.004	0.002	0.003
RMS(angles)	0.62	0.53	0.65
Ramachandran favored (%)	92.63	94.20	94.00
Ramachandran allowed (%)	6.45	5.80	5.52
Ramachandran outliers (%)	0.92	0.00	0.48
Rotamer outliers (%)	0.05	1.59	2.88
Clashscore	14.56	24.22	23.84
Average B-factor	38.58	47.80	74.91
macromolecules	38.46	47.89	75.36
ligands	67.33		58.23
solvent	34.05	32.71	47.82

Supplementary Table 3. Identified LC-MS peaks.

List of identified LC-MS peaks relevant to this study. The retention time (RT) of each peak is indicated based on the C18 chromatography method described in Methods.

m/z	RT (min)	Ion type	Metabolite	Confirmed by
273.076	7.50	$[M+H]^+$	naringenin	pure standard
273.076	7.40	$[M+H]^+$	naringenin chalcone	pure standard
273.076	5.86	$[M+H]^+$	*p*-coumaroyltriacetic acid lactone (CTAL)	MS/MS
229.086	9.43	$[M+H]^+$	desmethoxyyangonin	pure standard
231.102	8.85	$[M+H]^+$	(±)-kavain	pure standard
233.117	8.95	$[M+H]^+$	7,8-dihydrokavain	pure standard
259.096	9.10	$[M+H]^+$	*cis*-yangonin	pure standard, cis/trans determined based on published elution order[9]
259.096	9.50	$[M+H]^+$	*trans*-yangonin	pure standard, cis/trans determined based on published elution order[9]
275.091	8.65	$[M+H]^+$	methysticin	pure standard
277.107	8.67	$[M+H]^+$	7,8-dihydromethysticin	pure standard
275.091	8.65	$[M+H]^+$	5,6-dehydro-7,8-dihydromethysticin	MS/MS, see Supplementary Note 6
315.123	10.76	$[M+H]^+$	flavokavain A	pure standard
285.112	10.80	$[M+H]^+$	flavokavain B	pure standard
301.107	10.12	$[M+H]^+$	flavokavain C	pure standard
215.070	8.00	$[M+H]^+$	4-hydroxy-6-styryl-2-pyrone	accurate mass, *in vitro* enzyme assay
231.065	6.28	$[M+H]^+$	bisnoryangonin	MS/MS, see Supplementary Note 1
193.086	6.22	$[M+H]^+$	dehydrozingerone	pure standard
344.170	4.10	$[M+NH_4]^+$	raspberryketone glucoside	pure standard
349.125	4.10	$[M+Na]^+$	raspberryketone glucoside	pure standard
247.060	5.60	$[M+H]^+$	hispidin	accurate mass, *in vitro* enzyme assay
261.076	6.60	$[M+H]^+$	11-methoxybisnoryangonin	accurate mass, *in vitro* enzyme assay
247.060	5.90	$[M+H]^+$	10-hydroxybisnoryangonin	accurate mass, *in vitro* enzyme assay
217.086	7.60	$[M+H]^+$	5,6-dihydro-4-hydroxy-6-styryl-2-pyrone	accurate mass, *in vitro* enzyme assay
217.086	7.80	$[M+H]^+$	7,8-dihydro-4-hydroxy-6-styryl-2-pyrone	accurate mass, *in vitro* enzyme assay
233.081	6.40	$[M+H]^+$	5,6-dihydrobisnoryangonin	accurate mass, *in vitro* enzyme assay
233.081	5.84	$[M+H]^+$	7,8-dihydrobisnoryangonin	accurate mass, *in vitro* enzyme assay
579.170	5.20	$[M+H]^+$	kaempferol 3-*O*-α-l-rhamnopyranoside-7-*O*-α-l-rhamnopyranoside[1]	accurate mass
595.166	4.70	$[M+H]^+$	kaempferol 3-*O*-β-d-glucopyranoside-7-*O*-α-l-rhamnopyranoside[1]	accurate mass

Supplementary Table 4. List of oligonucleotide primers used in this study.

Primer	Used for	Sequence (5′−3′)
TP0005	Pm4CL1 → pHis8-4 (F)	GAAAACTTGTACTTCCAGGCCCATGGCATGAAGATGGTAGTAGACACTATTGCTACT
TP0006	Pm4CL1 → pHis8-4 (R)	CTCGAATTCGGATCCGCCATGGCTACTTGGGTATACCAGCTGCCAATCTTCC
TP0001	PmSPS1 → pHis8-4 (F)	GAAAACTTGTACTTCCAGGCCCATGGCATGTCGAAGACGGTGGAGGATCGGGCAGCG
TP0002	PmSPS1 → pHis8-4 (R)	CTCGAATTCGGATCCGCCATGGTTATCTGCCTTCATCGGCTTTGGGGATGGG
TP0003	PmSPS2 → pHis8-4 (F)	GAAAACTTGTACTTCCAGGCCCATGGCATGTCGAAGATGGTGGAGGAGCATTGGGCA
TP0004	PmSPS2 → pHis8-4 (R)	CTCGAATTCGGATCCGCCATGGTTAATTGGCTGTCGAGATGGGTACGCTACG
TP0080	PmCHS → pHis8-4 (F)	GTGAAAACTTGTACTTCCAGGCCCATGGCATGTCGAAGACCGTAGAGGAG
TP0081	PmCHS → pHis8-4 (R)	GGAGCTCGAATTCGGATCCGCCATGGTTAGTTGGCCTCGGCGATG
ADA0032	PmSPS1 → pEarleyGate 100 (F)	TTCATTTCATTTGGAGAGGACACGCTCGAGATGTCGAAGACGGTGGAGG
ADA0033	PmSPS1 → pEarleyGate 100 (R)	CTAGTCCCGGGTCTTAATTAACTCTAGATTATCTGCCTTCATCGGCTTTGG
ADA0034	PmSPS2 → pEarleyGate 100 (F)	TTCATTTCATTTGGAGAGGACACGCTCGAGATGTCGAAGATGGTGGAGGAGC
ADA0035	PmSPS2 → pEarleyGate 100 (R)	CTAGTCCCGGGTCTTAATTAACTCTAGATTAATTGGCTGTCGAGATGGGTACG
ADA0036	PmCHS → pEarleyGate 100 (F)	TTCATTTCATTTGGAGAGGACACGCTCGAGATGTCGAAGACCGTAGAGGAGATTTGG
ADA0037	PmCHS → pEarleyGate 100 (R)	CTAGTCCCGGGTCTTAATTAACTCTAGATTAGTTGGCCTCGGCGATGGG
TP0154	PmSPS1 qRT-PCR (F)	CTGGGCTTGTGTCGATGAA
TP0155	PmSPS1 qRT-PCR (R)	GACTTGAGCTTCTCCTCCTTTAG
TP0254	PmSPS2 qRT-PCR (F)	CTGCTGGAGAAGAACCCTAAC
TP0255	PmSPS2 qRT-PCR (R)	CCTTGGTTGCAGCTTCTTTG
TP0258	PmCHS qRT-PCR (F)	CGACGACTGGAACTCGATATTT
TP0259	PmCHS qRT-PCR (R)	TGTCTTCAGCTTCTCCACTTTC
JKW0444	At1g13320 qRT-PCR (F)	TAACGTGGCCAAAATGATGC
JKW0445	At1g13320 qRT-PCR (R)	GTTCTCCACAACCGCTTGGT
TP0106	Pm4CL1 → pEAQ-HT (F)	TTCTGTATATTCTGCCCAAATTCGCGACCGGTATGAAGATGGTAGTAGACACTATTG
TP0107	Pm4CL1 → pEAQ-HT (R)	AATTTAATGAAACCAGAGTTAAAGGCCTCGAGCTACTTGGGTATACCAGCTGCCAATC
TP0100	PmSPS1 → pEAQ-HT (F)	TTCTGTATATTCTGCCCAAATTCGCGACCGGTATGTCGAAGACGGTGGAGGATC
TP0101	PmSPS1 → pEAQ-HT (R)	AATTTAATGAAACCAGAGTTAAAGGCCTCGAGTTATCTGCCTTCATCGGCTTTGG
TP0102	PmSPS2 → pEAQ-HT (F)	TTCTGTATATTCTGCCCAAATTCGCGACCGGTATGTCGAAGATGGTGGAGGAGCATTG
TP0103	PmSPS2 → pEAQ-HT (R)	AATTTAATGAAACCAGAGTTAAAGGCCTCGAGTTAATTGGCTGTCGAGATGGGTAC
TP0014	PmCOMTL1 → pHis8-4 (F)	GAAAACTTGTACTTCCAGGCCCATGGCATGAATGATCAAGAGTTGCATGGATACTCAC
TP0015	PmCOMTL1 → pHis8-4 (R)	CTCGAATTCGGATCCGCCATGGCTAAGGGTAGAGCTCGATAAGCGAACTCGTTC
TP0016	PmCOMTL2 → pHis8-4 (F)	GAAAACTTGTACTTCCAGGCCCATGGCATGGCCGCAATGACAATATCCAAGG
TP0017	PmCOMTL2 → pHis8-4 (R)	CTCGAATTCGGATCCGCCATGGTCAATGATACAGCTCAATAACGGATCGAAC
TP0018	PmCOMTL3 → pHis8-4 (F)	GAAAACTTGTACTTCCAGGCCCATGGCATGGAAGCCCACAGCGATGGAAGC
TP0019	PmCOMTL3 → pHis8-4 (R)	CTCGAATTCGGATCCGCCATGGTTAGTAGGGAAAAGCCTCTATGACAGATTGGGC
TP0061	PmCOMTL4 → pHis8-4 (F)	GAAAACTTGTACTTCCAGGCCCATGGCATGGAGCAAGCTGTGTTCAAAGACC
TP0062	PmCOMTL4 → pHis8-4 (R)	CTCGAATTCGGATCCGCCATGGTTATTTGTGAAATTCCATAACCCACACGCCATAG
TP0065	PmCOMTL5 → pHis8-4 (F)	GAAAACTTGTACTTCCAGGCCCATGGCATGAGCTTGGAGAGGGGTAACATTG
TP0066	PmCOMTL5 → pHis8-4 (R)	CTCGAATTCGGATCCGCCATGGTCATGGGTACAATTCGATCACAGATCGGAG
TP0067	PmCOMTL6 → pHis8-4 (F)	GAAAACTTGTACTTCCAGGCCCATGGCATGAGTTCCGTTGTTGCACCAAATG
TP0068	PmCOMTL6 → pHis8-4 (R)	CTCGAATTCGGATCCGCCATGGTTACCCCCTGTGATTGCCATCATCAC
TP0116	PmKOMT1 → pEAQ-HT (F)	TTCTGTATATTCTGCCCAAATTCGCGACCGGTATGGAGCAAGCTGTGTTCAAAGACC
TP0117	PmKOMT1 → pEAQ-HT (R)	AATTTAATGAAACCAGAGTTAAAGGCCTCGAGTTATTTGTGAAATTCCATAACCCAC
TP0118	PmRDCT01 → pEAQ-HT (F)	TTCTGTATATTCTGCCCAAATTCGCGACCGGTATGGCAGAGAAGAGCAAGATCTTG
TP0119	PmRDCT01 → pEAQ-HT (R)	AATTTAATGAAACCAGAGTTAAAGGCCTCGAGTCACACTAATCGGTCGAGGAATTC
TP0120	PmRDCT02 → pEAQ-HT (F)	TTCTGTATATTCTGCCCAAATTCGCGACCGGTATGGAGAAAAGCAGGGTTCTGGTGG

TP0121	PmRDCT02 → pEAQ-HT (R)	AATTTAATGAAACCAGAGTTAAAGGCCTCGAGCTATATATATCTTTTGAGATAGTC
TP0150	PmRDCT03 → pEAQ-HT (F)	TTCTGTATATTCTGCCCAAATTCGCGACCGGTATGGCCGACATGAAAGACAGGGTGC
TP0123	PmRDCT03 → pEAQ-HT (R)	AATTTAATGAAACCAGAGTTAAAGGCCTCGAGTTAAACGAAGCGCTTGAGGTACTCC
TP0124	PmRDCT05 → pEAQ-HT (F)	TTCTGTATATTCTGCCCAAATTCGCGACCGGTATGGCCGCCGGTGAAGGGAAGATCG
TP0125	PmRDCT05 → pEAQ-HT (R)	AATTTAATGAAACCAGAGTTAAAGGCCTCGAGTTAAAAACTCACGAATTTCTTCTCC
TP0126	PmRDCT09 → pEAQ-HT (F)	TTCTGTATATTCTGCCCAAATTCGCGACCGGTATGGCCGACGACAAGAGATCCTGC
TP0127	PmRDCT09 → pEAQ-HT (R)	AATTTAATGAAACCAGAGTTAAAGGCCTCGAGTTAAACCATTTCCCTCTTCGTTATAC
TP0128	PmRDCT10 → pEAQ-HT (F)	TTCTGTATATTCTGCCCAAATTCGCGACCGGTATGGAGACTGAGAGGAAGTCCAGG
TP0129	PmRDCT10 → pEAQ-HT (R)	AATTTAATGAAACCAGAGTTAAAGGCCTCGAGCTATTCATTTTCCAGCAAACCATGC
TP0142	PmRDCT14 → pEAQ-HT (F)	TTCTGTATATTCTGCCCAAATTCGCGAATGGCGACGAAAATTCCAGAGATG
TP0143	PmRDCT14 → pEAQ-HT (R)	AATTTAATGAAACCAGAGTTAAAGGCCTTATACGTCCCCATCCCAAAACTCG
TP0144	PmRDCT15 → pEAQ-HT (F)	TTCTGTATATTCTGCCCAAATTCGCGAATGGGCACGAAATTGGTCGG
TP0145	PmRDCT15 → pEAQ-HT (R)	AATTTAATGAAACCAGAGTTAAAGGCCCTACATTTCTCCATCCCAGAACTCTTCAG
TP0146	PmRDCT16 → pEAQ-HT (F)	TTCTGTATATTCTGCCCAAATTCGCGAATGAGAGAAAGAGGAGGCACATGTTC
TP0147	PmRDCT16 → pEAQ-HT (R)	AATTTAATGAAACCAGAGTTAAAGGCCCTAAGATTGTTCATCACCATCTGTCTCCG
TP0148	PmRDCT27 → pEAQ-HT (F)	TTCTGTATATTCTGCCCAAATTCGCGAATGGCGGCTGAAGTTGGTAGG
TP0149	PmRDCT27 → pEAQ-HT (R)	AATTTAATGAAACCAGAGTTAAAGGCCTCATTGTCCCTTCCATGAAGCTAGTG
TP0168	PmRDCT42 → pEAQ-HT (F)	TTCTGTATATTCTGCCCAAATTCGCGACCGGTATGGCCGAGGTGCCAAACAAGC
TP0169	PmRDCT42 → pEAQ-HT (R)	AATTTAATGAAACCAGAGTTAAAGGCCTCGAGTCAATTCTCATATGCTACTCGGATTAC
TP0170	PmRDCT44 → pEAQ-HT (F)	TTCTGTATATTCTGCCCAAATTCGCGACCGGTATGGGAGCAGTGGTGCCCAAC
TP0171	PmRDCT44 → pEAQ-HT (R)	AATTTAATGAAACCAGAGTTAAAGGCCTCGAGTCACTCTGGTGCAATCACAACAACCTG
TP0094	PmKLR1 → pHis8-4 (F)	GTGAAAACTTGTACTTCCAGGCCCATGGCATGGAGACTGAGAGGAAGTCCAG
TP0095	PmKLR1 → pHis8-4 (R)	CTCGAATTCGGATCCGCCATGGCTATTCATTTTCCAGCAAACCATGCTC
ADA0014	PmCYP719A26 → pYeDP60 (F)	CACAAATACACACACTAAATTACCGGATCCATGGAGCAAGCTCAATGG
ADA0015	PmCYP719A26 → pYeDP60 (R)	TAGAGACATGGGAGATCCCCCGCGGAATTCTCATGCACGTGGAACGATTTTG
TP0132	PmCYP719A26 → pEAQ-HT (F)	TTCTGTATATTCTGCCCAAATTCGCGACCGGTATGGAGCAAGCTCAATGGGTCGAC
TP0133	PmCYP719A26 → pEAQ-HT (R)	AATTTAATGAAACCAGAGTTAAAGGCCTCGAGTCATGCACGTGGAACGATTTTGGC
CHS009	Pm4CL1 → pJKW1565 (F)	CTTTAAGAAGGAGAACTGCATCTCGAGGATGAAGATGGTAGTAGACACTATTGCTACTG
CHS010	Pm4CL1 → pJKW1565 (R)	CAACCGTACCGGATCCCCTGCAGGGCTACTTGGGTATACCAGCTGCCAATC
CHS011	PmSPS1 → pJKW1565 (F)	CTTTAAGAAGGAGACTGCCATAATGTCGAAGACGGTGGAGGATCG
CHS012	PmSPS1 → pJKW1565 (R)	GATCAAGCTTTTAATTAAGAGCTCATTATCTGCCTTCATCGGCTTTGGG
CHS050	PmSPS2 → pJKW1565 (F)	GTTTAACTTTAAGAAGGAGACTGCCATAATGTCGAAGATGGTGGAGGAG
CHS051	PmSPS2 → pJKW1565 (R)	GGCGATCAAGCTTTTAATTAAGAGCTCATTAATTGGCTGTCGAGATGGGTAC
MP0567	Pm4CL1 → p425TEF (F)	CAATCTAATCTAAGTTTTCTAGAACTAGTATGAAGATGGTAGTAGACACTATTGCTACTG
MP0568	Pm4CL1 → p425TEF (R)	CAGCCCGGGGGATCCACTAGTCTACTTGGGTATACCAGCTGCCAATC
TP0217	PmSPS1 → p426TEF (F)	GCATAGCAATCTAATCTAAGTTTTCTAGAACTAGTATGTCGAAGACGGTGGAGGATC
TP0218	PmSPS1 → p426TEF (R)	CAGCCCGGGGGATCCACTAGTTTATCTGCCTTCATCGGCTTTGGGGATGG
CHS001	PmSPS2 → p426TEF (F)	GCATAGCAATCTAATCTAAGTTTTCTAGAACTAGTATGTCGAAGATGGTGGAGGAGC
CHS002	PmSPS2 → p426TEF (R)	CAGCCCGGGGGATCCACTAGTTCATTAATTGGCTGTCGAGATGGGTAC

Supplementary Table 5. List of expression plasmids constructed in this study that are available from Addgene.

Plasmid	Backbone	Insert(s)
pJKW0825	pHis8-4	*Pm*4CL1
pJKW0770	pHis8-4	*Pm*SPS1
pJKW0776	pHis8-4	*Pm*SPS2
pJKW1077	pHis8-4	*Pm*CHS
pJKW0957	pHis8-4	*Pm*KOMT1
pJKW0826	pHis8-4	*Pm*KOMT2
pJKW1131	pHis8-4	*Pm*KLR1
pJKW1148	pYeDP60[14]	*Pm*CYP719A26
pJKW1254	pEAQ-HT[15]	*Pm*4CL1
pJKW1251	pEAQ-HT[15]	*Pm*SPS1
pJKW1252	pEAQ-HT[15]	*Pm*SPS2
pJKW1253	pEAQ-HT[15]	*Pm*CHS
pJKW1259	pEAQ-HT[15]	*Pm*KOMT1
pJKW1258	pEAQ-HT[15]	*Pm*KOMT2
pJKW1265	pEAQ-HT[15]	*Pm*KLR1
pJKW1270	pEAQ-HT[15]	*Pm*CYP719A26
pJKW1413	p425TEF[16]	*Pm*4CL1
pJKW1538	p426TEF[16]	*Pm*SPS1
pJKW1547	p426TEF[16]	*Pm*SPS2
pJKW1548	p426TEF[16]	*Pm*CHS
pJKW1574	pJKW1565	*Pm*4CL1 + *Pm*SPS1
pJKW1658	pJKW1565	*Pm*4CL1 + *Pm*SPS2
pJKW1659	pJKW1565	*Pm*4CL1 + *Pm*CHS

Supplementary References

1. Veit, M. & Pauli, G. F. Major flavonoids from Arabidopsis thaliana leaves. *J. Nat. Prod.* **62**, 1301–1303 (1999).
2. Murrell, B. *et al.* Detecting individual sites subject to episodic diversifying selection. *PLoS Genet.* **8**, e1002764 (2012).
3. Böcker, S., Letzel, M. C., Lipták, Z. & Pervukhin, A. SIRIUS: decomposing isotope patterns for metabolite identification. *Bioinformatics* **25**, 218–224 (2009).
4. Kim, S. *et al.* PubChem Substance and Compound databases. *Nucleic Acids Res.* **44**, D1202–13 (2016).
5. Dührkop, K., Shen, H., Meusel, M., Rousu, J. & Böcker, S. Searching molecular structure databases with tandem mass spectra using CSI:FingerID. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 12580–12585 (2015).
6. Pluskal, T., Castillo, S., Villar-Briones, A. & Oresic, M. MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics* **11**, 395 (2010).
7. Lê Cao, K.-A., Boitard, S. & Besse, P. Sparse PLS discriminant analysis: biologically relevant feature selection and graphical displays for multiclass problems. *BMC Bioinformatics* **12**, 253 (2011).
8. Xia, J., Sinelnikov, I. V., Han, B. & Wishart, D. S. MetaboAnalyst 3.0—making metabolomics more meaningful. *Nucleic Acids Res.* **43**, W251–W257 (2015).
9. Smith, R. M., Thakrar, H., Arowolo, T. A. & Shafi, A. A. High-performance liquid chromatography of kava lactones from piper methysticum. *J. Chromatogr. A* **283**, 303–308 (1984).
10. Boutet, E., Lieberherr, D., Tognolli, M., Schneider, M. & Bairoch, A. UniProtKB/Swiss-Prot. in *Plant Bioinformatics: Methods and Protocols* (ed. Edwards, D.) 89–112 (Humana Press, 2007).
11. Saito, K. *et al.* The flavonoid biosynthetic pathway in Arabidopsis: structural and genetic diversity. *Plant Physiol. Biochem.* **72**, 21–34 (2013).
12. Weng, J.-K. & Chapple, C. The origin and evolution of lignin biosynthesis. *New Phytol.* **187**, 273–285 (2010).
13. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).
14. Urban, P., Cullin, C. & Pompon, D. Maximizing the expression of mammalian cytochrome P-450 monooxygenase activities in yeast cells. *Biochimie* **72**, 463–472 (1990).
15. Peyret, H. & Lomonossoff, G. P. The pEAQ vector series: the easy and quick way to produce recombinant proteins in plants. *Plant Mol. Biol.* **83**, 51–58 (2013).
16. Mumberg, D., Müller, R. & Funk, M. Yeast vectors for the controlled expression of heterologous proteins in different genetic backgrounds. *Gene* **156**, 119–122 (1995).