

## Article

# Analysing a Group of Homologous BAHD Enzymes Provides Insights into the Evolutionary Transition of Rosmarinic Acid Synthases from Hydroxycinnamoyl-CoA:Shikimate/Quinate Hydroxycinnamoyl Transferases

Jiali Zhou, Xiaofang Zou, Zixin Deng and Lian Duan \* 

Key Laboratory of Combinatory Biosynthesis and Drug Discovery, Ministry of Education, School of Pharmaceutical Science, Wuhan University, Wuhan 430071, China; 2021203060022@whu.edu.cn (J.Z.); 2022283060061@whu.edu.cn (X.Z.); zxdeng@sjtu.edu.cn (Z.D.)

\* Correspondence: lianduan@whu.edu.cn

**Abstract:** The interplay of various enzymes and compounds gives rise to the intricate secondary metabolic networks observed today. However, the current understanding of their formation and expansion remains limited. BAHD acyltransferases play important roles in the biosynthesis of numerous significant secondary metabolites. In plants, they are widely distributed and exhibit a diverse range of activities. Among them, rosmarinic acid synthase (RAS) and hydroxycinnamoyl-CoA:shikimate/quinate hydroxycinnamoyl transferase (HCT) have gained significant recognition and have been extensively investigated as prominent members of the BAHD acyltransferase family. Here, we conducted a comprehensive study on a unique group of RAS homologous enzymes in *Mentha longifolia* that display both catalytic activities and molecular features similar to HCT and Lamiaceae RAS. Subsequent phylogenetic and comparative genome analyses revealed their derivation from expansion events within the HCT gene family, indicating their potential as collateral branches along the evolutionary trajectory, leading to Lamiaceae RAS while still retaining certain ancestral vestiges. This discovery provides more detailed insights into the evolution from HCT to RAS. Our collective findings indicate that gene duplication is the driving force behind the observed evolutionary pattern in plant-specialized enzymes, which probably originated from ancestral enzyme promiscuity and were subsequently shaped by principles of biological adaptation.

**Keywords:** homologous protein; hydroxycinnamoyl-CoA:shikimate/quinate hydroxycinnamoyl transferase; promiscuity; rosmarinic acid synthase; evolution



**Citation:** Zhou, J.; Zou, X.; Deng, Z.; Duan, L. Analysing a Group of Homologous BAHD Enzymes Provides Insights into the Evolutionary Transition of Rosmarinic Acid Synthases from Hydroxycinnamoyl-CoA:Shikimate/Quinate Hydroxycinnamoyl Transferases. *Plants* **2024**, *13*, 512. <https://doi.org/10.3390/plants13040512>

Academic Editors: Javier Palazon, Hsiaohang Chung and Sarana Sommano

Received: 18 January 2024

Revised: 1 February 2024

Accepted: 5 February 2024

Published: 13 February 2024



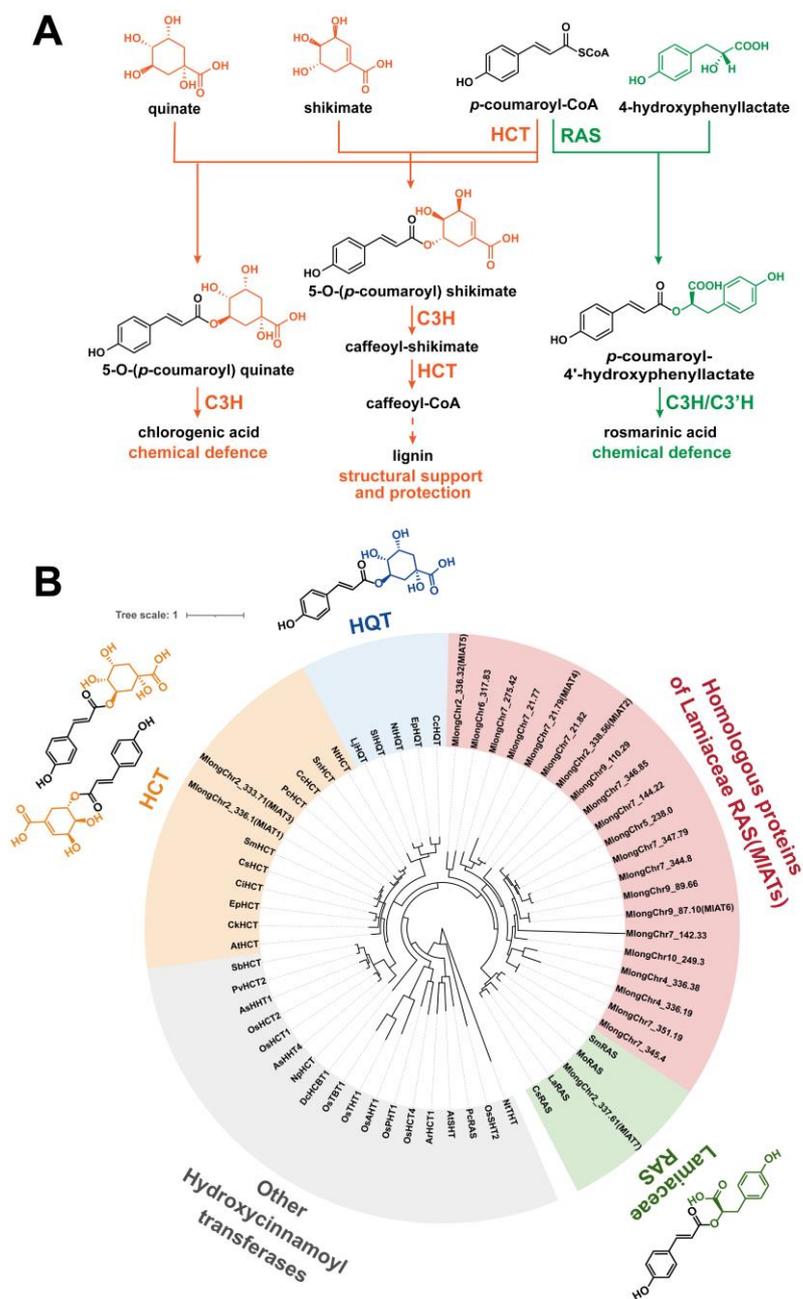
**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Numerous plants belonging to the Lamiaceae family have been extensively employed in traditional medicinal practices for the treatment of diverse ailments, including inflammation, infection, depression, weakness, and indigestion [1–3]. The results of further investigations have revealed that some of the therapeutic properties of these plants primarily stem from their abundant content of phenolic compounds, with a specific emphasis on rosmarinic acid (RA) [4,5]. Pharmacological studies have demonstrated that RA exhibits a wide range of activities encompassing anti-inflammatory effects [6], antioxidative properties [7], antidiabetic potential [8], antitumor activity [9], antiviral effects [10], neuroprotective actions [11], and hepatoprotective benefits [12]. Meanwhile, within the plant kingdom, RA is commonly recognized as an adaptive mechanism to cope with challenging environments and primarily functions in mediating interspecific interactions from a defensive perspective [13].

In order to further investigate and exploit the potential of RA, researchers conducted an analysis of the biosynthetic pathway in plants, leading to the discovery of a crucial enzyme

known as rosmarinic acid synthase (RAS). RAS facilitates the transfer of *p*-coumaroyl acyl groups from *p*-coumaroyl-CoA to the 4-hydroxyphenyllactate compound, resulting in the production of *p*-coumaroyl-4'-hydroxyphenyllactate, which serves as a precursor for RA (Figure 1A). Several RASs have been isolated from various species belonging to Lamiaceae and Boraginaceae families, renowned for their extensive distribution and high accumulation of RA. Interestingly, these enzymes were believed to have evolved from two distinct pathways. Lamiaceae RASs exhibit significant sequence homology with shikimate/quinic acid hydroxycinnamoyl transferases (HCTs), a kind of early-evolved enzyme [13,14], while Boraginaceae RASs were thought to originate from a spermidine hydroxycinnamoyl transferase (SHT) ancestor [15]. Notably, SHT was also considered as an enzyme family that has expanded and evolved from HCT [16,17].



**Figure 1.** (A) In plants, HCT primarily catalyses the transfer of the *p*-coumaroyl acyl group from *p*-coumaroyl-CoA to shikimate or quinate to produce *p*-coumaroyl shikimate or *p*-coumaroyl quinate,

while RAS mainly accepts *p*-coumaroyl-CoA and 4-hydroxyphenyllactate to produce *p*-coumaroyl-4'-hydroxyphenyllactate, a main precursor of rosmarinic acid. The final products of these metabolic pathways assume roles in phytochemical defence or structural support. HCT, shikimate/quinic hydroxycinnamoyl transferase; RAS, rosmarinic acid synthase; C3H, *p*-coumaroyl 3-hydroxylase; C3'H, *p*-coumaroyl 3'-hydroxylase. **(B)** The phylogenetic analysis of MIATs. This phylogenetic tree displays the evolutionary connections between MIATs and diverse hydroxycinnamoyl transferases originating from distinct plant taxa. MIAT1 and MIAT3 are grouped within the clade of the HCT family. Other MIATs are adjacent to the Lamiaceae RASs. HQT, quinic hydroxycinnamoyl transferase. MIATs, a general term for RAS homologous proteins derived from *Mentha longifolia*.

As an early-evolved plant-specialized metabolic enzyme, HCTs exhibited functional conservation over 500 million years of embryophyte development, and plays a crucial role in the phenylpropanoid pathway [18]. This pathway generates a vast array of specialized metabolites and precursors of significant extracellular biopolymers that facilitated early adaptation to terrestrial environments [17,19,20]. Members of the HCT family belong to clade V of BAHD acyltransferases [21] and consistently utilize shikimate and quinate as preferred acyl acceptors to form *p*-coumaroyl shikimate esters and *p*-coumaroyl quinate esters (Figure 1A) [22–25]. Furthermore, investigations into the catalytic properties of HCT have revealed its broad substrate permissiveness towards acceptor molecules [22,26–28]. The promiscuous catalytic property may serve as a foundation for the subsequent emergence of hydroxycinnamoyl transferases with novel functionalities.

Previous investigations have suggested that Lamiaceae RAS might share a pedigree with HCT [14]. Notably, these studies also highlighted significant divergence in the substrate binding centres between the two types of enzymes, which accounts for their distinct substrate selectivity [13,14,27]. However, the preliminary discussion of evolution at the sequence level does not provide substantial evidence to determine the specific differentiation relationship between RAS and HCT. Although current hypotheses lean towards Lamiaceae RAS being derived from an ancient HCT ancestor [15,17], further investigations are required to draw definitive conclusions and explore more intricate details. This study focuses on elucidating the evolutionary hypothesis of Lamiaceae RAS by integrating structural, functional, and evolutionary aspects through the incorporation of additional intermediate sequences to identify deeper insights into these hydroxycinnamoyl transferases.

## 2. Results

### 2.1. A Large Number of RAS Homologous Genes Distributed in the Genome of *Mentha longifolia*

The amino acid sequences of RAS enzymes contributing to the formation of RA in Lamiaceae plants exhibit a significant degree of correlation. For instance, CsRAS (from *Coleus scutellarioides*) shares a pairwise sequence identity of 79.5% with SmRAS (from *Salvia miltiorrhiza*), 79.6% with LaRAS (from *Lavandula angustifolia*), and 82.3% with MoRAS (from *Melissa officinalis*). To gain deeper insights into the origin and development of RAS, these enzymes were employed as baits to identify their homologues in the genome of *Mentha longifolia*, a mint species that exhibits a high level of RA accumulation [29] and has been established as a model species for genetic and genomic research on mint [30]. However, our genome-based exploration did not yield strong related candidate genes. Interestingly, numerous RAS homologous genes are distributed throughout the *Mentha longifolia* genome, primarily concentrated on chromosomes 2, 7, and 9. The enzymes encoded by these RAS homologous genes were designated as MIATs for the sake of simplicity. Preliminary analysis indicates that all MIATs belong to the plant BAHD acyltransferase family.

To elucidate these RAS homologous genes, we constructed a phylogenetic tree encompassing various types of hydroxycinnamoyl transferases (Figure 1B). The phylogenetic tree demonstrated that most of hydroxycinnamoyl transferases were grouped based on physiological/biochemical functions. HCTs and Lamiaceae RASs were classified into distinct clades. The Boraginaceae PcRAS (from *Phacelia campanularia*) was divided into an individual clade, which was consistent with a previous report of independent evolution [15].

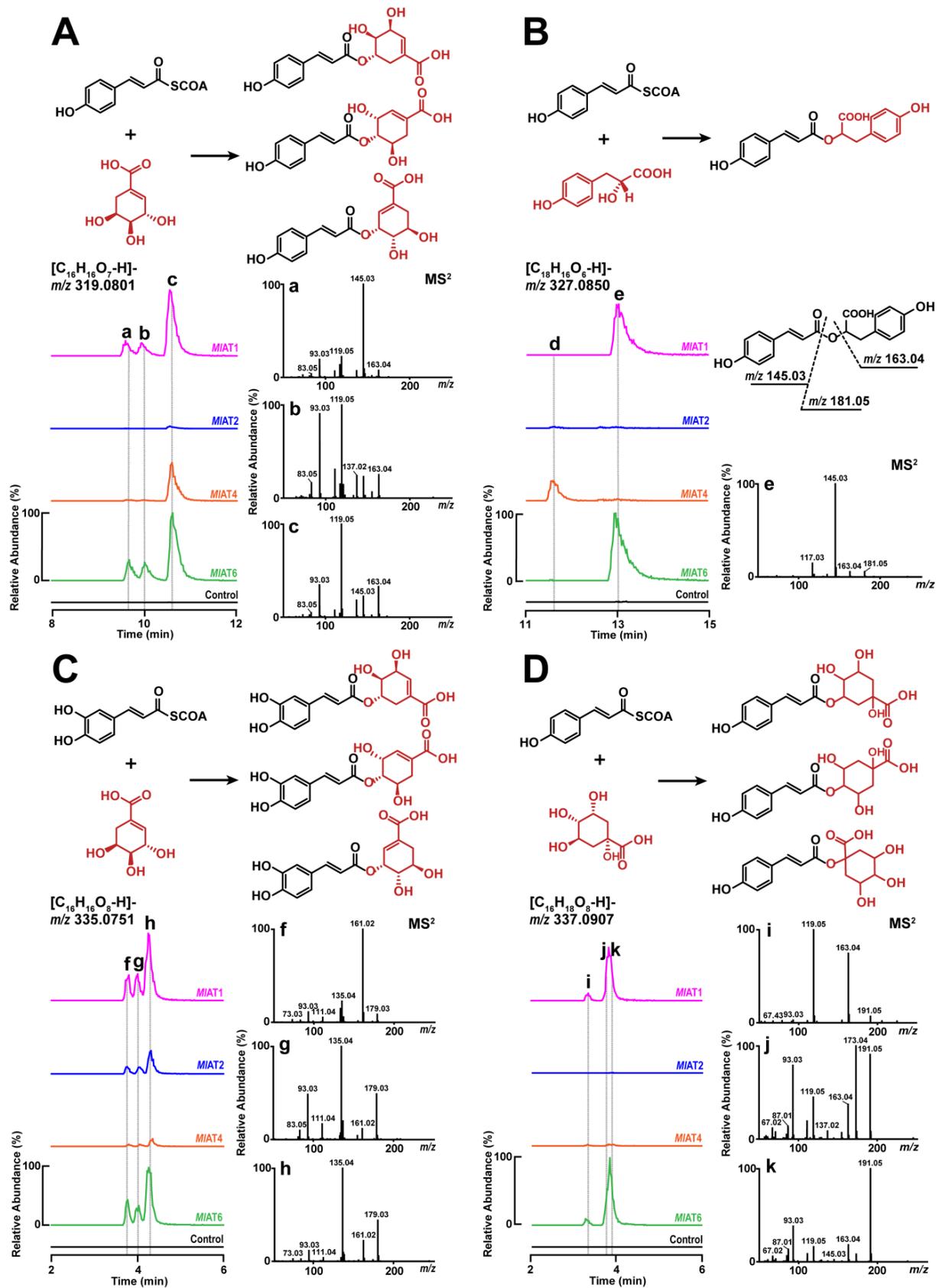
In the *MIATs* population, *MIAT1* and *MIAT3* were located in the clade of HCT, while the majority of other *MIATs* were grouped into a separate clade neighbour to the Lamiaceae RAS family. The conservation level of the sequences within this clade is relatively low compared with other types of enzymes on the tree, with certain members displaying only 50% to 60% identity. *MIAT7* appears to belong to the Lamiaceae RAS family; however, the enzyme encoded by this gene consists of only 160 amino acids, lacking a complete acyltransferase structure.

## 2.2. Revealing the Primary Activity of *MIATs* through Comprehensive Characterization

As homologous enzymes of RAS, *MIATs* may also be involved in the biosynthesis of RA. Therefore, we selected several enzymes (*MIAT1*, *MIAT2*, *MIAT4*, and *MIAT6*) for in vitro functional verification. These selected enzymes exhibit a certain degree of evolutionary divergence and originate from distinct chromosomes. For testing purposes, salvianolic acid A, 4-hydroxyphenyllactate, shikimate, and quinate were chosen as prioritized acyl acceptors, while *p*-coumaroyl-CoA and caffeoyl-CoA were chosen as acyl donors in the enzymatic reactions.

The results revealed that different activities were exhibited by *MIAT1*, *MIAT2*, *MIAT4*, and *MIAT6* when incubated with different substrate pairs (Figure 2). An intriguing observation is that *MIAT1* and *MIAT6* exhibit nearly indistinguishable activity, despite belonging to distinct clades. They efficiently facilitated the transfer of acyl groups from acyl donors to shikimate or quinate, resulting in three *p*-coumaroyl shikimate isomers, three caffeoyl shikimate isomers, and three *p*-coumaroyl quinate isomers in our tests (Figure 2A,C,D). Moreover, they also catalysed the formation of *p*-coumaroyl-4'-hydroxyphenyllactate (a precursor of RA) by accepting the substrate pair consisting of *p*-coumaroyl-CoA and 4-hydroxyphenyllactate, as confirmed through precise molecular ion mass and tandem mass spectrometry (MS<sup>2</sup>) analysis (Figure 2B), in accordance with previous studies [31–33]. In the case of *MIAT2*, only trace products were detected via mass spectrometry spectrum when using *p*-coumaroyl-CoA as the acyl donor. However, it readily accepted caffeoyl-CoA and shikimate to generate three isomeric forms of caffeoyl shikimate (Figure 2C). *MIAT4* exhibited a preference for transferring the acyl group from *p*-coumaroyl-CoA to shikimate to form an isomer type of *p*-coumaroyl shikimate (Figure 2A). The regional selectivity of this phenomenon appears to be more pronounced compared with its counterparts. It also displayed some ability in accepting other substrate pairs, such as promoting the conversion of *p*-coumaroyl-CoA and 4-hydroxyphenyllactate into an unknown isomer of *p*-coumaroyl-4'-hydroxyphenyllactate (Figure 2B), catalysing the generation of either *p*-coumaroyl-3',4'-dihydroxyphenyllactate or its isomer when reacting with *p*-coumaroyl-CoA and salvianolic acid A (Figure S1), and facilitating the production of various forms of caffeoyl-shikimates by utilizing caffeoyl-CoA and shikimate. Regarding *MIAT6*, it exhibited comparable activities to *MIAT1* (Figure 2) by accommodating various substrate pairs resulting in diverse enzyme-catalyzed products.

Overall, in the aforementioned activity tests, *MIAT1* was confirmed to be an HCT enzyme, while phylogenetically closer homologous proteins to RAS, such as *MIAT2*, *MIAT4*, and *MIAT6*, also exhibited significant HCT activity. Additionally, all enzymes demonstrated detectable RAS activity, suggesting potential promiscuous activities.



**Figure 2.** Enzyme activity test of *MIATs*. Acyl acceptors, namely 4-hydroxyphenyllactate, shikimate, and quinate, can be utilized by *MIAT1* (an HCT), *MIAT2*, *MIAT4*, and *MIAT4* to generate their respective

esters with *p*-coumaroyl-CoA or caffeoyl-CoA at varying degrees of catalytic efficiency. (A) The substrate pair consists of *p*-coumaroyl-CoA and shikimate, with “a”, “b”, and “c” representing three *p*-coumaroyl shikimate isomers. (B) The substrate pair consists of *p*-coumaroyl-CoA and 4-hydroxyphenyllactate, where “e” represents *p*-coumaroyl-4'-hydroxyphenyllactate molecule, while “d” may be an isomer with an unknown structure. (C) The substrate pair consists of caffeoyl-CoA and shikimate, with “f”, “g”, and “h” representing three caffeoyl shikimate isomers. (D) The substrate pair consists of *p*-coumaroyl-CoA and quinate, with “i”, “j”, and “k” representing three *p*-coumaroyl quinate isomers. Control groups: substrate pairs and buffer; enzyme groups: substrate pairs, enzymes, and buffer. Structural formulas in black represent acyl donors, whereas red formulas represent acyl acceptors. LC-MS was used to detect products, molecular ion peaks were extracted in ESI-mode, and fragmentation of the ions assisted in the qualitative analysis.

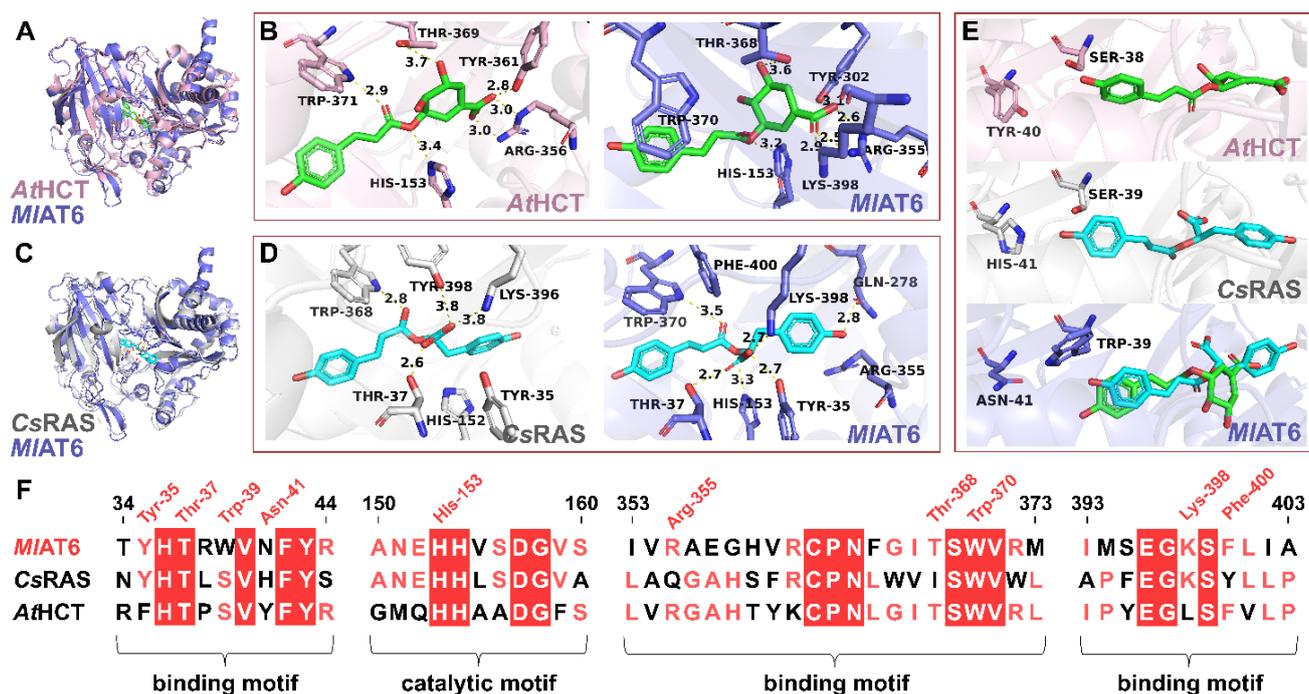
### 2.3. Homology Modelling Revealed That MIAT6 Shares Multiple Active Residues Both with HCT and RAS in the Catalytic Cavity

In the aforementioned experimental tests, we observed a fascinating phenomenon in which MIAT1 and MIAT6, despite sharing only 51% amino acid sequence identity, exhibited identical catalytic functions. They not only demonstrate superior HCT activity, but also possess remarkable RAS catalytic capabilities. As a typical HCT enzyme with clear evolutionary status and function, the active residues in MIAT1 catalytic pockets are basically consistent with other identified HCTs (Figure S2) [28]. However, the underlying catalytic principle of MIAT6 is still unknown. To gain deeper insights into the binding and catalysis mechanisms of MIAT6, we further investigated these processes at the molecular level.

The protein structure models of MIAT6 were generated using the AlphaFold2 online server [34]. The molecular model revealed that MIAT6 possessed an active pocket located at the intersection of two pseudosymmetrical domains, consistent with the characteristic features of BAHD acyltransferases [35–37]. Subsequently, virtual docking experiments were performed to dock a *p*-coumaroyl-4'-hydroxyphenyllactate molecule and a *p*-coumaroyl shikimate molecule into the active cavity of MIAT6. The resulting models from molecular docking were compared with crystal structures of CsRAS (PDB code: 6MK2) [15] and AtHCT (PDB code: 5KJU) [24] for validation purposes (Figure 3A,C). Our findings demonstrated that, in addition to the conserved active residues involved in the catalytic cycle shared by all BAHD acyltransferases, such as catalytic histidine (His-152 in CsRAS, His-153 in AtHCT, and His-153 in MIAT6), as well as the tryptophan residue responsible for stabilizing negatively charged tetrahedral transition-state intermediates (Trp-368 in CsRAS, Trp-371 in AtHCT, and Trp-370 in MIAT6), according to previous studies [38], multiple active sites within acyl receptor binding domain have been identified between MIAT6, CsRAS, and AtHCT.

For example, in CsRAS, there are four polar residues (Tyr-35, Thr-37, Lys-396, and Tyr-398) that potentially serve as binding sites for the carboxyl group of 4-hydroxyphenyllactate (Figure 3D) and are conserved in Lamiaceae RASs. Molecular docking analysis suggested that MIAT6 also has a similar acyl acceptor recognition mechanism. Tyr-35 and Thr-37 in MIAT6 may interact with the 4-hydroxyphenyllactate unit through hydrogen bonds, while Lys-398 may form a salt bridge with the carboxylate group of *p*-coumaroyl-4'-hydroxyphenyllactate (Figure 3D). There was only one non-polar substitution at those four positions (Phe-400 in MIAT6).

Similarly, MIAT6 also possesses multiple active residues similar to AtHCT, such as Arg-355 and Thr-368 in MIAT6 and Arg-356 and Thr-369 in AtHCT (Figure 3B), which are favourable for binding shikimate molecules [28]. In the docking model of MIAT6, Arg-355 appeared to form multiple salt bridges with the carboxyl group of the shikimate unit (Figure 3B). It has been confirmed that arginine at the same position as HCT is a prominent site for recognizing specific acyl acceptors because it can build stable double salt bridges with the carboxyl group of a shikimate or its analogues [28,38]. Additionally, compared with AtHCT, MIAT6 contains certain extra residues likely involved in contact with the shikimate unit, including Tyr-302 and Lys-398.



**Figure 3.** Structural comparison and sequence alignment of *MIAT6* with *AtHCT* and *CsRAS*. (A) Structure comparison of *MIAT6* (the docking model, purple) and *AtHCT* (the crystal structure, pink). (B) The interaction of *p*-coumaroyl-5-*O*-shikimate with *AtHCT* and *MIAT6*. In *AtHCT*, Arg-356 forms double salt bridges with the carboxyl group of shikimate moiety, while Tyr-361 and Thr-369 form hydrogen bonds with the ligand. In the case of *MIAT6*, Tyr-302, Arg-355, and Lys-398 interact with the carboxyl moiety of *p*-coumaroyl-5-*O*-shikimate, while Thr-368 forms a hydrogen bond with the hydroxyl group. (C) Structural comparison of *MIAT6* (the docking model, purple) and *CsRAS* (the crystal structure, grey). (D) The interaction of *p*-coumaroyl-4'-hydroxyphenyllactate with *CsRAS* and *MIAT6*. Tyr-35, Thr-37, Lys-396, and Tyr-398 in *CsRAS* serve as potential interaction sites surrounding the carboxyl group of *p*-coumaroyl-4'-hydroxyphenyllactate, while in *MIAT6*, Tyr-35, Thr-37, and Lys-398 act together on the ligand. In addition, Gln-278 in *MIAT6* is prone to forming a hydrogen bond with phenolic hydroxyl groups. (E) Potential residues in *MIAT6*, *CsRAS*, and *AtHCT* that may be implicated in indirect interactions with acyl donors. (F) Three primary binding motifs and a catalytic motif in *MIAT6*, *CsRAS*, and *AtHCT*. Regions highlighted in red with white text represent sequence identity among the three enzymes, while regions with light red text represent sequence identity between each two of three. Additionally, in the above catalytic pockets, Trp and His are active residues associated with the catalytic cycle, conservatively present in all plant BAHD acyltransferases. The *p*-coumaroyl-5-*O*-shikimate is marked in green, and the *p*-coumaroyl-4'-hydroxyphenyllactate is in cyan.

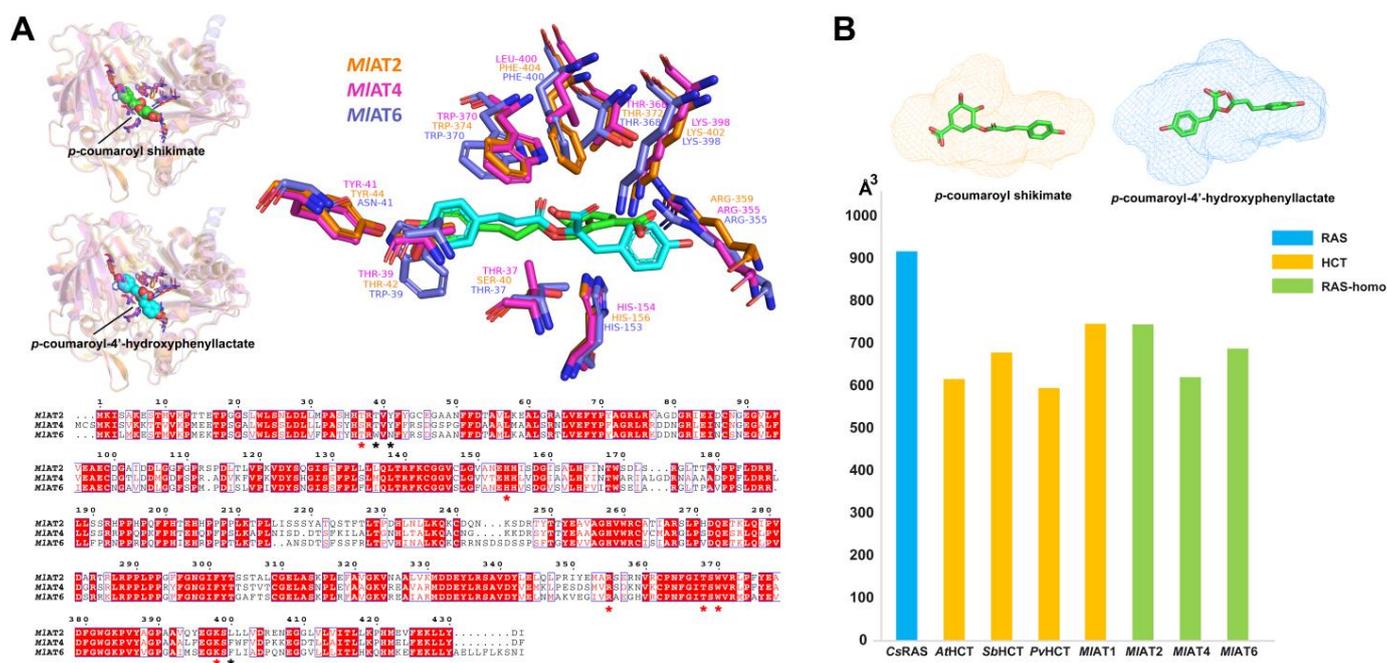
As previously mentioned, *MIAT6* shares multiple active residues both in the acyl receptor binding region of the active pocket with *CsRAS* and *AtHCT*. In contrast, the binding domains of acyl donors exhibit several substitutions of active residues that may have indirect contact with acyl-CoA thioesters [28,38] (Figure 3E). Specifically, two acidic residues, Ser-38 and Tyr-40 in *AtHCT*, are substituted by two basic residues, Trp-39 and Asn-41 in *MIAT6*, while at the same place as *CsRAS*, there is the acidic residue Ser-39 and the basic residue His-41.

Moreover, by identifying these key residues in the active pockets of *MIAT6*, *CsRAS*, and *AtHCT*, we obtained three binding motifs along with one catalytic motif (Figure 3F). We observed that a majority of the members of *MIATs* universally exhibited significant characteristics of both HCT and RAS at these specific motifs, especially in the acyl-acceptor-binding region (Figure S3), which indicates that *MIAT6* probably has a specific acyl acceptor recognition mechanism similar to both RAS and HCT mechanisms, providing an explana-

tion for its concurrent RAS and HCT activity. In subsequent experiments, we introduced mutations to two key active residues (Arg-355 and Lys-398) in the *MIAT6* active pocket. In vitro activity assays demonstrated their indispensability for protein function, as mutation of either R355A or K398L resulted in a complete loss of catalytic activity (Figure S4).

#### 2.4. The Narrow Catalytic Cavities of MIATs May Impose Constraints on Substrate Binding

The structural models of *MIAT2* and *MIAT4* were obtained using the AlphaFold2 program, and their comparison with *MIAT6* was performed (Figure 4A). In terms of the distribution of active residues in catalytic pockets, *MIAT2* and *MIAT4* are similar to *MIAT6*, especially in acyl receptor binding regions. However, in the acyl donor binding sites, two basic residues (Trp-39 and Asn-41 in *MIAT6*) were substituted by two acidic residues (Thr-39 and Tyr-41 in *MIAT2*, Thr-42 and Tyr-44 in *MIAT4*).



**Figure 4.** (A) Structural comparison and multiple sequence alignment of *MIAT2*, *MIAT4*, and *MIAT6*. They are conserved in the acyl receptor binding domain and exhibit similar molecular characteristics overall. Regions highlighted in red with white text represent sequence identity among the three enzymes. The active residues in the catalytic centre are indicated with an asterisk; red indicates good sequence identity; and black indicates that they are not conservative. (B) Calculated volumes of proteins' active cavities. The volume of the active pocket of CsRAS was significantly larger than that of HCTs and *MIATs*. It should be pointed out that the CsRAS and HCTs used in the calculation are crystal structures, whereas *MIATs* correspond to structural models.

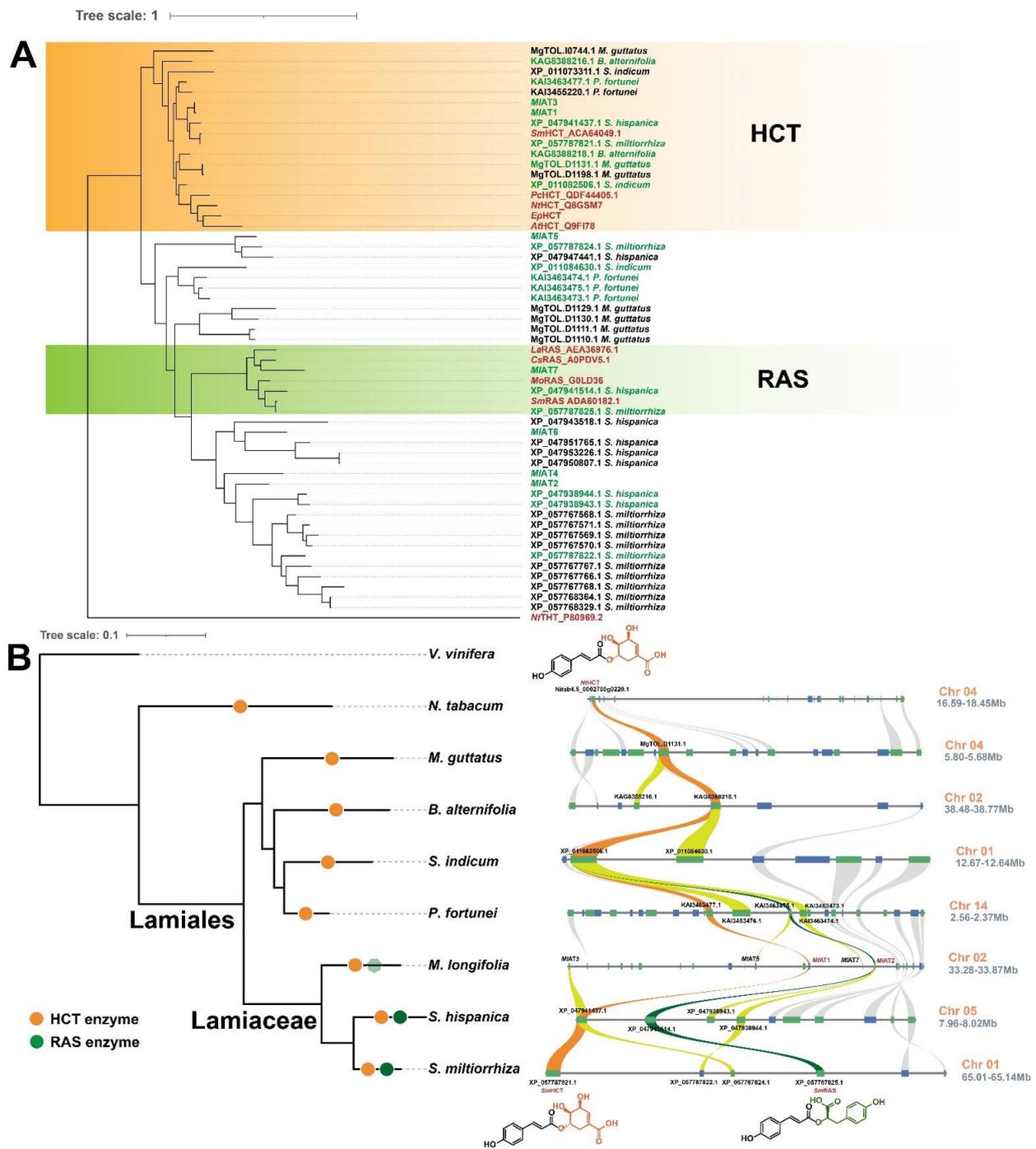
Although phylogenetically and molecularly close to RAS, the ability of these *MIATs* to accept 4-hydroxyphenyllactate or 3,4-dihydroxyphenyllactate was found to be insignificant based on the available data. To investigate the potential constraints on their catalytic activity, we calculated the active pocket volumes for several protein crystals and models. The results revealed an expanded CsRAS active pocket with a cavity volume exceeding 900 Å<sup>3</sup> compared with an average active pocket volume of less than 700 Å<sup>3</sup> for HCTs or *MIATs* (Figure 4B).

### 2.5. Tandem Duplication of the HCT Gene Serves a Prerequisite for the Emergence of RAS

The available sequence and functional data supported a potentially close evolutionary relationship between MIATs and the existing Lamiaceae RASs and HCTs. To gain further insights into these aspects, we conducted an extensive search for additional evidence from plant genomes. Previous studies have demonstrated the conservation of catalytic properties and physiological function of HCT throughout the 500 million years of embryophyte evolution [18]. This suggests that many existing HCT homologous proteins, including RAS and MIATs, may have originated from ancestral HCT duplications. To validate this hypothesis, we investigated plant genomes of Lamiaceae and its relatives to identify all genes homologous to HCT. The resulting encoding proteins were aligned with several well-characterized hydroxycinnamoyl CoA transferases for phylogeny reconstruction (Figure 5A). The phylogenetic analysis revealed a strongly supported clade consisting of one or more members from each species at the root, representing the HCT clade. Other proteins formed distinct clades as sister groups to HCT, with the RAS family being one of them.

Subsequently, a species tree of Lamiaceae and its sister families was constructed, and the syntenic segments harbouring HCT genes on the chromosomes of these species were identified (Figure 5B). The findings demonstrated that the tandem duplication of HCT synthesis genes in certain species led to the presence of multiple homologous genes. The MIAT synthesis genes on chromosome 2 of the *Mentha longifolia* genome are arranged in collinear regions, where MIAT1 and MIAT2 exhibit close evolutionary relationships to HCT and RAS, respectively. Consistent with previous hypotheses, RAS synthesis genes in Lamiaceae were derived from the duplication and differentiation of the ancestral HCT gene, and the genetic correlation of *SmRAS* (Accession: ADA60182.1) can be traced back as distant as *NtHCT* (Accession: Q8GSM7.1) in tobacco.

To be clear, we did not find intraspecific and interspecific collinear associations of RAS homologous proteins on chromosomes 7 and 9 of *Mentha longifolia*. These homologous proteins at other chromosomal locations may have been acquired through dispersive duplication or other mechanisms, but we currently lack sufficient insights into these processes. Moreover, our study did not identify a functionally and evolutionarily intact MIRAS synthesis gene in *Mentha longifolia*. The closest evolutionary relative to known RAS genes is MIAT7; however, it displays an incomplete genic structure, resembling pseudogenes.



**Figure 5.** Expansion and diversification of the HCT enzyme family in plants. **(A)** Phylogenetic analysis of HCT and its homologous enzymes in Lamiaceae and its relatives. *NtTHT* (Hydroxycinnamoyl-CoA: tyramine N-hydroxycinnamoyltransferase) as the outgroup. Enzymes highlighted in red indicate verified enzymes that have undergone functional validation and serve as a point of reference. Enzymes highlighted in green indicate those that are located within the collinear block of chromosomes. **(B)** Microsynteny analysis of HCT regions and RAS regions. Red gene names signify that the gene has undergone functional validation, while black denotes the annotation name. The interconnections among all HCT genes on the local chromosomal region are visually emphasized in orange, while the green lines depict the genetic correlation of the *SmRAS* gene. The links to other homologous genes of HCT are highlighted in a yellow-green. Enzymes in the species tree are depicted with solid circles of distinct colours, where orange represents HCT enzymes and green represents RAS enzymes. “Chr” is an abbreviation for chromosome.

### 3. Discussion

Plants employ various specialized enzymes to synthesize a wide range of structurally and functionally diverse metabolites, enabling them to adapt to changing environmental stresses. Investigating the origin and evolution of these specialized metabolic enzymes has long been a crucial aspect in plant evolutionary research. Up to now, extensive efforts have been dedicated to comprehend the generation and development of plant-specialized metabolism [39–41], which not only contributes to enhancing our understanding of plant evolution, but also aids in elucidating the existing constraints on metabolic function and devising improvement strategies [42,43]. In recent years, there has been increased focus among biologists on the evolution of functional diversity in plant-specialized enzyme families that contribute to metabolic diversity [44–46]. However, understanding these processes remains challenging due to the parallel elimination of most evolutionary intermediates by distinct lineages during extensive historical development [47]. The reality is that our exploration of evolutionary footprints is limited to only a small fraction of the available historical phylogenetic samples. Fortunately, significant advancements in plant genome sequencing over the past decade have provided a substantial volume of sequence data, enabling a comprehensive understanding of enzyme evolutionary pathways by connecting related yet discrete biological features. Herein, we employed an interdisciplinary approach integrating chemical, structural, functional, and sequence data within a phylogenetic framework to gain some insights into the evolutionary patterns exhibited by plant-specialized metabolic enzymes.

In general, homologous proteins from closely related species serve as important reference materials for studying the evolutionary patterns of metabolic enzymes [48]. By analysing the genome sequences of *M. longifolia*, we successfully identified one HCT (*MIAT1*) and several RAS homologous enzymes. In vitro functional verification and molecular modelling suggest that they possess catalytic activities and molecular features similar to both HCT and RAS (Figures 2–4). Interestingly, despite their closer phylogenetic affinity to Lamiaceae RASs (Figure 1B), they appear to demonstrate a slightly stronger overall HCT function (Figure 2). Structural analysis of these enzymes' active sites further revealed an acyl-acceptor-binding region resembling both HCT and RAS (Figure 4). Overall, this group of enzymes exhibits a convergence of molecular mechanisms between HCT and RAS, suggesting their unique evolutionary position. Investigation into the genomes of Lamiaceae and closely related species has significantly expanded the populations of homologous enzymes for both RAS and HCT (Figure 5A). Comparative genomic analysis revealed that RAS homologous genes on chromosome 2 of *Mentha longifolia* emerged through tandem duplication events of HCT genes, with *MIAT2* exhibiting a close genetic relationship with the *SmRAS* (Figure 5B). Despite the convergence in sequence and structure, *MIAT2* exhibits limited RAS activity (Figure 2). The genuine RAS synthesis gene was absent in *Mentha longifolia*.

The production and accumulation of some beneficial compounds can confer a competitive advantage in specific ecological niches [49]. Thus, a universally observed phenomenon is that evolution readily selects those promiscuous activities with biological adaptive advantages from existing proteins as the starting point for new functions [50–53]. In previous reports, RA has been confirmed to play an important role in plant defence, including photoprotection and insect resistance [54,55]. We inferred that the emergence of RAS and its homologous enzymes in Lamiaceae can likely be attributed to the substrate permissiveness of progenitor HCTs towards 4-hydroxyphenyllactate and its analogues. The resulting RA or its analogues might have been favoured by plants for ecological benefits, leading to the subsequent evolution of favourable enzyme activities. It should be emphasized that a trade-off often exists between the enhancement of novel functions and the retention of original functionalities, as these processes typically involve distinct molecular mechanisms. In this case, gene duplication is considered an indispensable prerequisite that creates physical conditions for adaptive protein evolution [56]. Gene tandem duplications of HCT in Lamiaceae have not only expanded this enzyme group, but also relieved restrictive pressure

on nucleotide sequences, allowing the accumulation of mutations. In general, the forces of natural selection tend to maintain mutations that confer substantial adaptation advantages, resulting in a collective change in a specific direction at the sequence level. As a result, it is not surprising that these duplicated HCT genes exhibit sequence characteristics that progressively approach the RAS function throughout evolution. Previous studies have shown that the functional specialization of members of the BAHD acyltransferase family may result in sequence variants that are concentrated in specific motifs, rather than globally distributed variants throughout the enzyme structure [57]. Our study also observed this phenomenon and emphasized the connections and distinctions between these homologous enzymes with HCT and RAS in binding and catalytic motifs. As a clade that preserves evolutionary traits, they provide a link for understanding functional differentiation between HCT and RAS.

However, it is worth noting that the majority of these ultimately undergo loss rather than adaption, since evolutionary outcomes are often accidental [58,59]; this aligns with the weaker function observed in some enzymes. Despite not being in an optimal evolutionary form, enzyme genes can persist within a population if they are capable of producing specific compounds that offer adaptive advantages to plants [60]. An intriguing concept is that these non-specific homologous enzymes may possess enhanced evolutionary plasticity as a result of lacking strict functional constraints, particularly regarding functional promiscuity and its potential for enhancing adaptive evolution [50,61,62]. Notably, our study was unable to ascertain the ancestral origin of all RAS homologous proteins discovered in peppermint. There are various types of gene duplication events caused by genetic aberrations, including replication slippage, retrotransposition, ectopic recombination, aneuploidy, and polyploidy [56]. Despite their associations with HCT and RAS based on sequence, structure, and function, we lack qualitative evidence regarding the source of other proteins scattered throughout the genome, including *MIAT4* and *MIAT6*. They could have been acquired through decentralized repetition mechanisms, or might have functioned as enzymes assuming alternative roles.

Furthermore, in addition to plants belonging to the Lamiaceae and Boraginaceae families, sporadic traces of the RA chemotype have also been found in various taxa of flowering plants, as well as some early plant groups such as hornworts and ferns [13,14,63]. While existing evidence supports the independent evolution of RA biosynthesis in different plant lineages, our study suggests that enzyme promiscuity may also contribute to the accumulation of RA, which is consistent with the trace amounts observed in certain plant taxa. Overall, the findings of our study offer novel evidence regarding the origin and evolutionary trajectory of RAS in Lamiaceae, achieved through a comprehensive identification process of RAS homologous enzymes. Additionally, our research provides some insights into the intricate mechanisms underlying the metabolism and accumulation of RA in *Mentha longifolia*.

#### 4. Conclusions

In this study, we conducted a comprehensive investigation and characterization of several RAS homologues to unravel their divergence from the HCT family, leading to the emergence and enhancement of the RA metabolic pathway. Our analysis suggests that gene tandem duplications of ancestral HCTs during Lamiaceae plant evolution resulted in extensive homologous enzymes. Subsequent selection pressures likely drove them to explore new functions for enhanced biological adaptability. In this process, a small number of enzymes have completed functional divergence and acquired specific RAS functions. Most of the other enzymes eventually aggregate in the adjacent branches of the Lamiaceae RASs in phylogenetic trees, exhibiting molecular characteristics similar to RAS but with less activity. As an offshoot to the RAS developmental trajectory, the analysis of these enzymes has provided a glimpse into their evolution.

Our work enhances understanding of the developmental trajectory from HCT to RAS in Lamiaceae by elucidating functional and structural features shared among different

homologous enzymes. This provides some insights into the evolution of specialized metabolic enzymes in plants. The observed evolutionary pattern reflects a natural order resulting from random mutations guided by function to meet adaptive needs. The study also provides a novel case in favour of the fundamental theory of enzymatic evolution. The expansion of enzyme families weakens functional constraints and creates favourable conditions for new enzyme emergence, shaping a diverse and adaptable evolutionary landscape. The acquisition of new functions does not necessarily disrupt old mechanisms completely; there may be an adaptive transitional stage before switching occurs. In future research, increasingly sophisticated experimental techniques and enriched omics data will provide opportunities for the better understanding, more efficient utilization, and more precise development of these enzymes.

## 5. Materials and Methods

### 5.1. Genome Mining

Genome data of *Mentha longifolia*. (Horse mint) are stored in the Mint Genomics Resource (<http://langelabtools.wsu.edu/mgr/> (accessed on 7 October 2022)). The database provides an online BLAST [64] server. The TBLASTN search was used to perform a homologous search for RAS, identifying database sequences encoding proteins similar to the query. Proteins encoded by these candidate genes were preliminarily annotated using the NCBI database (<https://www.ncbi.nlm.nih.gov/> (accessed on 8 October 2022)) and the UniProtKB/Swiss-Prot database (<https://www.uniprot.org> (accessed on 8 October 2022)).

The BLAST search was used to investigate orthologous genes of HCT in the genomes of Lamiaceae and related species, including *Buddleja alternifolia* (GCA\_019426215.1), *Mimulus guttatus* (MguttatusTOL\_551\_v5.0), *Paulownia fortune* (GCA\_019321725.1), *Salvia hispanica* (GCF\_023119035.1), *Sesamum indicum* (GCF\_000512975.1), and *Salvia miltiorrhiza* (GCF\_028751815.1). These plant genomes were downloaded from Phytozome (<https://phytozome-next.jgi.doe.gov/> (accessed on 12 January 2024)) and NCBI. In addition, the HCT gene (*NtHCT*) in the genome of *Nicotiana tabacum* was searched and located, aiming to provide a reference for the evolutionary origins of HCT homologous genes in other species. The tobacco genome (Nitab-v4.5) was downloaded from Sol Genomics Network, <https://solgenomics.net/> (accessed on 13 January 2024).

### 5.2. Phylogenetic Analysis of Candidate Homologous Genes of RAS

The amino acid sequences were aligned using the MUSCLE algorithm [65]. Subsequently, phylogenetic analyses were conducted in MrBayes (v3.2.6) [66] with the following parameters: average standard deviation = 0.1; minimum generation = 500,000; maximum generation = 0 ( $\infty$ ); and detection every 5000 generations. Other BAHD-AT sequences for analysis can be downloaded from the NCBI database. Enzymes used to build the phylogenetic trees can be queried in Table S1.

### 5.3. Molecular Biology

The CDSs of *MIAT1*, *MIAT2*, *MIAT4*, and *MIAT6* were synthesized through the gene service company (Tsingke, Nanjing, China) and inserted into the pET-28a(+) plasmid with a 6×His tag at the N-terminus (completed by the company). All sequences were optimized to make them more suitable for expression in *E. coli* (the process of codon optimization does not alter the amino acid sequence). Additionally, *MIAT1* and *MIAT4* were further inserted into the pET-28a(+)-SUMO plasmid with a 6×His tag and a SUMO tag at the N-terminus to promote protein solubility. The CDSs were prepared through PCR amplification from the pET-28a(+)-*MIAT1* and pET-28a(+)-*MIAT4* plasmids using I-5™ 2X High-Fidelity Master Mix (MCLAB, Beijing, China) and then cloned into the target vector using the Uniclone One-Step Seamless Cloning Kit (Genesand, Beijing, China). Mutant plasmids of R355A, K398L, and R355A+K398L were prepared from the pET-28a(+)-*MIAT6* plasmid by designing primer pairs at specific mutant sites and using the multi-gene fragment recombination strategy.

Plasmids were linearized using FastDigest restriction enzymes purchased from Thermo Scientific as follows: pET-28a(+) was linearized with Nhe I and BamH I; pET-28a(+)-SUMO was linearized with Nhe I and BamH I.

The primers utilized in the experiments are listed in Table S2. The coding sequences (CDSs) of *MIATs* are presented in Information S1, while the optimized sequences are available in Information S2. Additionally, the mutation sequences of *MIAT6* can be found in Information S3.

#### 5.4. Recombinant Protein Expression and Purification

The plasmids of pET-28a(+)-SUMO\_*MIAT1*, pET-28a(+)-*MIAT2*, and pET-28a(+)-SUMO\_*MIAT4* were transformed into pGro7/BL21(DE3) chaperone competent cells; pET-28a(+)-*MIAT6* were transformed into Rosetta (DE3) competent cells. The cells were grown in terrific broth medium at 37 °C for 2–3 h, followed by the induction of expression with 0.1 mM IPTG at 18 °C for 16–18 h. For the strains of pGro7/BL21(DE3), an additional 1 mg/mL *L*-arabinose was added to the medium to induce the expression of chaperone proteins. After centrifugation at 9000× *g* for 30 min, the cells were resuspended in lysis buffer (50 mM Tris-HCl, 300 mM NaCl, 25 mM imidazole, pH 8.0) and lysed using an Ultrasonic Homogenizer (Scientz, Ningbo, China) at 4 °C. The lysate was clarified by centrifugation at 15,000× *g* for 1 h at 4 °C and filtered to remove insoluble cell fragments. The clarified lysate was immediately affinity-purified using a HisTrap FF crude 5 mL column (Amersham Biosciences, Slough, UK) prepacked with precharged Ni Sepharose™ 6 Fast Flow for the preparative purification of histidine-tagged recombinant proteins through immobilized metal affinity chromatography. A large amount of binding buffer (50 mM Tris-HCl, 1 M NaCl, 25 mM imidazole, pH 8.0) was used to wash off the heteroproteins. Elution buffer (50 mM Tris-HCl, 200 mM NaCl, and 300 mM imidazole, pH 8.0) was used to wash off the histidine-tagged recombinant proteins. Recombinant proteins containing a SUMO tag needed to have the label removed by using SUMO protease digestion. The processed proteins did not have an affinity for nickel ions; they could flow out with the binding buffer. The collected solution was dialyzed overnight at 4 °C in dialysis buffer (50 mM Tris-HCl, 50 mM NaCl, and 10% glycerol) to remove imidazole. The dialyzed protein was concentrated to 5–20 mg/mL in an Amicon Ultra-15 mL 10 kDa Centrifugal Filter Unit (Millipore, Burlington, MA, USA), flash-frozen in liquid nitrogen, and stored at –80 °C until use. The preparation of *MIAT6* mutant proteins followed the same protocol as that for the *MIAT6* proteins.

#### 5.5. Enzyme Assays

Recombinant proteins were taken out from –80 °C and thawed in the ice. The thawed solution needed to have glycerol removed in order to reduce side reactions. This was performed by using an Amicon Ultra-0.5 mL 10 kDa Centrifugal Filter Unit (Millipore) to replace the storage buffer with the reaction buffer, which consisted of 50 mM Tris-HCl, pH 8.0, 1 mM DTT, and 0.5 mM ascorbic acid. Each protein was diluted to 5 mg/mL.

The enzymatic activity of recombinant proteins (*MIAT1*, *MIAT2*, *MIAT4*, and *MIAT6*) was determined by incubating protein fractions in a total volume of 125 µL, consisting of 110 µL reaction buffer, 2.5 µL 100 mM acyl donor (*p*-coumaroyl-CoA or caffeoyl-CoA), 2.5 µL 10 mM acyl acceptor (4-hydroxyphenyllactate, 3,4-dihydroxyphenyllactate, shikimate, or quinate), along with 10 µL recombinant proteins. All reactions were carried out at 30 °C for 3 h, and then 20 µL 6 N HCl was added to terminate them. The reaction products were extracted three times with 500 µL ethyl acetate each time, and then evaporated to remove the solvent. The extracts were redissolved in 150 µL 50% methanol and filtered for LC-MS analysis. We detected significant hydrolysis of hydroxycinnamoyl-CoA under the catalysis of enzymes in tests; therefore, we did not explore the enzyme kinetics of the acyl transfer reaction.

The enzymatic activity of *MIAT6* and its mutant proteins were determined in total volumes of 50 µL, consisting of 40 µL reaction buffer, 1 µL 100 mM *p*-coumaroyl-CoA, 1 µL

10 mM shikimate, and 8  $\mu$ L recombinant proteins. The reaction was carried out at 28 °C for 2 h, and then 50  $\mu$ L MeOH was added to terminate. The solution was centrifuged at  $17,000\times g$  for 15 min, and then the supernatant was used for LC-MS analysis. Each reaction was repeated three times.

(Supplementary instruction: These reaction conditions, including the reaction time, temperature, and termination conditions, were all established based on the hydroxycinnamoyl transferase reactions reported in the literature. Additionally, we conducted a time gradient analysis of the reaction and observed consistent production of multiple isomers, regardless of whether the reaction duration was 30 min or 3 h.)

#### 5.6. UHPLC-Q-Executive Focus Mass Spectrometry

LC was performed on a Dionex UltiMate 3000 UHPLC system (Thermo Fisher Scientific, Waltham, MA, USA) using Milli-Q water (solvent A) and methanol (solvent B). Samples, including the groups of *p*-coumaroyl-CoA and shikimate, *p*-coumaroyl-CoA and 4-hydroxyphenyllactate, and caffeoyl-CoA and shikimate, were analysed using an Eclipse XDB C18 80 Å column (length 250 mm, diameter 3 mm, particle size 5  $\mu$ m, Agilent, Santa Clara, CA, USA). The elution gradient was as follows: 10–30%B for 5 min, 30–70%B for 5 min, 70–90%B for 6 min, 90%B for 2 min, 90–10%B for 2 min, and 10%B for 5 min at a flow rate of 1 mL/min. For samples of the group of *p*-coumaroyl-CoA and quinate, the elution gradient was modified as follows: 10–50%B for 5 min, 50–70%B for 7 min, 70–90%B for 2 min, 90–10%B for 1 min, and 10%B for 5 min at a flow rate of 1 mL/min. The same conditions were applied to the samples of hydrolysis reactions.

Additionally, samples of *MIAT6* and its mutant proteins were analysed using a ZORBAX 300SB-C8 column (length 150 mm, diameter 4.6 mm, particle size 5  $\mu$ m, Agilent), equipped with a protective column (Phenomenex SecurityGuard KJ0-4282, Guangzhou, China). The elution gradient was set as follows: 10%–30%B for 3 min, 30%–90%B for 12 min, 90%B for 5 min, 90%–10%B for 1 min, and 10%B for 4 min at a flow rate of 0.3 mL/min.

All compounds were detected using a high-resolution Q-Executive Focus Orbitrap Mass Spectrometer (Thermo Fisher Scientific), operating in negative ionization mode (ESI-) with a full scan range of 120–1200 *m/z*. All data were analysed using the Xcalibur 4.0 Qual Browser (Thermo Fisher Scientific).

#### 5.7. Prediction of Protein Structural Models

The homologous models of acyltransferases were generated by submitting amino acid sequences to the AlphaFold2 Protein Structure Prediction Online Server (<https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/AlphaFold2.ipynb> (accessed on 25 July 2022)). The process was executed using default settings, and the rank1 output was utilized for comparative structural analysis. Additionally, we compared the prediction result with those obtained from other protein structure prediction servers, namely Robetta (<https://rosetta.bakerlab.org/> (accessed on 6 July 2022)) and Phyre2 (<http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index> (accessed on 6 July 2022)). The root-mean-squared deviation (RMSD) values for the same protein predicted by different programs were less than 2 Å, as calculated in PyMOL (2.5.2).

#### 5.8. Molecular Docking

The crystal structures of CsRAS (PDB ID: 6MK2) and *AtHCT* (PDB ID: 5KJU) were downloaded from Uniprot (<https://www.uniprot.org/> (accessed on 12 July 2022)). All ligands were drawn using ChemBio3D Ultra 14.0 and saved in PDB format. Due to an incorrect connection between the Arg-355 and the Tyr-302 in the rank1 output of *MIAT6*, the rank2 output was selected for docking. Docking was performed using AutoDock4, and the results were analysed using AutoDockTools (v.1.5.6). We employed a semi-flexible docking method where the protein was treated as rigid, except for Arg-355 in *MIAT6*, which was set as a flexible amino acid residue. The grid covered the active cavity with an appropriate cube size. Docking conformations were generated using a Lamarckian genetic algorithm

(LGA), and the results were ordered based on the interaction energy between the substrate and target protein, from lowest to highest. Low-energy conformations were our focus. In addition, references in the literature provided positions and conformations for some active residues of acyltransferases. All results were visualized on PyMOL, including aligned structures, measured distances, calculated RMSD, labelled residues, and figure preparation. Finally, Adobe Illustrator 2023 was used to combine and enhance these images.

### 5.9. Multiple Sequence Alignment

Multiple sequence alignments were performed and visualized in Jalview (v.2.11.2.6) using the MUSCLE algorithm with default parameters. The displayed results were enhanced using Adobe Illustrator 2023. The brighter the colour, the more conserved the site. In addition, we used ESPript 3.0 to help with visualization.

### 5.10. Calculation of Volumes of Protein Activity Pockets

The active pocket volumes of protein crystals and structural models were analysed using parKVFinder v1.0.2, with the cavity detection modes set to “ligand adjustment”, the “probe out” adjusted to 10.0, the “removal distance” set to 0.0, and all other parameters defaulted.

### 5.11. Species Tree

Orthologs were identified using OrthoFinder v2.5.4 software among nine plant species: *Vitis vinifera*, *Nicotiana tabacum* [67], *Buddleja alternifolia* [68], *Sesamum indicum* [69], *Paulownia fortunei* [70], *Mimulus guttatus*, *Mentha longifolia* [30], *Salvia miltiorrhiza* [71], and *Salvia hispanica* [72]. The protein sequences of single-copy orthologous genes were used to construct a phylogenetic tree.

### 5.12. Chromosomal Collinearity Analysis

The nine plant species used to construct the tree were analysed: *Buddleja alternifolia* (GCA\_019426215.1), *Mimulus guttatus* (MguttatusTOL\_551\_v5.0), *Paulownia fortunei* (GCA\_019321725.1), *Salvia hispanica* (GCF\_023119035.1), *Sesamum indicum* (GCF\_000512975.1), *Salvia miltiorrhiza* (GCF\_028751815.1), *Nicotiana tabacum* (Nitab-v4.5), and *V. vinifera* (*Vvinifera\_457\_v2.1*). Protein sequences and GFF/GFF3 attachments from fully sequenced genomes were downloaded from the Phytozome, NCBI, and Sol Genomics Network. Genomic collinearity was detected using MCScan (Python version) ([https://github.com/tanghaibao/jcvi/wiki/MCscan-\(Python-version\)](https://github.com/tanghaibao/jcvi/wiki/MCscan-(Python-version))) (accessed on 15 January 2024)) with default parameters, resulting in the construction of a library dataset that included all possible syntenic gene pairs among the nine plant genomes. Chromosomal collinearity fragments containing HCT and its homologous genes were selected for mapping and visualization.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/plants13040512/s1>, Figure S1: *MIAT4* has RAS catalytic potential; Figure S2: The homology model of *MIAT1*; Figure S3: The multiple sequence alignment of HCTs, RAS-homo proteins in *Mentha longifolia*, and Lamiaceae RASs; Figure S4: Validation of two key active residues; Table S1: Enzymes used for the phylogenetic tree; Table S2: PCR primers used for protein expression; Information S1: The CDS sequences of *MIAT1*-*MIAT7*; Information S2: The optimized sequences of *MIAT1*, *MIAT2*, *MIAT4* and *MIAT6*; Information S3: The nucleotide sequences of *MIAT6* mutants.

**Author Contributions:** J.Z. was the main executor of the experiments and data collection; she also wrote the manuscript. X.Z. contributed to the experiments. Z.D. provided suggestions and oversaw the research. L.D. provided funding and revised the paper. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by National Key R&D Program of China (2021YFA0910500) and National Key R&D Program of China (2018YFA0903200).

**Data Availability Statement:** The Supplementary Materials and PDB files are available within the article. Other raw data are available on request from the corresponding author.

**Acknowledgments:** We are grateful to Yi Yu for providing excellent suggestions for revising the article.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Vieira, A. A comparison of traditional anti-inflammation and anti-infection medicinal plants with current evidence from biomedical research: Results from a regional study. *Pharmacogn. Res.* **2010**, *2*, 195–293. [[CrossRef](#)]
2. Yeung, K.S.; Hernandez, M.; Mao, J.J.; Haviland, I.; Gubili, J. Herbal medicine for depression and anxiety: A systematic review with assessment of potential psycho-oncologic relevance. *Phytother. Res.* **2018**, *32*, 865–891. [[CrossRef](#)]
3. Ahmed, H.M. Ethnomedicinal, Phytochemical and Pharmacological Investigations of *Perilla frutescens* (L.) Britt. *Molecules* **2018**, *24*, 102. [[CrossRef](#)]
4. Hitl, M.; Kladar, N.; Gavarić, N.; Božin, B. Rosmarinic Acid-Human Pharmacokinetics and Health Benefits. *Planta Med.* **2022**, *87*, 273–282. [[CrossRef](#)]
5. Mascoloti, S.R.; Caleja, C.; Pinela, J.; Finimundy, T.C.; Calhelha, R.C.; Kostić, M.; Sokovic, M.; Prieto, M.A.; Pereira, E.; Amaral, J.S.; et al. Comparative study on the phenolic composition and in vitro bioactivity of medicinal and aromatic plants from the Lamiaceae family. *Food Res. Int.* **2022**, *161*, 111875. [[CrossRef](#)] [[PubMed](#)]
6. Jang, A.H.; Kim, T.H.; Kim, G.D.; Kim, J.E.; Kim, H.J.; Kim, S.S.; Jin, Y.-H.; Park, Y.S.; Park, C.-S. Rosmarinic acid attenuates 2,4-dinitrofluorobenzene-induced atopic dermatitis in NC/Nga mice. *Int. Immunopharmacol.* **2011**, *11*, 1271–1277. [[CrossRef](#)] [[PubMed](#)]
7. Lin, C.; Xiao, J.; Xi, Y.; Zhang, X.; Zhong, Q.; Zheng, H.; Cao, Y.; Chen, Y. Rosmarinic acid improved antioxidant properties and healthspan via the IIS and MAPK pathways in *Caenorhabditis elegans*. *Biofactors* **2019**, *45*, 774–787. [[CrossRef](#)] [[PubMed](#)]
8. Runtuwene, J.; Cheng, K.C.; Asakawa, A.; Amitani, H.; Amitani, M.; Morinaga, A.; Takimoto, Y.; Kairupan, B.H.R.; Inui, A. Rosmarinic acid ameliorates hyperglycemia and insulin sensitivity in diabetic rats, potentially by modulating the expression of PEPCK and GLUT4. *Drug Des. Devel. Ther.* **2016**, *10*, 2193–2202. [[PubMed](#)]
9. Venkatachalam, K.; Gunasekaran, S.; Jesudoss, V.A.; Namasivayam, N. The effect of rosmarinic acid on 1,2-dimethylhydrazine induced colon carcinogenesis. *Exp. Toxicol. Pathol.* **2013**, *65*, 409–418. [[CrossRef](#)] [[PubMed](#)]
10. Hsieh, C.F.; Jheng, J.R.; Lin, G.H.; Chen, Y.L.; Ho, J.Y.; Liu, C.J.; Hsu, K.-Y.; Chen, Y.-S.; Chan, Y.F.; Yu, H.-M.; et al. Rosmarinic acid exhibits broad anti-enterovirus A71 activity by inhibiting the interaction between the five-fold axis of capsid VP1 and cognate sulfated receptors. *Emerg. Microbes Infect.* **2020**, *9*, 1194–1205. [[CrossRef](#)] [[PubMed](#)]
11. Khamse, S.; Sadr, S.S.; Roghani, M.; Hasanzadeh, G.; Mohammadian, M. Rosmarinic acid exerts a neuroprotective effect in the kainate rat model of temporal lobe epilepsy: Underlying mechanisms. *Pharm. Biol.* **2015**, *53*, 1818–1825. [[CrossRef](#)]
12. Lin, S.Y.; Wang, Y.Y.; Chen, W.Y.; Liao, S.L.; Chou, S.T.; Yang, C.P.; Chen, C.-J. Hepatoprotective activities of rosmarinic acid against extrahepatic cholestasis in rats. *Food Chem. Toxicol.* **2017**, *108*, 214–223. [[CrossRef](#)]
13. Petersen, M. Rosmarinic acid: New aspects. *Phytochem. Rev.* **2013**, *12*, 207–227. [[CrossRef](#)]
14. Petersen, M.; Abdullah, Y.; Benner, J.; Eberle, D.; Gehlen, K.; Hücherig, S.; Janiak, V.; Kim, K.H.; Sander, M.; Weitzel, C.; et al. Evolution of rosmarinic acid biosynthesis. *Phytochemistry* **2009**, *70*, 1663–1679. [[CrossRef](#)] [[PubMed](#)]
15. Levsh, O.; Pluskal, T.; Carballo, V.; Mitchell, A.J.; Weng, J.K. Independent evolution of rosmarinic acid biosynthesis in two sister families under the Lamiids clade of flowering plants. *J. Biol. Chem.* **2019**, *294*, 15193–15205. [[CrossRef](#)]
16. Elejalde-Palmett, C.; de Bernonville, T.D.; Glevarec, G.; Pichon, O.; Papon, N.; Courdavault, V.; St-Piere, B.; Giglioli, N.; Lanoue, A.; Besseau, S. Characterization of a spermidine hydroxycinnamoyltransferase in *Malus domestica* highlights the evolutionary conservation of trihydroxycinnamoyl spermidines in pollen coat of core Eudicotyledons. *J. Exp. Bot.* **2015**, *66*, 7271–7285. [[CrossRef](#)]
17. Weng, J.K. The evolutionary paths towards complexity: A metabolic perspective. *New Phytol.* **2014**, *201*, 1141–1149. [[CrossRef](#)] [[PubMed](#)]
18. Kriegshäuser, L.; Knosp, S.; Grienberger, E.; Tatsumi, K.; Gütle, D.D.; Sørensen, I.; Herrgott, L.; Zumsteg, J.; Rose, J.K.C.; Reski, R.; et al. Function of the HYDROXYCINNAMOYL-CoA:SHIKIMATE HYDROXYCINNAMOYL TRANSFERASE is evolutionarily conserved in embryophytes. *Plant Cell* **2023**, *33*, 1472–1491. [[CrossRef](#)]
19. Vogt, T. Phenylpropanoid biosynthesis. *Mol. Plant* **2010**, *3*, 2–20. [[CrossRef](#)]
20. Weng, J.K.; Chapple, C. The origin and evolution of lignin biosynthesis. *New Phytol.* **2010**, *187*, 273–285. [[CrossRef](#)] [[PubMed](#)]
21. D’Auria, J.C. Acyltransferases in plants: A good time to be BAHD. *Curr. Opin. Plant Biol.* **2006**, *9*, 331–340. [[CrossRef](#)]
22. Hoffmann, L.; Maury, S.; Martz, F.; Geoffroy, P.; Legrand, M. Purification, cloning, and properties of an acyltransferase controlling shikimate and quinate ester intermediates in phenylpropanoid metabolism. *J. Biol. Chem.* **2003**, *278*, 95–103. [[CrossRef](#)]
23. Lallemand, L.A.; Zubieta, C.; Lee, S.G.; Wang, Y.; Acajjaoui, S.; Timmins, J.; McSweeney, S.; Jez, J.M.; McCarthy, J.G.; McCarthy, A.A. A structural basis for the biosynthesis of the major chlorogenic acids found in coffee. *Plant Physiol.* **2012**, *160*, 249–260. [[CrossRef](#)]

24. Levsh, O.; Chiang, Y.C.; Tung, C.F.; Noel, J.P.; Wang, Y.; Weng, J.K. Dynamic Conformational States Dictate Selectivity toward the Native Substrate in a Substrate-Permissive Acyltransferase. *Biochemistry* **2016**, *55*, 6314–6326. [[CrossRef](#)] [[PubMed](#)]
25. Chiang, Y.C.; Levsh, O.; Lam, C.K.; Weng, J.K.; Wang, Y. Structural and dynamic basis of substrate permissiveness in hydroxycinnamoyltransferase (HCT). *PLoS Comput. Biol.* **2018**, *14*, e1006511. [[CrossRef](#)] [[PubMed](#)]
26. Hoffmann, L.; Besseau, S.; Geoffroy, P.; Ritzenthaler, C.; Meyer, D.; Lapierre, C.; Pollet, B.; Legrand, M. Silencing of hydroxycinnamoyl-coenzyme A shikimate/quinic acid hydroxycinnamoyltransferase affects phenylpropanoid biosynthesis. *Plant Cell* **2004**, *16*, 1446–1465. [[CrossRef](#)] [[PubMed](#)]
27. Sander, M.; Petersen, M. Distinct substrate specificities and unusual substrate flexibilities of two hydroxycinnamoyltransferases, rosmarinic acid synthase and hydroxycinnamoyl-CoA:shikimate hydroxycinnamoyl-transferase, from *Coleus blumei* Benth. *Planta* **2011**, *233*, 1157–1171. [[CrossRef](#)] [[PubMed](#)]
28. Eudes, A.; Pereira, J.H.; Yogiswara, S.; Wang, G.; Teixeira Benites, V.; Baidoo, E.E.; Lee, T.S.; Adams, P.; Kesling, J.D.; Loque, D. Exploiting the Substrate Promiscuity of Hydroxycinnamoyl-CoA: Shikimate Hydroxycinnamoyl Transferase to Reduce Lignin. *Plant Cell Physiol.* **2016**, *57*, 568–579. [[CrossRef](#)] [[PubMed](#)]
29. Shekarchi, M.; Hajimehdipour, H.; Saeidnia, S.; Gohari, A.R.; Hamedani, M.P. Comparative study of rosmarinic acid content in some plants of Labiatae family. *Pharmacogn. Mag.* **2012**, *8*, 37–41. [[PubMed](#)]
30. Vining, K.J.; Pandelova, I.; Lange, I.; Parrish, A.N.; Lefors, A.; Kronmiller, B.; Liachko, I.; Kronenberg, Z.; Srividya, N.; Lange, B.M. Chromosome-level genome assembly of *Mentha longifolia* L. reveals gene organization underlying disease resistance and essential oil traits. *G3* **2022**, *12*, jkac112. [[CrossRef](#)]
31. Wang, X.; Qian, Y.; Li, X.; Jia, X.; Yan, Z.; Han, M.; Qiao, M.; Ma, X.; Chu, Y.; Zhou, S.; et al. Rapid determination of rosmarinic acid and its two bioactive metabolites in the plasma of rats by LC-MS/MS and application to a pharmacokinetics study. *Biomed. Chromatogr.* **2021**, *35*, e4984. [[CrossRef](#)]
32. Di, P.; Zhang, L.; Chen, J.; Tan, H.; Xiao, Y.; Dong, X.; Zhou, X.; Chen, W. <sup>13</sup>C tracer reveals phenolic acids biosynthesis in hairy root cultures of *Salvia miltiorrhiza*. *ACS Chem. Biol.* **2013**, *8*, 1537–1548. [[CrossRef](#)] [[PubMed](#)]
33. Landmann, C.; Hücherig, S.; Fink, B.; Hoffmann, T.; Dittlein, D.; Coiner, H.A.; Schwab, W. Substrate promiscuity of a rosmarinic acid synthase from lavender (*Lavandula angustifolia* L.). *Planta* **2011**, *234*, 305–320. [[CrossRef](#)] [[PubMed](#)]
34. Tunyasuvunakool, K.; Adler, J.; Wu, Z.; Green, T.; Zielinski, M.; Židek, A.; Bridgland, A.; Cowie, A.; Meyer, C.; Laydon, A.; et al. Highly accurate protein structure prediction for the human proteome. *Nature* **2021**, *596*, 590–596. [[CrossRef](#)]
35. Ma, X.; Koepke, J.; Panjikar, S.; Fritzsche, G.; Stöckigt, J. Crystal structure of vinorine synthase, the first representative of the BAHD superfamily. *J. Biol. Chem.* **2005**, *280*, 13576–13583. [[CrossRef](#)]
36. Unno, H.; Ichimaida, F.; Suzuki, H.; Takahashi, S.; Tanaka, Y.; Saito, A.; Nishino, T.; Kusunoki, M.; Nakayama, T. Structural and mutational studies of anthocyanin malonyltransferases establish the features of BAHD enzyme catalysis. *J. Biol. Chem.* **2007**, *282*, 15812–15822. [[CrossRef](#)]
37. Yamane, M.; Takenoya, M.; Yajima, S.; Sue, M. Crystal structure of barley agmatine coumaroyltransferase, an N-acyltransferase from the BAHD superfamily. *Acta Crystallogr. F Struct. Biol. Commun.* **2020**, *76*, 590–596. [[CrossRef](#)]
38. Walker, A.M.; Hayes, R.P.; Youn, B.; Vermerris, W.; Sattler, S.E.; Kang, C. Elucidation of the structure and reaction mechanism of sorghum hydroxycinnamoyltransferase and its structural relationship to other coenzyme a-dependent transferases and synthases. *Plant Physiol.* **2013**, *162*, 640–651. [[CrossRef](#)] [[PubMed](#)]
39. Jacobowitz, J.R.; Weng, J.K. Exploring Uncharted Territories of Plant Specialized Metabolism in the Postgenomic Era. *Annu. Rev. Plant Biol.* **2020**, *71*, 631–658. [[CrossRef](#)]
40. Moghe, G.D.; Leong, B.J.; Hurney, S.M.; Daniel Jones, A.; Last, R.L. Evolutionary routes to biochemical innovation revealed by integrative analysis of a plant-defense related specialized metabolic pathway. *eLife* **2017**, *6*, e28468. [[CrossRef](#)]
41. Kliebenstein, D.J.; Osbourn, A. Making new molecules—Evolution of pathways for novel metabolites in plants. *Curr. Opin. Plant Biol.* **2012**, *15*, 415–423. [[CrossRef](#)]
42. Maeda, H.A.; Fernie, A.R. Evolutionary History of Plant Metabolism. *Annu. Rev. Plant Biol.* **2021**, *72*, 185–216. [[CrossRef](#)] [[PubMed](#)]
43. Milo, R.; Last, R.L. Achieving diversity in the face of constraints: Lessons from metabolism. *Science* **2012**, *336*, 1663–1667. [[CrossRef](#)] [[PubMed](#)]
44. Glasner, M.E.; Truong, D.P.; Morse, B.C. How enzyme promiscuity and horizontal gene transfer contribute to metabolic innovation. *FEBS J.* **2010**, *287*, 1323–1342. [[CrossRef](#)] [[PubMed](#)]
45. Kusano, H.; Li, H.; Minami, H.; Kato, Y.; Tabata, H.; Yazaki, K. Evolutionary Developments in Plant Specialized Metabolism, Exemplified by Two Transferase Families. *Front. Plant Sci.* **2019**, *10*, 794. [[CrossRef](#)]
46. Weng, J.K.; Noel, J.P. The remarkable pliability and promiscuity of specialized metabolism. *Cold Spring Harb. Symp. Quant. Biol.* **2012**, *77*, 309–320. [[CrossRef](#)]
47. Smock, R.G.; Gierasch, L.M. Finding the fittest fold: Using the evolutionary record to design new proteins. *Cell* **2005**, *122*, 832–834. [[CrossRef](#)]
48. Wellner, A.; Raitses Gurevich, M.; Tawfik, D.S. Mechanisms of protein sequence divergence and incompatibility. *PLoS Genet.* **2013**, *9*, e1003665. [[CrossRef](#)]
49. Isah, T. Stress and defense responses in plant secondary metabolites production. *Biol. Res.* **2019**, *52*, 39. [[CrossRef](#)]

50. Depristo, M.A. The subtle benefits of being promiscuous: Adaptive evolution potentiated by enzyme promiscuity. *HFSP J.* **2007**, *1*, 94–98. [[CrossRef](#)]
51. Leong, B.J.; Last, R.L. Promiscuity, impersonation and accommodation: Evolution of plant specialized metabolism. *Curr. Opin. Struct. Biol.* **2017**, *47*, 105–112. [[CrossRef](#)]
52. Pandya, C.; Farelli, J.D.; Dunaway-Mariano, D.; Allen, K.N. Enzyme promiscuity: Engine of evolutionary innovation. *J. Biol. Chem.* **2014**, *289*, 30229–30236. [[CrossRef](#)]
53. Copley, S.D. An evolutionary biochemist’s perspective on promiscuity. *Trends Biochem. Sci.* **2015**, *40*, 72–78. [[CrossRef](#)]
54. Casal-Porras, I.; Muñoz, K.; Ortega, M.J.; Brun, F.G.; Zubía, E. Rosmarinic Acid and Flavonoids of the Seagrass *Zostera noltei*: New Aspects on Their Quantification and Their Correlation with Sunlight Exposure. *Plants* **2023**, *12*, 4078. [[CrossRef](#)] [[PubMed](#)]
55. Khan, S.; Taning, C.N.T.; Bonneure, E.; Mangelinckx, S.; Smagghe, G.; Ahmad, R.; Fatima, N.; Asif, M.; Shah, M.M. Bioactivity-guided isolation of rosmarinic acid as the principle bioactive compound from the butanol extract of *Isodon rugosus* against the pea aphid, *Acyrtosiphon pisum*. *PLoS ONE* **2019**, *14*, e0215048.
56. Weng, J.K.; Lynch, J.H.; Matos, J.O.; Dudareva, N. Adaptive mechanisms of plant specialized metabolism connecting chemistry to function. *Nat. Chem. Biol.* **2021**, *17*, 1037–1045. [[CrossRef](#)] [[PubMed](#)]
57. Moghe, G.; Kruse, L.H.; Petersen, M.; Scossa, F.; Fernie, A.R.; Gaquerel, E.; D’Auria, J.C. BAHD Company: The Ever-Expanding Roles of the BAHD Acyltransferase Gene Family in Plants. *Annu. Rev. Plant Biol.* **2023**, *74*, 165–194. [[CrossRef](#)] [[PubMed](#)]
58. Jayaraman, V.; Toledo-Patiño, S.; Noda-García, L.; Laurino, P. Mechanisms of protein evolution. *Protein Sci.* **2022**, *31*, e4362. [[CrossRef](#)]
59. Lynch, M.; O’Hely, M.; Walsh, B.; Force, A. The probability of preservation of a newly arisen gene duplicate. *Genetics* **2001**, *159*, 1789–1804. [[CrossRef](#)] [[PubMed](#)]
60. Weng, J.K.; Philippe, R.N.; Noel, J.P. The rise of chemodiversity in plants. *Science* **2012**, *336*, 1667–1670. [[CrossRef](#)]
61. Amitai, G.; Gupta, R.D.; Tawfik, D.S. Latent evolutionary potentials under the neutral mutational drift of an enzyme. *HFSP J.* **2007**, *1*, 67–78. [[CrossRef](#)]
62. Wroe, R.; Chan, H.S.; Bornberg-Bauer, E. A structural model of latent evolutionary potentials underlying neutral networks in proteins. *HFSP J.* **2007**, *1*, 79–87. [[CrossRef](#)]
63. Guan, H.; Luo, W.; Bao, B.; Cao, Y.; Cheng, F.; Yu, S.; Fan, Q.; Zhang, L.; Wu, Q.; Shan, M. A Comprehensive Review of Rosmarinic Acid: From Phytochemistry to Pharmacology and Its New Insight. *Molecules* **2022**, *27*, 3292. [[CrossRef](#)]
64. Ye, J.; McGinnis, S.; Madden, T.L. BLAST: Improvements for better sequence analysis. *Nucleic Acids Res.* **2006**, *34*, W6–W9. [[CrossRef](#)]
65. Edgar, R.C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **2004**, *32*, 1792–1797. [[CrossRef](#)]
66. Huelsenbeck, J.P.; Ronquist, F. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **2001**, *17*, 754–755. [[CrossRef](#)]
67. Edwards, K.D.; Fernandez-Pozo, N.; Drake-Stowe, K.; Humphry, M.; Evans, A.D.; Bombarely, A.; Allen, F.; Hurst, R.; White, B.; Kernodle, S.P.; et al. A reference genome for *Nicotiana tabacum* enables map-based cloning of homeologous loci implicated in nitrogen utilization efficiency. *BMC Genom.* **2017**, *18*, 448–461. [[CrossRef](#)] [[PubMed](#)]
68. Ma, Y.P.; Wariss, H.M.; Liao, R.L.; Zhang, R.G.; Yun, Q.Z.; Olmstead, R.G.; Chau, J.; Milne, R.I.; de Peer, Y.; Sun, W. Genome-wide analysis of butterfly bush (*Buddleja alternifolia*) in three uplands provides insights into biogeography, demography and speciation. *New Phytol.* **2021**, *232*, 1463–1476. [[CrossRef](#)] [[PubMed](#)]
69. Wang, X.; Wang, S.; Lin, Q.; Lu, J.; Lv, S.; Zhang, Y.; Wang, X.; Fan, W.; Liu, W.; Zhang, L.; et al. The Wild Allotetraploid Sesame Genome Provides Novel Insights into Evolution and Lignan Biosynthesis. *J. Adv. Res.* **2022**, *50*, 13–24. [[CrossRef](#)] [[PubMed](#)]
70. Cao, Y.; Sun, G.; Zhai, X.; Xu, P.; Ma, L.; Deng, M.; Zhao, Z.; Yang, H.; Dong, Y.; Shang, Z.; et al. Genomic insights into the fast growth of paulownias and the formation of Paulownia witches’ broom. *Mol. Plant* **2021**, *14*, 1668–1682. [[CrossRef](#)] [[PubMed](#)]
71. Pan, X.; Chang, Y.; Li, C.; Qiu, X.; Cui, X.; Meng, F.; Zhang, S.; Li, X.; Lu, S. Chromosome-level genome assembly of *Salvia miltiorrhiza* with orange roots uncovers the role of Sm2OGD3 in catalyzing 15,16-dehydrogenation of tanshinones. *Hortic. Res.* **2023**, *10*, uhad069. [[CrossRef](#)] [[PubMed](#)]
72. Wang, L.; Lee, M.; Sun, F.; Song, Z.; Yang, Z.; Yue, G.H. A chromosome-level genome assembly of chia provides insights into high omega-3 content and coat color variation of its seeds. *Plant Commun.* **2022**, *3*, 100326. [[CrossRef](#)] [[PubMed](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.