Independent evolution of rosmarinic acid biosynthesis in two sister families under the Lamidi clade of flowering plants

Olesya Levsh§§, Tomáš Pluská§§, Valentina Carballo§§, Andrew J. Mitchell§§, and Jing-Ke Weng§§‡‡

From the §Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139 and the ‡Whitehead Institute for Biomedical Research, Cambridge, Massachusetts 02142

As a means to maintain their sessile lifestyle amid challenging environments, plants produce an enormous diversity of compounds as chemical defenses against biotic and abiotic insults. The underpinning metabolic pathways that support the biosynthesis of these specialized chemicals in divergent plant species provide a rich arena for understanding the molecular evolution of complex metabolic traits. Rosmarinic acid (RA) is a phenolic natural product first discovered in plants of the mint family (Lamiaceae) and is recognized for its wide range of medicinal properties and potential applications in human dietary and medical interventions. Interestingly, the RA chemotype is present sporadically in multiple taxa of flowering plants as well as some hornworts and ferns, prompting the question whether its biosynthesis arose independently across different lineages. Here we report the elucidation of the RA biosynthetic pathway in Phacelia campanularia (desert bells). This species represents the borage family (Boraginaceae), an RA-producing family closely related to the Lamidi within the Lamidi clade. Using a multi-omics approach in combination with functional characterization of candidate genes both in vitro and in vivo, we found that RA biosynthesis in P. campanularia involves specific activities of a BAHD acyltransferase and two cytochrome P450 hydroxylases. Further phylogenetic and comparative structure–function analyses of the P. campanularia RA biosynthetic enzymes clearly indicate that RA biosynthesis has evolved independently at least twice in the Lamidi, an exemplary case of chemotypic convergence through disparate evolutionary trajectories.

Plants of the mint family (Lamiaceae, order Lamiales), such as basil, rosemary, lavender, and mint, are used worldwide as popular aromatic culinary herbs. A number of Lamiaceae species have also been used medicinally for centuries in countries such as China, Lebanon, and Uzbekistan (1–3). Some reported medicinal properties of Lamiaceae plants, including antimicrobial, anxiolytic, anti-inflammatory and anti-cancer properties, have been attributed to the rich and diverse specialized metabolites present in these plants (4–7).

Rosmarinic acid (RA),3 a specialized phenolic ester abundantly present in most Lamiaceae plants (Fig. 1), has garnered particular attention because of its multifaceted medicinal properties and potential for treating various human diseases (8–11). For example, RA is recognized for its antioxidant and neuroprotective properties (11, 12). Multiple animal studies have demonstrated the anti-inflammatory activity of RA through repression of the NF-κB and STAT3 signaling pathways (13, 14). Additionally, RA is an inhibitor of GABA transaminase and therefore serves as an anxiolytic agent by modulating GABA levels in the central nervous system (15). RA is also an inhibitor of indoleamine-pyrdrole 2,3-dioxygenase, which is an immune checkpoint molecule and has recently emerged as an important target for cancer immunotherapy (16). The potential utility of RA as a therapeutic agent and in other biotechnological applications has propelled research regarding RA biosynthesis in Lamiaceae (17).

The resolved RA biosynthetic pathway from Coleus blumei (Lamiaceae) consists of a specialized acyltransferase, rosmarinic acid synthase (RAS), and a cytochrome P450 meta-hydroxylase, CYP98A14 (18, 19). First, RAS catalyzes the transfer of the 4-coumaroyl acyl group from 4-coumaroyl-CoA to the acyl acceptor molecule (R)-3-(4-hydroxyphenyl)lactate to produce 4-coumaroyl-(R)-3-(4-hydroxyphenyl)lactate, and CYP98A14 subsequently catalyzes a single meta-hydroxylation reaction on each of the aromatic rings of 4-coumaroyl-(R)-3-(4-hydroxyphenyl)lactate to yield RA (Fig. 2A) (18, 19).

Although common in the Lamiaceae, RA is a specialized metabolite that is distributed only sporadically throughout the plant kingdom (Fig. 1), primarily within angiosperms but also in some ferns and hornworts (20). This scattered taxonomic distribution pattern in land plants suggests the complex evolutionary history of RA biosynthesis. RA biosynthesis could have

This work was supported by Pew Scholars Program in the Biomedical Sciences Grant 27345, Searle Scholars Program Grant 15-SSP-162, the Family Larsson-Rosenquist Foundation, and the Keck Foundation. J.-K. W. is a co-founder, member of the Scientific Advisory Board, and shareholder of DoubleRainbow Biosciences, which develops biotechnologies related to natural products.

This article contains Figs. S1–S6, Tables S1–S6, and supporting Methods. The atomic coordinates and structure factors (code 6MK2) have been deposited in the Protein Data Bank (http://wwpdb.org/).

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EBI Data Bank with accession number(s) MH878831, MH878832, MH878833, MH878834, MH878835, MH878836, MH878837.

1 A Simons Foundation Fellow of the Helen Hay Whitney Foundation.

‡ To whom correspondence should be addressed. Tel.: 617-324-4921; E-mail: wengj@wi.mit.edu.

3 The abbreviations used are: RA, rosmarinic acid; Pc, Phacelia campanularia; Cb, Coleus blumei; RAS, rosmarinic acid synthase; HCT, hydroxycinnamoyl-CoA:shikimate hydroxycinnamoyltransferase; SHT, spermidine hydroxycinnamoyltransferase; HQT, hydroxycinnamoyl-CoAquinase hydroxycinnamoyltransferase; C3′H, 4-coumaroyl ester 3′-hydroxylase; VIGS, virally induced gene silencing; TRV, tobacco rattle virus; IPTG, isopropyl-1-thio-β-D-galactopyranoside.
emerged once in early land plants and subsequently been lost in many taxa. Independent loss of ancestral metabolic traits has been well-documented in the case of pale beach mice, where the light pigment phenotypes of beach mouse populations from different parts of the world have been found to be caused by independently arising mutations (21). Alternatively, RA biosynthesis might have evolved independently multiple times during land plant evolution, which has been a common evolutionary phenomenon in plant-specialized metabolism (22). To date, there is no sufficient evidence to distinguish between these two evolutionary scenarios.

Here we examined a family of plants within the Lamiids clade, closely related to Lamiaceae: the borage family (Boraginaceae, order Boraginales). Previous studies have demonstrated that many sampled members of the Boraginaceae also produce RA, like the Lamiaceae (20, 23–25). In contrast, plants from Solanales, an order immediately sister to Boraginales (Fig. 1), are not RA producers (20). Notably, no other family under the order Lamiales besides the Lamiaceae has been shown to produce RA (20). This intriguing distribution of RA in taxa closely related to and within the Lamiids led us to investigate the biosynthetic origin of RA in Boraginaceae. Using a candidate-based approach, we discovered the RA biosynthetic pathway in the Boraginaceae plant Phacelia campanularia and provide substantial evidence supporting independent evolution of RA in two closely related Lamiid families.

**Results**

The Boraginaceae RAS is a highly expressed BAHD acyltransferase paralogous to spermidine hydroxycinnamoyltransferase

To study RA biosynthesis in the Boraginaceae family, a little-studied family of nonmodel plants, we selected *P. campanularia*, colloquially known as desert bells or California bluebells, as our experimental species (Fig. 1). *P. campanularia* grows easily in standard greenhouses, and its transcriptome has been sequenced previously through the 1000 Plants (1KP) initiative (26). Using LC high-resolution accurate-mass MS-based metabolic profiling, we confirmed that laboratory-grown *P. campanularia* produces RA (Fig. S1A).

We hypothesized that the Boraginaceae RAS, like that of Lamiaceae, is a relatively highly expressed BAHD acyltransferase and mined the *P. campanularia* transcriptome for genes that fit this criteria (Table S1) (18, 27, 28). Of the top 20 most highly expressed BAHD acyltransferase-encoding genes, the 15 that harbored full-length ORFs were cloned and transiently expressed in the leaf tissue of *Nicotiana benthamiana* using Agrobacterium-mediated infiltration (29). 5 days after infiltration, we assessed the accumulation of the immediate RAS product 4-coumaroyl-(R)-3-(4-hydroxyphenyl)lactate in the infiltrated *N. benthamiana* leaves by LC high-resolution accurate-mass MS. Of all 15 candidates tested, only the second most highly expressed *P. campanularia* BAHD gene yielded the RAS
product; it is hereafter referred to as PcRAS (Table S1 and Fig. 3, A and C).

Molecular phylogenetic analysis of PcRAS together with other related BAHD acyltransferases from a variety of plant species revealed a Boraginaceae-specific RAS clade distinct from the Lamiaceae RAS clade (Fig. 2B). Although Lamiaceae RAS enzymes appear to have evolved from the hydroxycinnamoyl-CoA:shikimatehydroxycinnamoyltransferases (HCTs), which produce 4-coumaroyl shikimate, a key intermediate in the conserved plant phenylpropanoid metabolism, Boraginaceae RAS enzymes share an ancestor with the spermidine hydroxycinnamoyltransferases (SHTs), which produce hydroxycinnamoylspermidine conjugates reported to accumulate in the pollen coat of Arabidopsis thaliana (30–34). Compared with most of the BAHD enzymes represented in the phylogeny, which accept negatively charged acyl acceptor substrates, SHT is unique because it prefers the positively charged acyl acceptor spermidine. Although BAHD acyltransferases are well-recognized for their relaxed substrate selectivity, they tend to be specific with regard to the charge identity of their acyl acceptor substrates, as this property is generally dictated by the electrostatic properties of the acyl acceptor binding pocket within the enzyme active site (35, 36). It is therefore intriguing that Boraginaceae RAS, which has evolved to accommodate the negatively charged acyl acceptor substrate (R)-3-(4-hydroxyphenyl)lactate, is closely related to SHT. Functional analysis demonstrated that PcRAS does not retain SHT function, suggesting that it has lost its ancestral activity (Fig. S2).

Our phylogenetic analysis also revealed that P. campanularia contains a presumed ortholog of hydroxycinnamoyl-CoA:quinane hydroxycinnamoyltransferase (HQT), which is involved in chlorogenic acid biosynthesis (37). Indeed, metabolic profiling of P. campanularia leaf tissue revealed the presence of chlorogenic acid (Fig. S1B). Our enzymatic assays further confirmed the activity of PcHQT (Fig. 4A).

PcRAS, PcCYP98A112, and PcCYP98A113 are sufficient to reconstitute RA biosynthesis in N. benthamiana

To complete our model of RA biosynthesis, we took an example from the Lamiaceae pathway, hypothesizing that the Boraginaceae 4-coumaroyl-(R)-3-(4-hydroxyphenyl)lactate meta-hydroxylases descended from the highly conserved 4-coumaroyl ester 3′-hydroxylase (C3′H) (19). C3′H belongs to the CYP98A family of P450s and catalyzes the first meta-hydroxylation of the key phenylpropanoid pathway intermediate 4-coumaroyl shikimate (34). Our phylogenetic analyses of P450s related to C3′H from a variety of plant species revealed two Boraginaceae-specific clades (Fig. 2C) distinct from the conserved C3′H clade and the clade containing known Lamiaceae 4-coumaroyl-(R)-3-(4-hydroxyphenyl)lactate meta-hydroxylases (19). We designated the two presumed P. campanularia orthologs from each of these Boraginaceae-specific clades PcCYP98A112 and PcCYP98A113, and the presumed P. campanularia ortholog of C3′H was designated PcCYP98A111 (Table S2).
To test the function of these newly identified *P. campanularia* P450s, we coinfiltulated combinations of *PcRAS*, *PcCYP98A112*, and *PcCYP98A113*, as well as *PcHQT* and *PcCYP98A111* as negative controls, into *N. benthamiana* leaves in the presence or absence of the RA precursors 4-coumarate and (R)-3-(4-hydroxyphenyl)lactate (*Fig. 3*). Although neither *PcHQT* nor *PcCYP98A111* produced any on-pathway RA intermediates, transgenic expression of *PcRAS* was sufficient to produce 4-coumaroyl-(R)-3-(4-hydroxyphenyl)lactate (*Fig. 3*, *A and B*). The in planta function of *PcRAS* was further corroborated by stable transgenic expression of *PcRAS* in *A. thaliana* (*Fig. 3*). Interestingly, the combination of *PcRAS* and *PcCYP98A113*, excluding *PcCYP98A112*, was sufficient to produce small quantities of RA. This observation suggests that an endogenous *N. benthamiana* meta-hydroxylase, perhaps the conserved C3/H11032, may be able to perform some degree of *PcCYP98A112* function, especially in the presence of excess substrate. This is consistent with the observation that *PcCYP98A112* is phylogenetically more closely related to C3’H (*Fig. 2C*). *PcCYP98A113*, on the other hand, exhibits a more distant phylogenetic pattern, reflected in the relative length of the branch to the CYP98A113 clade, and also appears to have a more specialized biochemical function.

**In vitro characterization of *P. campanularia* RA biosynthetic enzymes**

We expressed and purified recombinant *PcHCT*, *PcHQT*, and *PcRAS* from *E. coli* and carried out Michaelis–Menten kinetics assays (*Fig. 4A* and Table 1) to examine the substrate specificity of these enzymes toward a panel of acyl acceptor compounds in the presence of 4-coumaroyl-CoA as the common acyl donor at a saturating concentration, chosen based on previously reported *K_m* values for HCT and RAS enzymes (27, 38). As expected, *PcHCT*, *PcHQT*, and *PcRAS* were specific toward shikimate, quinate, and (R)-3-(4-hydroxyphenyl)lactate, respectively. However, we observed that both the *K_m* and
$V_{\text{max}}$ measured for PcRAS against (R)-3-(4-hydroxyphenyl)lactate were inferior relative to CbRAS (Table S3) and those reported in the literature for other Lamiaceae RASs, demonstrating that PcRAS may not be as catalytically efficient under our experimental conditions (27, 35). To rule out the possibility that an alternative substrate is preferred by PcRAS, we screened many possible variations in acyl donor and acyl acceptor compounds in accordance with possible RA pathway intermediates (Fig. S3). Additionally, we quantified the absolute RA content across multiple tissues of C. blumei (Lamiaceae) and P. campanularia, ruling out the possibility that P. campanularia produces significantly lower quantities of RA than a Lamiaceae plant (Table S4). Collectively, these results suggest that, despite kinetic parameters differing from those of Lamiaceae RASs, PcRAS is likely sufficient to support RA accumulation in vivo. Alternatively, there may be additional levels of regulation that improve the catalytic efficiency of PcRAS in vivo.

We additionally expressed PcCYP98A111, PcCYP98A112, and PcCYP98A113 in yeast, and purified yeast microsomes containing these recombinant P450s to carry out kinetics experiments against a panel of possible substrates (Fig. 4B). As expected, only PcCYP98A111 was able to convert 4-coumaroyl shikimate into caffeoyl shikimate, demonstrating the validity of our microsome assays and indicating that the two other P. campanularia CYP98 P450s have lost their ancestral function. Of the remaining enzymes, both exhibited 4-coumaroyl-(R)-3-(4-hydroxyphenyl)lactate meta-hydroxylase activity, although PcCYP98A112 and PcCYP98A113 were specific toward 4-coumaroyl-(R)-3-(3,4-dihydroxyphenyl)lactate and caffeoyl-(R)-3-(4-hydroxyphenyl)lactate, respectively (Fig. 4B and Fig. S4A). Additionally, we demonstrated the lack of activity of PcCYP98A112 and PcCYP98A113 against 4-coumarate or (R)-3-(4-hydroxyphenyl)lactate (Fig. S4B). Because neither enzyme is able to undergo two rounds of hydroxylation to produce substantial quantities of RA from 4-coumaroyl-(R)-3-(3,4-dihydroxyphenyl)lactate and caffeoyl-(R)-3-(4-hydroxyphenyl)lactate, respectively (Fig. 4B), these results suggest that PcCYP98A112 specifically catalyzes 3-hydroxylation of the acyl donor-derived phenolic ring, whereas PcCYP98A113 catalyzes 3′-hydroxylation of the ring derived from the acyl acceptor (Fig. 4C). Moreover, PcCYP98A112 or PcCYP98A113 can accept a substrate lacking any meta-hydroxylation as well as a substrate that has already been meta-hydroxylated once, suggesting lack of a precise order of catalysis downstream of the 4-coumaroyl-(R)-3-(4-hydroxyphenyl)lactate intermediate.

**Virally induced gene silencing (VIGS) of PcRAS in P. campanularia decreases RA content in leaves**

To our knowledge, no molecular genetics experimentation has been performed previously on any Boraginaceae plant. Therefore, we referred to VIGS, a technique used in other non-model plant systems, to assess gene function. VIGS leverages...
the host’s endogenous RNAi pathway to allow knockdown of a target gene of interest (39–42). We were further encouraged by a reported instance of tobacco mosaic virus replication in P. campanularia (43). Because of its availability and widespread use, we started with the tobacco rattle virus (TRV) VIGS system. To facilitate viral replication and maximize VIGS efficiency, we utilized the mechanical transmission technique, where we first propagated viral VIGS vectors in N. benthamiana and subsequently applied the viral leaf extract onto P. campanularia leaves (Fig. 5A) (44).

To assess whether TRV VIGS is a viable gene knockdown method in P. campanularia, we first attempted to silence P. campanularia phytoene desaturase (PcPDS), a highly conserved gene involved in carotenoid biosynthesis (45). Silencing PDS in other plants causes an albino phenotype, serving as an apparent visual cue of successful gene silencing (39). In our initial optimization of the PcPDS silencing experiment, we found that knockdown was most effective in 4-week-old P. campanularia plants and that efficacy ranged from nearly uniform knockdown throughout the leaf to, more commonly, sparse knockdown on a background of healthy leaf tissue (Fig. 5B). Therefore, although lacking desirable robustness, TRV VIGS offers a functional approach to assess gene function in P. campanularia.

Next we designed a VIGS construct to silence PcRAS as well as a construct containing GFP to use as a control. Because depleted RA accumulation lacks an associated visible phenotype as that of PcPDS knockdown, we could not target specific parts of the affected tissues for analysis. We therefore harvested VIGS-treated whole leaves for metabolomics analysis. Leaves inoculated with PcRAS VIGS produced significantly lower quantities of RA than those inoculated with a control, although PcRAS VIGS did not cause complete depletion of RA accumulation (Fig. 5C). It is possible that the existing RA pool prior to VIGS depletes slowly when PcRAS expression is knocked down. Additionally, the knockdown nature of RNAi and the lack of robustness in tissue penetrance of TRV VIGS in P. campanularia may contribute to this.

### Structural features of independently evolved Boraginaceae and Lamiales RAS enzymes

To elucidate the structural elements dictating the common biochemistry of two independently evolved RAS enzymes, we solved the X-ray crystal structure of C. blumei RAS (CbRAS) in complex with 4-coumaroyl-(R)-3-(4-hydroxyphenyl)lactate at 3.35 Å resolution (Table S5 and Fig. 6, A and B). CbRAS consists of two pseudosymmetrical domains, the N-terminal domain (residues 1–171 and 371–391) and C-terminal domain (residues 218–370 and 392–429) connected by a long intervening loop (residues 172–217) that lacks secondary structure (Fig. 6B). Therefore, although lacking desirable robustness, TRV VIGS offers a functional approach to assess gene function in P. campanularia.

Next we designed a VIGS construct to silence PcRAS as well as a construct containing GFP to use as a control. Because depleted RA accumulation lacks an associated visible phenotype as that of PcPDS knockdown, we could not target specific parts of the affected tissues for analysis. We therefore harvested VIGS-treated whole leaves for metabolomics analysis. Leaves inoculated with PcRAS VIGS produced significantly lower quantities of RA than those inoculated with a control, although PcRAS VIGS did not cause complete depletion of RA accumulation (Fig. 5C). It is possible that the existing RA pool prior to VIGS depletes slowly when PcRAS expression is knocked down. Additionally, the knockdown nature of RNAi and the lack of robustness in tissue penetrance of TRV VIGS in P. campanularia may contribute to this.

### Structural features of independently evolved Boraginaceae and Lamiales RAS enzymes

To elucidate the structural elements dictating the common biochemistry of two independently evolved RAS enzymes, we solved the X-ray crystal structure of C. blumei RAS (CbRAS) in complex with 4-coumaroyl-(R)-3-(4-hydroxyphenyl)lactate at 3.35 Å resolution (Table S5 and Fig. 6, A and B). CbRAS consists of two pseudosymmetrical domains, the N-terminal domain (residues 1–171 and 371–391) and C-terminal domain (residues 218–370 and 392–429) connected by a long intervening loop (residues 172–217) that lacks secondary structure (Fig. 6B). Therefore, although lacking desirable robustness, TRV VIGS offers a functional approach to assess gene function in P. campanularia.

Next we designed a VIGS construct to silence PcRAS as well as a construct containing GFP to use as a control. Because depleted RA accumulation lacks an associated visible phenotype as that of PcPDS knockdown, we could not target specific parts of the affected tissues for analysis. We therefore harvested VIGS-treated whole leaves for metabolomics analysis. Leaves inoculated with PcRAS VIGS produced significantly lower quantities of RA than those inoculated with a control, although PcRAS VIGS did not cause complete depletion of RA accumulation (Fig. 5C). It is possible that the existing RA pool prior to VIGS depletes slowly when PcRAS expression is knocked down. Additionally, the knockdown nature of RNAi and the lack of robustness in tissue penetrance of TRV VIGS in P. campanularia may contribute to this.
CbRAS and PcRAS share overall low sequence identity (high degree of tertiary structure similarity (BAHDs and contain many active-site amino acid substitutions compared with other well-characterized orthologous groups of characteristic of BAHD acyltransferases), respectively. Because of unsuccessful attempts to obtain a crystal structure among four presumed Boraginaceae RAS orthologs (P. campanularia, Mertensia paniculata, Ehretia acuminata, and Lennoa madreporeoides), we produced the structural model of PcRAS using Robetta (Fig. 6A) (46).

The structural alignment of the two RAS structures reveals a high degree of tertiary structure similarity (Fig. S5A), characteristic of BAHD acyltransferases (36, 47, 48). Nevertheless, CbRAS and PcRAS share overall low sequence identity (Fig. 6E) compared with other well-characterized orthologous groups of BAHDs and contain many active-site amino acid substitutions (Fig. S5) (30). Interestingly, a number of substitutions are functionally similar (Fig. 6C). For instance, in CbRAS, His-5-41, which appears to play a role in constructing the binding pocket for the 4-coumaryl moiety of the ligand, is substituted in PcRAS by another aromatic amino acid, Tyr-53, which retains a similar side-chain volume and the capability to hydrogen-bond with the 4-OH group of the ligand. Additionally, Ser-39 in CbRAS is substituted by Thr-5-1 in PcRAS, indicating selection for a small, hydrophilic residue at this position, whereas conservation of hydrophobic residues is reflected by the substitution of Val-148 and Ile-373 in CbRAS to Met-163 and Met-403 in PcRAS, respectively. Moreover, the aromatic Tyr-398 in CbRAS is substituted by the similarly aromatic Phe-428 in PcRAS. These trends were supported across a more extensive multiple sequence alignment comparing Boraginaceae and Lamiaceae RAS enzymes and are consistent with the notion that RAS evolved independently in Boraginaceae and Lamiaceae (Fig. S5).

The structures also shared several conserved residues (Fig. 6D). As expected, both enzymes contain residues that are known to be involved in the catalytic cycle of a typical BAHD acyltransferase, such as the catalytic histidine (His-152 in CbRAS and His-167 in PcRAS) and the tryptophan involved in stabilizing the negatively charged, tetrahedral transition state intermediate (Trp-368 in CbRAS and Trp-398 in PcRAS) (36, 48). Additionally, we observed the conservation of an active-site lysine that appears to form a salt bridge with the carboxylate group of the ligand (Lys-396 in CbRAS and Lys-426 in PcRAS) (Fig. S5). Recombinant alanine mutants of PcRAS at these three conserved positions all displayed less than 5% activity for 4-coumaryl-(R)-3-(4-hydroxyphenyl)lactate production compared with WT PcRAS in vitro (Fig. S6). In contrast, mutations to the p-coumaryl binding pocket of PcRAS did not result in a drastic reduction of catalysis (Fig. S6), suggesting that selective binding of the acyl acceptor molecule is more crucial for an efficient RAS protein. Previously, we identified an arginine residue in HCT that serves as a handle to attract and bind the acyl acceptor substrate shikimate into the proper position and orientation within the active site (36). It is likely that this lysine residue acts analogously throughout the RAS catalytic cycle, facilitating binding of (R)-3-(4-hydroxyphenyl)lactate by engaging in a guiding electrostatic interaction with the acyl acceptor substrate. It is of particular interest that, despite the independent origins of the Boraginaceae and Lamiaceae RASs from nonorthologous progenitors, both sets of enzymes converged toward an active-site lysine to mediate the specific acyl-acceptor-binding mechanism (Fig. S5).

Discussion

The RA biosynthetic pathway reported here is the first to be characterized from any non-Lamiaceae species. Notably, although Lamiaceae RASs and Boraginaceae RASs both belong to the same subfamily of the BAHD acyltransferases, indicating...
their related evolutionary trajectories, Lamiaceae RASs were derived from an HCT progenitor, whereas we infer that the Boraginaceae RASs evolved from an ancestral SHT (Fig. 2A) (49). Likewise, although the newly evolved P450s involved in RA biosynthesis in both Boraginaceae and Lamiaceae are members of the CYP98 family, Boraginaceae contains two separate clades of meta-hydroxylases with further specialized catalytic specificity, whereas Lamiaceae only contains a single dual-function meta-hydroxylase (Fig. 2B). The simplest explanation for the phylogenetic pattern of these related CYP98 P450s is parallel evolution of neofunctionalized RA biosynthetic P450s in Boraginaceae and Lamiaceae from the highly conserved ancestral C3’H. Collectively, this evidence supports that the Boraginaceae RA biosynthetic pathway evolved independent from that of the Lamiaceae.

From a biophysical perspective, the fact that Boraginaceae RAS evolved from an ancestral SHT is intriguing because such an evolutionary trajectory calls for a switch in the charge specificity of the acyl acceptor substrate; SHT binds the positively charged spermidine, whereas RAS binds the negatively charged (R)-3-(4-hydroxyphenyl)lactate. This key event is further implied by the structural observation of the electrostatic nature of the acyl-acceptor-proximal region of the RAS active site. Interestingly, despite the disparate evolutionary histories of PcRAS and CbRAS, it appears that these enzymes converged upon the same molecular mechanism to establish a positive electrostatic identity within the acyl-acceptor-binding pocket and, in turn, promote efficient substrate binding. Our structural analysis and mutagenesis studies revealed an active-site lysine (Lys-426 in PcRAS and Lys-396 in CbRAS) that appears to engage in an ionic interaction with the carboxylate functional group of (R)-3-(4-hydroxyphenyl)lactate (Fig. 6C), similar to the mechanism by which an active-site arginine assists shikimate binding in HCT (36).

Our study is also the first to resolve two regioselective P450s catalyzing the two meta-hydroxylation steps of RA biosynthesis in a non-Lamiaceae species. A previous study characterizing CYP98A6 from Lithospermum erythrorhizon (Boraginaceae), which is a presumed ortholog of PcCYP98A112, noted this enzyme’s ability to 3-hydroxylate the 4-coumaroyl moiety of 4-coumaroyl-(R)-3-(4-hydroxyphenyl)lactate or 4-coumaroyl-(R)-3-(3,4-dihydroxyphenyl)lactate, whereas a study of C. blumei revealed that CYP98A14 is a bifunctional P450 capable of hydroxylating both the 3 and 3’ positions of 4-coumaroyl-(R)-3-(4-hydroxyphenyl)lactate (19, 50). Of note, we and others were unable to identify any additional Lamiaceae CYP98 genes putatively involved in RA biosynthesis. It seems plausible that the meta-hydroxylation steps of RA pathway intermediates evolved to be a one-enzyme process in Lamiaceae and a two-enzyme process in Boraginaceae (Fig. 7). Moreover, in Boraginaceae, it appears that 3-hydroxylation of the acyl acceptor moiety is functionally more derived than 3’-hydroxylation of the acyl donor moiety. This is implied by the more distant evolutionary relationship of PcCYP98A111 with the C3’H ancestral clade (Fig. 2C) and supported by the observation that PcCYP98A111 can catalyze the conversion of 4-coumaroyl-(R)-3-(3,4-dihydroxyphenyl)lactate to RA at low efficiency (Fig. S4A). Additionally, coinfiltration of PcRAS and PcCYP98A113 is sufficient to produce some quantity of RA, indicating that the endogenous N. benthamiana C3’H may be compensating for PcCYP98A112 function (Fig. 3B). Notably, the activity assay results indicate that there may not be any requisite order of
hydroxylation, suggesting the existence of a metabolic grid wherein 4-coumaroyl- (R)-3-[(4-hydroxyphenyl)]lactate is shuttled serendipitously between CYP98A112 and CYP98A113 until it is converted to RA (Fig. 4, B and C). Finally, although we observed that either PcCYP98A112 or PcCYP98A113 is sufficient to produce a small quantity of RA from 4-coumaroyl- (R)-3-[(4-hydroxyphenyl)]lactate in vitro, the results of our N. benthamiana infiltration experiments indicate that the majority of RA in vivo is likely produced by a combination of the two cytochrome P450 hydroxylases (Fig. S4B and Fig. 3B).

To adapt to challenging terrestrial environments with a sessile lifestyle, plants have evolved highly pliable specialized metabolic systems that continue to give rise to new metabolic traits. This evolutionary development is rooted in ancestral enzyme promiscuity and largely driven by gene duplication followed by sub- and neofunctionalization (51–57). Our findings reflect upon this remarkable plasticity of specialized metabolism and address a long-standing question regarding the evolutionary origins of RA in the plant kingdom. Our results reveal a scenario wherein 4-coumaroyl-(R)-3-[(4-hydroxyphenyl)]lactate is likely produced by a combination of the two cytochrome P450 hydroxylases (Fig. S4B and Fig. 3B).

Materials and methods

Transcriptome assembly

P. campanularia raw RNA-Seq reads (FASTQ files) were downloaded from the 1KP project (SRA ID ERR2040527) (26), trimmed for sequencing adapters using Trimmomatic (59), and assembled into a de novo transcriptome using Trinity v. 2.6.5 (60), resulting in 43,709 transcript sequences. Within those, 32,155 putative open reading frames were identified using Transdecoder (61). Gene expression statistics were determined using RSEM with bowtie2 as an aligner (62, 63). Transcripts and predicted protein sequences were annotated with transcript-per-million values and closest BLAST hits from the UniProtKB/Swiss-Prot database using in-house scripts. Transcriptome mining was performed on a local BLAST server (64).

Phylogenetic analysis of candidate RA biosynthetic enzymes

Candidates for evolutionary analysis were selected from the assembled P. campanularia transcriptome (26). Sequence alignments were performed in MEGA7 using the MUSCLE algorithm, and all phylogenetic analyses were conducted in MEGA7 (65, 66). Evolutionary history was inferred by using the maximum likelihood method based on the Le_Gascuel 2008 model (67). Evolutionary rate differences among sites were modeled with a discrete Gamma distribution with five categories. Bootstrap values were calculated using 500 replicates.

Molecular biology

Total RNA was extracted from P. campanularia leaf tissue using the RNeasy Plant Mini Kit (Qiagen). First-strand complementary DNA was synthesized from RNA using the SuperScript III First-Strand Synthesis System with the oligo(dT) primer (Thermo Fisher Scientific). ORFs were amplified from complementary DNA using Phusion High-Fidelity DNA Polymerase (New England Biolabs) and gene-specific primers. Plasmids were linearized using restriction enzymes purchased from New England Biolabs as follows: pHIS8–4 with Ncol, pEAQ-HT with AgeI and Xhol, pEarleyGate 100 with HindIII and EcoRI, pJKW 1145 with Xhol and Xbal, pTRV2 with BamHI and EcoRI, and pYEPD60 with BamHI and EcoRI. ORFs were cloned into target vectors using Gibson assembly (68). The accession number for CbRAS is CAK55166.1. PcRAS Y53F, T51A, and K426A mutant plasmids were prepared from the cloned PcRAS-pHIS8–4 plasmid using the Agilent QuickChange II site-directed mutagenesis protocol. PcRAS W398A and H167A mutant plasmids were prepared from the cloned PcRAS-pHIS8–4 plasmid by PCR amplification of two overlapping gene fragments using primers bearing the desired mutation and cloning both fragments into the pHIS8–4 vector using Gibson assembly. All primers used in this study are listed in Table S6.

Transient expression of candidate genes in N. benthamiana and metabolite analysis

pHCQT, PcRAS, PcCYP98A111, PcCYP98A112, PcCYP98A113, and additional genes listed in Table S1 were cloned into pEAQ-HT (29). Vectors were transformed into Agrobacterium tumefaciens strain LBA4404 (69). Bacteria were cultured at 30 °C in YM medium (0.4 g/liter yeast extract, 10 g/liter mannitol, 0.1 g/liter NaCl, 0.2 g/liter MgSO4·7H2O, and 0.5 g/liter K2HPO4·3H2O) to an A600 between 2.0 and 2.5, washed with 0.5× PBS, and then diluted with 0.5× PBS to an A600 of 0.8. For coinfiltration of multiple gene constructs, A. tumefaciens strains were mixed in equal concentrations to a final A600 of 0.8. 1 ml of bacterial suspension was infiltrated into leaves of N. benthamiana plants between 4 and 6 weeks of age. Leaves were harvested 5 days after infiltration.

Transgenic A. thaliana

To generate pJKW 1145, pEarleyGate 100 was digested with EcoRI and Xbal to remove the 3S promoter (p3SS) and ccdB cassette and then Gibson-assembled with the A. thaliana C4H (At4H) promoter amplified from plasmid pCC 0996 (70, 71). PcHCT and PcRAS were cloned into pEarleyGate 100, and PcRAS was additionally cloned into pJKW 1145. Cloned plasmids were transformed into A. tumefaciens strain GV3130. WT A. thaliana were transformed with the flower dip method, and transformants in the T1 generation were selected by spraying with a 1:500 dilution of Finale (Bayer) (72). Samples for metabolite extraction were harvested from T2 plants.

Mechanical transmission of VIGS to P. campanularia

A 268-bp fragment from the 3′ end of PcRAS and full-length GFP were cloned into pTRV2-MCS (73). Together with pTRV1, the cloned vectors were transformed into A. tumefaciens strain GV2260, and cultures were prepared according to a published protocol (74). TRV1 and TRV2 cul-
Evolution of rosmarinic acid biosynthesis

tures were mixed to a final $A_{600}$ of 0.5, and 1 ml of the mixture was infiltrated into the leaves of 4-week-old *N. benthamiana* plants. Leaves were harvested 3 days after infiltration, frozen in liquid nitrogen, and immediately ground in 0.1 M phosphate buffer (pH 7.0) (10 ml/1 g of fresh weight of tissue) and 400-mesh-particle-size silicon carbide (Sigma-Aldrich) abrasive (0.4 g/1 g of fresh weight of tissue). The mixture was gently applied to the surface of *P. campanulata* leaves, stirring frequently to ensure homogenous suspension of the abrasive throughout the procedure. Inoculated leaves were washed immediately under cold tap water for 5 s. Leaves were harvested 2 weeks following inoculation.

Metabolite extraction from plant tissue samples

Tissues from *N. benthamiana, A. thaliana,* and *P. campanulata* were harvested into grinding tubes containing 250 μl of 2.0 mm zirconia/silica disruption beads (Research Products International), frozen in liquid nitrogen, and stored at −80 °C until metabolite extraction. Frozen tissue was homogenized with a TissueLyser II (Qiagen), resuspended in 500 μl until metabolite extraction. Frozen tissue was homogenized with a TissueLyser II (Qiagen), resuspended in 500 μl of 50% methanol/100 mg of tissue, and incubated at 65 °C for 1 h. Extracts were centrifuged at 25,000 × g for 30 min, and the supernatant was collected for analysis.

Recombinant acyltransferase expression and purification

**PcHCT, PcHQT, PcRAS, and PcSHT** were cloned into pHIS8–4, an isopropyl 1-thio-β-d-galactopyranoside (IPTG)–inducible protein expression vector with an N-terminal His tag followed by a tobacco etch virus cleavage site. Cloned vectors were transformed into the Escherichia coli BL21(DE3) strain. Each strain was grown in Terrific Broth medium for 3 h at 37 °C, followed by induction of expression with 0.5 μM of IPTG at 18 °C overnight. Cells were harvested by centrifugation at 5,000 × g for 15 min, resuspended in lysis buffer (50 mM Tris (pH 8.0), 500 mM NaCl, 20 mM imidazole (pH 8.0), 10% glycerol, 1% Tween 20, and 10 mM 2-mercaptoethanol) and lysed at 4 °C with an M-110L microfluidizer (Microfluidics). Lysate was clarified by centrifugation at 25,000 × g for 45 min at 4 °C and immediately affinity-purified with a HisTrap HP histidine-tagged protein purification column (GE Healthcare Life Sciences). His-tagged tobacco etch virus protease was added to the eluate, and the solution was dialyzed overnight at 4 °C in 50 mM Tris (pH 8.0), 50 mM NaCl, and 10 mM 2-mercaptoethanol. Cleaved dialyzed protein was collected by gravity purification through nickel-nitritrocetic acid resin (Qiagen), concentrated to 5 ml using 30-kDa Amicon Ultra-15 centrifugal filter units (EMD Millipore), and purified via gel filtration on a HiLoad 16/600 Superdex 200 pg column (GE Healthcare Life Sciences). The eluate was collected and dialyzed overnight at 4 °C in 12.5 mM Tris (pH 8.0), 50 mM NaCl, and 2 mM DTT. Dialyzed protein was concentrated to 15 mg/ml in 30-kDa Amicon Ultra-15 centrifugal filter units, flash-frozen in liquid nitrogen, and stored at −80 °C until use. Both affinity purification and gel filtration were performed on ÅKTA Pure FPLC (GE Healthcare Life Sciences). PcRAS mutant proteins were prepared using the same protocol.

P450 expression and microsome purification in *Saccharomyces cerevisiae*

**PcCYP98A111, PcCYP98A112,** and **PcCYP98A113** were cloned into pYEPD60 and transformed into the *S. cerevisiae* WAT11 strain (75). Each strain was grown at 30 °C inYPGE medium (5 g/liter glucose, 10 g/liter peptone, 10 g/liter yeast extract, and 3% (v/v) ethanol) to an $A_{600}$ of 2.7, induced with 20 g/liter galactose, and grown at 30 °C to an $A_{600}$ of 3.5. Cells were washed three times by centrifugation at 1,000 × g for 5 min and resuspended in high-salt Tris buffer (50 mM Tris, 4 mM EDTA, 150 mM NaCl, and 20% glycerol (pH 7.4)). Washed, resuspended cells were lysed by vortexing with zirconia/silica disruption beads (Research Products International) and centrifuged at 1,000 × g for 10 min, and then the supernatant was transferred. 0.1 g/ml of PEG 3350 was added slowly to the supernatant, and the mixture was agitated for 15 min at 4 °C before centrifugation at 25,000 × g for 10 min. The pellet was resuspended in 20 ml of assay buffer (50 mM Tris, 4 mM EDTA, and 20% glycerol (pH 7.4)), homogenized with a Dounce homogenizer, flash-frozen in liquid nitrogen, and stored at −80 °C until use.

Enzyme assays

For acyltransferase enzyme kinetics assays, a master mix of 917 nM *A. thaliana* 4CL1 in 50 mM Tris-HCl (pH 7.5), 5 mM MgCl$_2$, 2 mM ATP, 2 μM p-coumaric acid, and 3 μM CoA was equilibrated overnight at room temperature to yield p-coumaryl CoA at an approximate concentration of 2 μM. The reactions consisted of 75 μl of master mix and 5 μl of substrate and initiated with 20 μl of enzyme. 2.09 nmol of PcHCT, 2.07 nmol of PcHQT, and 1.95 nmol of PcRAS were used in each reaction, with the following exceptions: 0.209 nmol of PcHCT was used in the shikimate reactions, and 0.207 nmol of PcHQT was used in the quinate reactions. Concentrations of shikimate, quinate, and (R)-3-(4-hydroxyphenyl)lactate, 4-coumaroylshikimate, 4-coumaroyl-3-(4-hydroxyphenyl)lactate, or caffeoyl-3-(4-hydroxyphenyl)lactate were 10, 50, 100, 250, 500, 1,000, 2,500, and 5,000 μM. (R)-3-(4-hydroxyphenyl)lactate was tested additionally at 25 μM. Reactions were timed for 10 min and quenched by adding 10 μl of 10 μM acetic acid and used immediately for LC-MS analysis. Time points were confirmed to be within the initial rate by running reactions to completion in the presence of excess enzyme over 1 h and ensuring that measured values were less than 5% of the total amount of product formed. PcRAS mutagenesis assays were performed likewise using 2 nmol of enzyme and 2 μM of (R)-3-(4-hydroxyphenyl)lactate and performed in triplicate. Standard curves were obtained by running reactions with 10, 25, 50, 100, and 250 μM substrate to completion and then used to absolutely quantify initial rate measurements. $K_m$ and $V_{max}$ were determined by fitting the data to the Michaelis–Menten equation using non-linear regression in Prism version 7.0a (GraphPad). To measure P450’s activity to produce 4-coumaroyl shikimate, 4-coumaroyl-(R)-3-(4-hydroxyphenyl)lactate, 4-coumaroyl-(R)-3-(3,4-dihydroxyphenyl)lactate, or caffeoyl-(R)-3-(4-hydroxyphenyl)lactate, master mixes (50 mM Tris (pH 7.5), 2 mM 4-coumaroyl-CoA or caffeoyl-CoA, 5 mM appropriate acyl acceptor, and 2 nmol of PcHCT or PcRAS) were set up at room temperature for 2 h. In a total volume of 100 μl, 30 μl of the master mix was suspended.
in 100 mM Tris (pH 7.5) and 5 mM NADPH, and 100 μl of thawed microsomes was added to start the reaction. Reactions were quenched after 15, 30, or 60 min with 100 μl of 10 mM acetic acid and used immediately for LC-MS analysis.

**LC-MS**

LC was performed on a Dionex UltiMate 3000 UHPLC system (Thermo Fisher Scientific) using water with 0.1% formic acid (solvent A) and acetonitrile with 0.1% formic acid (solvent B) and a Kinetex 2.6 μM C18 100 Å column (Phenomenex). To analyze enzyme assays, the gradient was 5% B for 0.5 min, 5%–50% B for 2.5 min, 50%–95% B for 0.1 min, 95% B for 0.9 min, 95%–5% B for 0.1 min, 5% B for 0.9 min, at a flow rate of 0.7 ml/min. Compounds were detected on a TQ5 Quantum Access Max Triple Quadrupole Mass Spectrometer (Thermo Fisher Scientific). Detection of ester compounds was performed in negative ionization mode with selected ion monitoring (SIM) methods set to detect product ions (319.090 m/z for 4-coumaroyl shikimate, 357.077 m/z for caffeoyl shikimate, 337.109 m/z for 4-coumaroyl quinate, 327.095 m/z for 4-coumaroyl-(R)-3-(4-hydroxyphenyl)lactate, 343.082 m/z for 4-coumaroyl-(R)-3-(3,4-dihydroxyphenyl)lactate/caffeoyl-(R)-3-(4-hydroxyphenyl)lactate, and 359.077 m/z for RA). Detection of amide compounds was performed in positive ionization mode with a SIM method set to detect product ions (292.200 m/z for 4-coumaroyl-spermidine, 438.239 m/z for dicoumaroyl-spermidine, and 584.276 m/z for tricoumaroyl-spermidine). Scan widths for SIM methods were set to 1.000 m/z. To analyze metabolite extracts, the gradient was 5% B for 0.5 min, 5%–60% B for 8.5 min, 60%–95% B for 0.2 min, 95% B for 2.6 min, 95%–5% B for 0.2 min, and 5% B for 3 min at a flow rate of 0.7 ml/min. Compounds were detected on a high-resolution Q-Exactive benchtop Orbitrap mass spectrometer (Thermo Fisher Scientific) using a full scan range of 100–1,000 m/z, with separate runs for positive and negative ionization modes. All data were analyzed using the Xcalibur Qual Browser (Thermo Fisher Scientific).

**Crystallization and structure determination of CbRAS**

CbRAS was codon-optimized for *E. coli* expression, ordered as a synthetic gene, and cloned into the pHIS8–3 vector, an IPTG-inducible protein expression vector with an N-terminal His_{6} tag followed by a thrombin cleavage site. Protein was expressed and purified as described previously. CbRAS 4-coumaroyl-(R)-3-(4-hydroxyphenyl)lactate crystals were grown by hanging drop vapor diffusion at 4 °C by mixing 2 μl of 10 mg/ml CbRAS with 1 μl of a reservoir solution containing 0.2 M MgCl_{2}, 0.1 M sodium 3-[[2-hydroxy-1,1-bis(hydroxymethyl)ethyl]-amino]-1-propanesulfonic acid (pH 8.5), 21% PEG 6000, and 15 mM 4-coumaroyl-(R)-3-(4-hydroxyphenyl)lactate. Single crystals were transferred to a cryoprotection solution containing 30% glycerol in reservoir solution, mounted in a cryoloop, and flash-frozen in liquid nitrogen. X-ray diffraction data were collected at beamline 8.2.1 of the Advanced Light Source at the Lawrence Berkeley National Laboratory equipped with ADSC Quantum 315 CCD detectors. Molecular replacement using *A. thaliana* HCT as a search model was done with Phaser under Phenix (76). Further structural refinement utilized Phenix programs (77). Coot was used for manual map inspection and model rebuilding (78). Crystallographic calculations were performed using Phenix. The Fo-Fc omit map for the ligand was calculated by removing the ligand coordinates from the structure and re-refining in Phenix. The atomic coordinates and structure factors of CbRAS have been deposited in the Protein Data Bank (www.rcsb.org) under PDB code 6MK2 (79).

**Structural modeling of PcRAS**

PcRAS was modeled by providing the primary amino acid sequence to the Robetta Full-Chain Protein Structure Prediction Server for Structure Prediction (46). The process was run using default settings, and the top output was used for comparative structural analysis. The root mean square deviation between the CbRAS crystal structure and the PcRAS model is 1.232 Å, as calculated in PyMOL.

**Acknowledgments**—We thank David Nelson for naming the *P. campanularia* P450s. This research used resources of the Advanced Light Source, which is a Department of Energy Office of Science User Facility under contract DE-AC02-05CH11231.

**References**

1. Guo, Y., Li, Y., Xue, L., Severino, R. P., Gao, S., Niu, J., Qin, L.-P., Zhang, D., and Brömme, D. (2014) *Salvia miltiorrhiza*: an ancient Chinese herbal medicine as a source for anti-osteoporotic drugs. *J. Ethnopharmacol.* 155, 1401–1416 CrossRef Medline

2. Khoury, M., Stien, D., Epapvier, V., Ouaini, N., and El Beyrouthy, M. (2016) Report on the medicinal use of eleven Lamiaceae species in Lebanon and rationalization of their antimicrobial potential by examination of the chemical composition and antimicrobial activity of their essential oils. *Evid. Based Complement. Alternat. Med.* 2016, 2547169 Medline

3. Mamadalieva, N. Z., Akramov, D. K., Ovidi, E., Tiezzi, A., Nahar, L., Azimova, S. S., and Sarkar, S. D. (2017) Aromatic medicinal plants of the Lamiaceae family from Uzbekistan: ethnopharmacology, essential oils composition, and biological activities. *Medicines* 10, pii: E8 CrossRef Medline

4. Shetty, K. (1997) Biotechnology to harness the benefits of dietary phenolics: focus on Lamiaceae. *Asia Pac. J. Clin. Nutr.* 6, 162–171 Medline

5. Siciliano, T., Bader, A., Vassallo, A., Braca, A., Morelli, I., Piazza, C., and De Tommasi, N. (2005) Secondary metabolites from *Ballota undulata* (Lamiaceae). *Biochem. Syst. Ecol.* 33, 341–351 CrossRef

6. Mimica-Dukic, N., and Bozin, B. (2008) *Mentha L.* species (Lamiaceae) as promising sources of bioactive secondary metabolites. *Curr. Pharm. Des.* 14, 3141–3150 CrossRef Medline

7. Mint Evolutionary Genomics Consortium. Electronic address: buellg@msu.edu, and Mint Evolutionary Genomics Consortium. (2018) Phylogenetic Mining of the Mints Reveals Multiple Mechanisms Contributing to the Evolution of Chemical Diversity in Lamiaceae. *Mol. Plant.* 11, 1084–1096 CrossRef Medline

8. del Baño, M. J., Lorente, J., Castillo, J., Benavente-García, O., del Río, J. A., Ortúñ, A., Quirin, K.-W., and Gerard, D. (2003) Phenolic diterpenes, flavones, and rosmarinic acid distribution during the development of leaves, flowers, stems, and roots of *Rosmarinus officinalis*: antioxidant activity. *J. Agric. Food Chem.* 51, 4247–4253 CrossRef Medline

4 Please note that the JBC is not responsible for the long-term archiving and maintenance of this site or any other third-party-hosted site.
Evolution of rosmarinic acid biosynthesis

9. Osakabe, N., Yasuda, A., Natsume, M., and Yoshihawa, T. (2004) Rosmarinic acid inhibits epidermal inflammatory responses: anticarcinogenic effect of Perilla frutescens extract in the murine two-stage skin model. Carcinogenesis 25, 549–557 CrossRef Medline

10. Furtado, M. A., de Almeida, L. C., Furtado, R. A., Cunha, W. R., and Tavares, D. C. (2008) Antimitogenicity of rosmarinic acid in Swiss mice evaluated by the micronucleus assay. Mutat. Res. 657, 150–154 CrossRef Medline

11. Frankel, E. N., Huang, S.-W., Aeschbach, R., and Prior, E. (1996) Antioxidant activity of a rosemary extract and its constituents, carnosic acid, carnosol, and rosmarinic acid, in bulk oil and oil-in-water emulsion. J. Agric. Food Chem. 44, 131–135 CrossRef

12. Grigoletto, J., Oliveira, C. V., Grauncke, A. C., Souza T. L., de Souto, N. S., Freitas M. L. de, Furian, A. F., Santos, A. R., and Oliveira, M. S. (2016) Rosmarinic acid is anticonvulsant against seizures induced by pentylenetetrazol and pilocarpine in mice. Epilepsy Behav. 62, 27–34 CrossRef Medline

13. Jin, B.-R., Chung, K.-S., Cheon, S.-Y., Lee, M., Hwang, S., Noh Hwang, S., Rhee, K.-I., and An, H.-J. (2017) Rosmarinic acid suppresses colonic inflammation in dextran sodium sulfate (DSS)-induced mice via dual inhibition of NF-κB and STAT3 activation. Sci. Rep. 7, 46252 CrossRef Medline

14. Jiang, K., Ma, X., Guo, S., Zhang, T., Zhao, G., Wu, H., Wang, X., and Deng, G. (2018) Anti-inflammatory effects of rosmarinic acid in lipopolysaccharide-induced mastitis in mice. Inflammation 41, 437–448 CrossRef Medline

15. Awad, R., Muhammad, A., Durst, T., Trudeau, V. L., and Arnason, J. T. (2009) Bioassay-guided fractionation of lemon balm (Melissa officinalis L.) using an in vitro measure of GABA transaminase activity. Phytother. Res. 23, 1075–1081 CrossRef Medline

16. Lee, H. J., Jeong, Y.-I., Lee, T.-H., Jung, I. D., Lee, J. S., Lee, C.-M., Kim, J.-I., Joo, H., Lee, J.-D., and Park, Y.-M. (2007) Rosmarinic acid inhibits indoleamine 2,3-dioxygenase expression in murine dendritic cells. Biochem. Pharmacol. 73, 1412–1421 CrossRef Medline

17. Bulgakov, V. P., Inyushkina, Y. V., and Fedoreyev, S. A. (2012) Rosmarinic acid and its derivatives: biotechnology and applications. Crit. Rev. Biotechnol. 32, 203–217 CrossRef Medline

18. Berger, A., Meinhard, J., and Petersen, M. (2006) Rosmarinic acid synthase is a new member of the superfamily of BAHD acyltransferases. Planta 224, 1503–1510 CrossRef Medline

19. Eberle, D., Ullmann, P., Wierc-Reichhart, D., and Petersen, M. (2009) cDNA cloning and functional characterisation of CYP98A14 and NADPH:cytochrome P450 reductase from Coleus blumei involved in rosmarinic acid biosynthesis. Plant Mol. Biol. 69, 239–253 CrossRef Medline

20. Petersen, M., Abdullah, Y., Benner, J., Eberle, D., Gehlen, K., Hücherig, S., Janiak, V., Kim, K. H., Sander, M., Weitzel, C., and Wolters, S. (2009) Evolution of rosmarinic acid biosynthesis. Phytochemistry. 70, 1663–1679 CrossRef Medline

21. Manucea, M., Domingues, V. S., Linnen, C. R., Rosenblum, E. B., and Hoekstra, H. E. (2010) Convergence in pigmentation at multiple levels: mutations, genes and function. Philos. Trans. R. Soc. Lond. B Biol. Sci. 365, 2439–2450 CrossRef Medline

22. Pichersky, E., and Lewinsohn, E. (2011) Convergent evolution in plant specialized metabolism. Annu. Rev. Plant Biol. 62, 549–566 CrossRef Medline

23. Tich, F. K., Hage, L. I., Cambraia, R. S., Pereira, P. S., Magro, A. J., Fontes, M. R., Stäbeli, R. G., Giglio, J. R., França, S. C., Soares, A. M., and Sampaio, S. V. (2005) Rosmarinic acid, a new snake venom phospholipase A2 inhibitor from Cordia verbenacea (Boraginaceae): antiserum action potential and molecular interaction. Toxicol. 46, 318–327 CrossRef Medline

24. Rossi, F., Jullian, V., Pawlowiec, R., Kumar-Roiné, S., Haddad, M., Darius, H. T., Gaertner-Mazoni, N., Chinain, M., and Laurent, D. (2012) Protective effect of Heliotropium foertherianum (Boraginaceae) folk remedy and its active component, rosmarinic acid, against a Pacific ciguatoxin. J. Ethnopharmacol. 143, 33–40 CrossRef Medline

25. Mizukami, H., Tabira, Y., and Ellis, B. E. (1993) Methyl jasmonate-induced rosmarinic acid biosynthesis in Lithospermum erythrorhizon cell suspension cultures. Plant Cell Rep. 12, 706–709 CrossRef Medline
Evolution of rosmarinic acid biosynthesis

45. Ye, X., Al-Bahili, S., Klöti, A., Zhang, J., Luca, P., Beyer, P., and Potrykus, I. (2000) Engineering the provitamin A (β-carotene) biosynthetic pathway into (carotenoid-free) rice endosperm. Science 287, 303–305 CrossRef Medline

46. Kim, D. E., Chivian, D., and Baker, D. (2004) Protein structure prediction and analysis using the Robetta server. Nucleic Acids Res. 32, W526 –W531 CrossRef Medline

47. Ma, X., Koepke, J., Panjikar, S., Fritsch, G., and Stöckigt, J. (2005) Crystal structure of vinoine synthase, the first representative of the BAHD superfamily. J. Biol. Chem. 280, 13576–13583 CrossRef Medline

48. Walker, A. M., Hayes, R. P., Youn, B., Vermerris, W., Sattler, S. E., and Kang, C. (2013) Elucidation of the structure and reaction mechanism of Sorghum bicolor hydroxycinnamoyltransferase and its structural relationship to other CoA-dependent transferases and synthases. Plant Physiol. 162, 640–651 CrossRef Medline

49. St-Pierre, B., and De Luca, V. (2000) Origin and diversification of the BAHD superfamily of acetyltransferases involved in secondary metabolism. Recent Adv. Phytochem. 34, 285–315 CrossRef Medline

50. Matsuno, M., Nagatsa, A., Oghara, Y., Ellis, B. E., and Mizukami, H. (2012) The remarkable pliability and promise of Arabidopsis CYP98A6 from S. bicolor to other CoA-dependent transferases and synthases. FEBS Lett. 514, 219–224 CrossRef Medline

51. Aharoni, A., Gaidukov, L., Khersonsky, O., McQ Gould, S., Roodveldt, C., St-Pierre, B., and De Luca, V. (2000) Origin and diversification of the BAHD superfamily of acetyltransferases involved in secondary metabolism. Recent Adv. Phytochem. 34, 285–315 CrossRef Medline

52. Otte, M., Orvis, J., Pochet, N., Strozzi, F., Bowden, J., Couger, M. B., Eccles, D., Li, B., Lieber, M., MacManes, M. D., Ott, M., Orvis, J., Poched, N., Strozzi, F., et al. (2013) De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat. Protoc. 8, 1494–1512 CrossRef Medline

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

54. Langmead, B., and Salzberg, S. L. (2012) Fast gapped-read alignment with Bowtie 2. Nat. Methods. 9, 357–359 CrossRef Medline

55. Priyam, A., Woodcroft, B. J., Rai, V., Munagala, A., Moghul, I., Ter, F., Gibbins, M. A., Moon, H., Leonard, G., Rumpf, W., and Wurm, Y. (2015) Sequenceserver: a modern graphical user interface for custom BLAST databases. bioRxiv 10.1101/033142 CrossRef Medline

56. Kumar, S., Stecher, G., and Tamura, K. (2016) MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Mol. Biol. Evol. 33, 1870–1874 CrossRef Medline

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

58. Le, S. Q., and Gascuel, O. (2008) An improved general amino acid replacement matrix. Mol. Biol. Evol. 25, 1307–1320 CrossRef Medline

59. Gibson, D. G., Young, L., Chuang, R.-Y., Venter, J. C., Hutchinson, C. A., 3rd, Smith, H. O. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat. Methods 6, 343–345 CrossRef Medline

60. Hoekema, A., Hirsch, P. R., Hooykaas, P. J. J., and Schilperoort, R. A. (1983) A binary plant vector system based on separation of vir- and T-region of the Agrobacterium tumefaciens Ti-plasmid. Nature 303, 179 CrossRef Medline

61. Earley, K. W., Haag, J. R., Pontes, O., Opper, K., Juehne, T., Song, K., and Pikaard, C. S. (2006) Gateway-compatible vectors for plant functional genomics and proteomics. Plant J. 45, 616 –629 CrossRef Medline

62. Bonawitz, N. D., Soltau, W. L., Blatchley, M. R., Powers, B. L., Hurlock, A. K., Seals, L. A., Weng, J.-K., Stout, J., and Chapple, C. (2012) REFRFRI, subunits of the transcriptional coregulatory complex mediator, are required for phenylpropanoid homeostasis in Arabidopsis. J. Biol. Chem. 287, 5434–5445 CrossRef Medline

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

64. Pédelaq, J.-D., Cabantous, S., Tran, T., Terverillger, T. C., and Waldo, G. S. (2006) Engineering and characterization of a superfolder green fluorescent protein. Nat. Biotechnol. 24, 79 –88 CrossRef Medline

65. Senthil-Kumar, M., and Mysore, K. S. (2014) Tobacco rattle virus-based virus-induced gene silencing in Nicotiana benthamiana. Nat. Protoc. 9, 1549–1562 CrossRef Medline

66. Pompon, D., Louerat, B., Bronine, A., and Urban, P. (1996) Yeast expression of animal and plant P450s in optimized redox environments. Methods Enzymol. 272, 51–64 CrossRef Medline

67. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software. J. Appl. Crystallogr. 40, 658 –674 CrossRef Medline

68. Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L.-W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., et al. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. D Biol. Crystallogr. 66, 213–221 CrossRef Medline

69. Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. Acta Crystallogr. D Biol. Crystallogr. 60, 2126–2132 CrossRef Medline

70. Berman, H. M., Westbrook, J., Feng, Z., Gililland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000) The Protein Data Bank. Nucleic Acids Res. 28, 235–242 CrossRef Medline