

# A Collaborative Classroom Investigation of the Evolution of SABATH Methyltransferase Substrate Preference Shifts over 120 My of Flowering Plant History

Nicole M. Dubs, Breck R. Davis, Victor de Brito, Kate C. Colebrook, Ian J. Tiefel, Madison B. Nakayama, Ruiqi Huang, Audrey E. Ledvina, Samantha J. Hack, Brent Inkelaar, Talline R. Martins, Sarah M. Aartila, Kelli S. Albritton, Sarah Almuhan, Ryan J. Arnoldi, Clara K. Austin, Amber C. Battle, Gregory R. Begeman, Caitlin M. Bickings, Jonathon T. Bradfield, Eric C. Branch, Eric P. Conti, Breana Cooley, Nicole M. Dotson, Cheyone J. Evans, Amber S. Fries, Ivan G. Gilbert, Weston D. Hillier, Pornkamol Huang, Kaitlin W. Hyde, Filip Jevtovic, Mark C. Johnson, Julie L. Keeler, Albert Lam, Kyle M. Leach, Jeremy D. Livsey, Jonathan T. Lo, Kevin R. Loney, Nich W. Martin, Amber S. Mazahem, Aurora N. Mokris, Destiny M. Nichols, Ruchi Ojha, Nnanna N. Okorafor, Joshua R. Paris, Thais Fuscaldi Reboucas, Pedro Beretta Sant'Anna, Mathew R. Seitz, Nathan R. Seymour, Lila K. Slaski, Stephen O. Stemaly, Benjamin R. Ulrich, Emile N. Van Meter, Meghan L. Young, and Todd J. Barkman\*

Department of Biological Sciences, Western Michigan University, Kalamazoo, MI, USA

\*Corresponding author: E-mail: todd.barkman@wmich.edu.

Associate editor: Belinda Chang

## Abstract

Next-generation sequencing has resulted in an explosion of available data, much of which remains unstudied in terms of biochemical function; yet, experimental characterization of these sequences has the potential to provide unprecedented insight into the evolution of enzyme activity. One way to make inroads into the experimental study of the voluminous data available is to engage students by integrating teaching and research in a college classroom such that eventually hundreds or thousands of enzymes may be characterized. In this study, we capitalize on this potential to focus on SABATH methyltransferase enzymes that have been shown to methylate the important plant hormone, salicylic acid (SA), to form methyl salicylate. We analyze data from 76 enzymes of flowering plant species in 23 orders and 41 families to investigate how widely conserved substrate preference is for SA methyltransferase orthologs. We find a high degree of conservation of substrate preference for SA over the structurally similar metabolite, benzoic acid, with recent switches that appear to be associated with gene duplication and at least three cases of functional compensation by paralogous enzymes. The presence of Met in active site position 150 is a useful predictor of SA methylation preference in SABATH methyltransferases but enzymes with other residues in the homologous position show the same substrate preference. Although our dense and systematic sampling of SABATH enzymes across angiosperms has revealed novel insights, this is merely the “tip of the iceberg” since thousands of sequences remain uncharacterized in this enzyme family alone.

**Key words:** enzyme evolution, plant methyltransferase evolution, paralog compensation, substrate preference evolution.

## Introduction

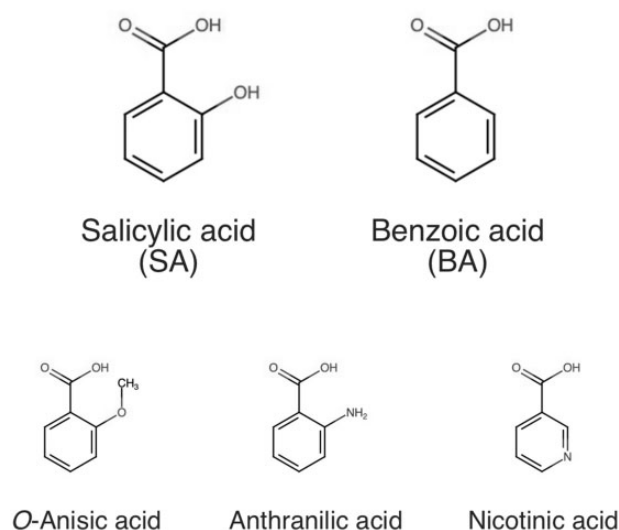
Studies of enzyme function have historically been based on model species which are limited to only a few lineages throughout the tree of life. This approach has yielded many important insights and formed a firm foundation upon which to understand variation and evolution of novelty. Yet, ideally, enzyme activity would be determined from throughout the tree of life such that comprehensive understanding of structure–function relationships and the evolutionary mechanisms of functional change would emerge (Mandoli and Olmstead 2000; Chang et al. 2016). Indeed, comparative

analyses of RUBISCO from over 100 distinct enzymes have helped to illuminate the nature of constraint on substrate discrimination (Bouvier et al. 2021). Although comprehensive studies were impractical 20 years ago, the genomic revolution has allowed for determination of sequences from nearly every major lineage of organisms. For example, the OneKP project generated transcriptomic data from more than 1,000 species of green plants (Leebens-Mack et al. 2019), yet most remain uncharacterized in terms of function. Such rich data sets provide unprecedented opportunities to experimentally study protein functional change among the many lineages that have colonized earth in a diversity of habitats.

To begin to utilize the massive data now available, we have taken an approach that integrates learning with discovery by developing a lab-based class to test evolutionary biochemical hypotheses. In this case, genes were cloned from previously unstudied plants and heterologous protein expression was used to test predictions related to enzyme substrate preference shifts. The class was designed to improve understanding of general evolutionary principles, promote phylogenetic perspectives to understand protein functional change, develop expertise with essential lab skills, and provide a research experience with unknown outcomes (see Materials and Methods for additional details). The data set and discoveries we present were largely generated from these classroom research projects.

Plants are well-known for their ability to produce a wide array of specialized metabolites including terpenoids, flavonoids, alkaloids, and phenylpropanoids that appear to have important and varied roles, particularly for interactions with the environment (Pichersky and Lewinsohn 2011). Methylation is a particularly common enzymatically catalyzed substitution in the biosynthetic pathways of these varied metabolites and often involves enzymes of the salicylic acid/benzoic acid/theobromine (SABATH) methyltransferase (MT) family (Noel et al. 2003). Numerous SABATH enzymes have been functionally characterized including those that methylate a wide array of carboxylic acids (Ross et al. 1999; Kolosova et al. 2001; Seo et al. 2001; Yang et al. 2006; Kapteyn et al. 2007; Varbanova et al. 2007; Murata et al. 2008; Hippauf et al. 2010; Petronikolou et al. 2018; Wu et al. 2018) as well as thiols and ring nitrogen atoms (Kato et al. 1996; Uefuji et al. 2003; McCarthy and McCarthy 2007; Zhao et al. 2012; Huang et al. 2016). The enzyme family appears particularly diverse in angiosperms with approximately 25–40 members in *Arabidopsis* and *Oryza*, whereas mosses and liverworts appear to have only approximately four to nine enzymes encoded in their genomes (D'Auria et al. 2003; Zhao et al. 2008, 2012; Zhang et al. 2019).

The most widely studied SABATH enzyme is salicylic acid carboxyl MT (SAMT) which produces methyl salicylate (MeSA) by methylation of salicylic acid (SA) (Ross et al. 1999) (fig. 1). MeSA has been identified as a component of floral scent and fruit flavor, is emitted after insect herbivory and acts as a vital signaling molecule in plant immune response (Knudsen et al. 2006; Park et al. 2007; Tieman et al. 2010). SAMT was first identified and characterized as a fragrance-producing enzyme in flowers of *Clarkia breweri* (Onagraceae) (Ross et al. 1999). Orthologous SAMT enzymes have since been identified in Plantaginaceae (Scrophulariaceae), Apocynaceae, Solanaceae, Fabaceae, Theaceae, and Salicaceae and variation in terms of substrate preference for structurally related molecules has been shown (Negre et al. 2002; Pott et al. 2004; Effmert et al. 2005; Hippauf et al. 2010; Lin et al. 2013; Han et al. 2018). Whereas some SAMTs are very specific in terms of affinity for SA, several studies have demonstrated activity with the structurally similar compound, benzoic acid (BA) (fig. 1), as well as numerous other molecules (Pott et al. 2004; Hippauf et al. 2010). SAMT appears to be known only from angiosperms; however, a



**FIG. 1.** Representative structures recognized as substrates by the SABATH family of MT enzymes. Most characterized family members are carboxyl MTs and usually exhibit a high degree of preference for one substrate.

recent study of the SABATH family in the gymnosperm genus, *Picea*, and the liverwort genus, *Conocephalum*, identified enzymes that prefer to methylate SA (Chaiprasongsuk et al. 2018; Zhang et al. 2019). Since neither of these enzymes appears to be closely related to SAMT in angiosperms, it is difficult to predict substrate preference from phylogenetic placement within the SABATH family tree, or primary amino acid sequence, alone.

Given the critical role for MeSA in pathogen and herbivore defense documented from model plants, it would be expected that species lacking an SAMT to produce the metabolite would be at a selective disadvantage. Yet, MeSA is only known to be produced from a few handfuls of species that are distantly related (Knudsen et al. 2006). Therefore, it remains to be determined if SAMT orthologs are found throughout the land plant lineage and whether or not specialized enzymatic activity for MeSA production has been maintained over time or has evolved multiple times independently by convergence. SAMT orthologs appear to be absent from the genomes of two model organisms, *Arabidopsis thaliana* and *Oryza sativa* (Chen et al. 2003; Zhao et al. 2010). In those species, benzoic/salicylic acid MT (BSMT), which may be paralogous to SAMT, is capable of BA as well as SA methylation, consistent with the evolutionary mechanism of paralog functional compensation (Hanada et al. 2011; Diss et al. 2014). This makes the extent to which the function of any SABATH family member would be conserved and predictably involved in the methylation of SA unclear.

Crystallographic characterization of SAMT from *C. breweri* yielded insight into the active site residues that bind the methyl donor, S-adenosylmethionine, as well as those that serve to orient SA within the binding pocket to allow for methyl transfer (Zubieta et al. 2003). Mutagenesis of one of these sites, Met150, in SAMT from *Datura wrightii* back to the putatively ancestral His showed an approximately 10-fold

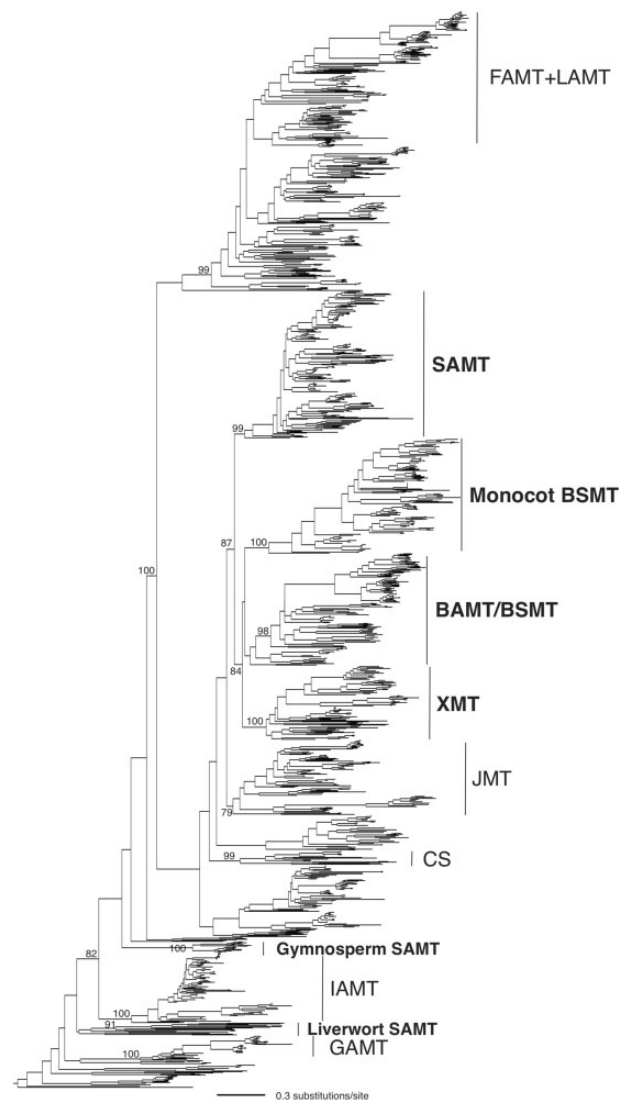
reduction of preference for SA over BA, thereby implicating this residue in substrate discrimination (Barkman et al. 2007). Subsequent studies used ancestral sequence resurrection and mutagenesis to show the evolution of an early preference shift toward SA was governed by a change from His150 to Met early in the history of the SAMT enzyme lineage (Huang et al. 2012). Additional studies have mutated Met150 to His (or vice versa) in several modern-day sequences and shown altered preference for SA (Zubieta et al. 2003; Han et al. 2018). Thus, this active site residue may be a robust predictor of substrate preference but it remains unclear how widespread it is since so few angiosperms have been investigated. Although the functional effects of amino acid substitutions on the evolution of protein function have been studied in SAMT and other plant enzymes for a few focal species (Kaltenegger et al. 2013; Smith et al. 2013), comparative approaches conducted across numerous diverse flowering plant lineages may reveal important variation and increase power for statistical analyses.

A comprehensive gene family history estimate coupled with functional characterization of enzymes has the potential to answer many long-standing questions regarding the evolution of SABATH activity in land plants. Specifically, our objectives in this study were to investigate: 1) how widely SAMT orthologs are found across land plants and to what extent gene duplication and/or loss occurred, 2) how enzyme preference for SA over BA has evolved over 120 My of angiosperm phylogeny in both SAMT orthologs as well as functionally compensating paralogs, and 3) whether Met150 is an accurate predictor of substrate preference throughout the SABATH enzyme family. To do this, we performed phylogenetic analyses of approximately 1,500 SABATH sequences and experimentally assayed enzyme function for more than 70 species from throughout the angiosperm tree of life in a college laboratory classroom setting to illustrate that students can meaningfully contribute to the generation of functional data to study protein evolution.

## Results and Discussion

### Angiosperm SABATH Enzymes That Methylate SA and/or BA Belong to Three Lineages: SAMT, BAMT, BSMT, and Xanthine Alkaloid MT

A maximum likelihood analysis of 1,578 SABATH sequences based on data from 62 complete genomes, in addition to functionally characterized cDNA sequences from select members of various clades, resulted in a single tree that was rooted with green algae sequences (log likelihood  $-328269.41$ ; fig. 2). Overall, branch support is low among the first diverging lineages making it difficult to draw confident conclusions about the most ancient branching patterns in the multigene family; however, there are several observations that are clear regarding the origins of SA- and BA-methylating enzymes. First, all known angiosperm enzymes that methylate SA and BA appear to be part of a well-supported clade (bootstrap proportion [BP] = 100) that is diversified in terms of substrate preference (fig. 2). Within this large group of enzymes is the SAMT lineage (BP=99), for which, preference for SA has been



**Fig. 2.** A phylogenetic analysis of 1,578 SABATH protein MT sequences shows overall enzyme family relationships. Enzyme names shown in bold have been demonstrated to methylate SA and/or benzoic acid (BA). Angiosperm enzymes that methylate SA and/or BA are found in the SAMT, BAMT, BSMT, and XMT clades. The Gymnosperm and Liverwort SAMT enzymes that methylate SA are not closely related to those of angiosperms. Bootstrap support is shown for selected nodes that separate the SA and BA-methylating clades of enzymes from others. SAMT, salicylic acid MT; BAMT, benzoic acid MT; BSMT, benzoic/salicylic acid MT; JMT, jasmonic acid MT; XMT, xanthine alkaloid MT; CS, caffeine synthase; IAMT, indole acetic acid MT; GA, Gibberellic acid MT; FAMT, farnesoic acid MT; LAMT, loganic acid MT.

shown (Ross et al. 1999). SAMT is closely related to a clade of enzymes that includes benzoic acid MT (BAMT), first studied in *Antirrhinum* and BSMT from *Arabidopsis* (BP=98) which has activity with both BA and SA (Murfitt et al. 2000; Chen et al. 2003). Sister to the BAMT/BSMT lineage (BP<50), is the monocot BSMT clade (BP=100), within which, enzymes with preference for benzoic acid, anthranilic acid, and SA have been reported (Kollner et al. 2010; Zhao et al. 2010). Also apparently related to SAMT and BAMT/BSMT is the xanthine



alkaloid MT (XMT) clade (BP=100) (fig. 2). This lineage includes enzymes that produce theobromine and caffeine in *Citrus* and *Coffea* as well as one from *Mangifera* that has high activity with BA, but not SA (Kato and Mizuno 2004; Huang et al. 2016). Second, the *Picea* SABATH enzyme that has been shown to have preference for SA (Chaiprasongsuk et al. 2018) is part of a clade of gymnosperm sequences (BP=100) sister to diverse flowering plant enzymes that include SAMT (fig. 2). Third, the *Conocephalum* MT that shows activity with SA (Zhang et al. 2019) is part of a lineage of liverwort sequences (BP=91) that likely diverged early in the history of the SABATH enzyme family but is not closely related to angiosperm SAMT (fig. 2). Whether these nonangiosperm enzymes represent cases of convergent gains of SA substrate preference or indicate that SA methylation is ancestral for land plants remains to be determined.

### SAMT Is an Ancient Gene Lineage That Has Been Retained in Most Orders of Flowering Plants

A phylogenetic analysis of 629 translated genomic and transcriptomic SAMT sequences, including those from 76 experimentally characterized enzymes (see below), was performed using a maximum likelihood approach (log likelihood  $-93006.968$ ; fig. 3). The estimated relationships among the SAMT proteins from 37 orders of flowering plants are largely as expected with “basal angiosperm” and “basal eudicot” orders estimated as early diverging and most eudicot sequences included in the large Rosid and Asterid clades (fig. 3). This widespread distribution and branching pattern of apparent SAMT orthologs suggests that the enzyme was present in early angiosperms and has been maintained over the approximately 120+ My history of flowering plants. Yet, at least three lineages are notable exceptions that appear to lack SAMT orthologs. First, no orthologous monocot SAMT sequences were obtained even though we queried genomic and transcriptomic data sets of numerous species. Thus, it would appear that SAMT was lost from the monocot lineage early during its divergence. Second, despite queries of multiple genome sequences and transcriptome data, no SAMT orthologs were recovered from Brassicaceae, the best-studied family of Brassicales in which the model organism *A. thaliana* belongs. However, an SAMT is present in *Carica* (papaya), an early branching member of Brassicales (family Caricaceae) (fig. 3), which indicates that the gene was lost later during the history of the order. Third, the absence of SAMT in complete genome sequences of three members of Rosaceae (*Fragaria*, *Prunus*, and *Malus*) suggests loss of the gene from the rose family because orthologs have been isolated from several other members of Rosales, the order to which it belongs (fig. 3). Lineage-specific gene deletion was also reported for GAMT (see fig. 2) which was shown to be lost numerous times throughout angiosperm history (Zhang et al. 2020). Of the 12 other orders/lineages that appear to be missing SAMT, genome sequences remain uncharacterized, making it difficult to discern whether they possess an ortholog or not. However, in the case of Geraniales, which will be discussed below, transcriptome data indicate that *Geranium*

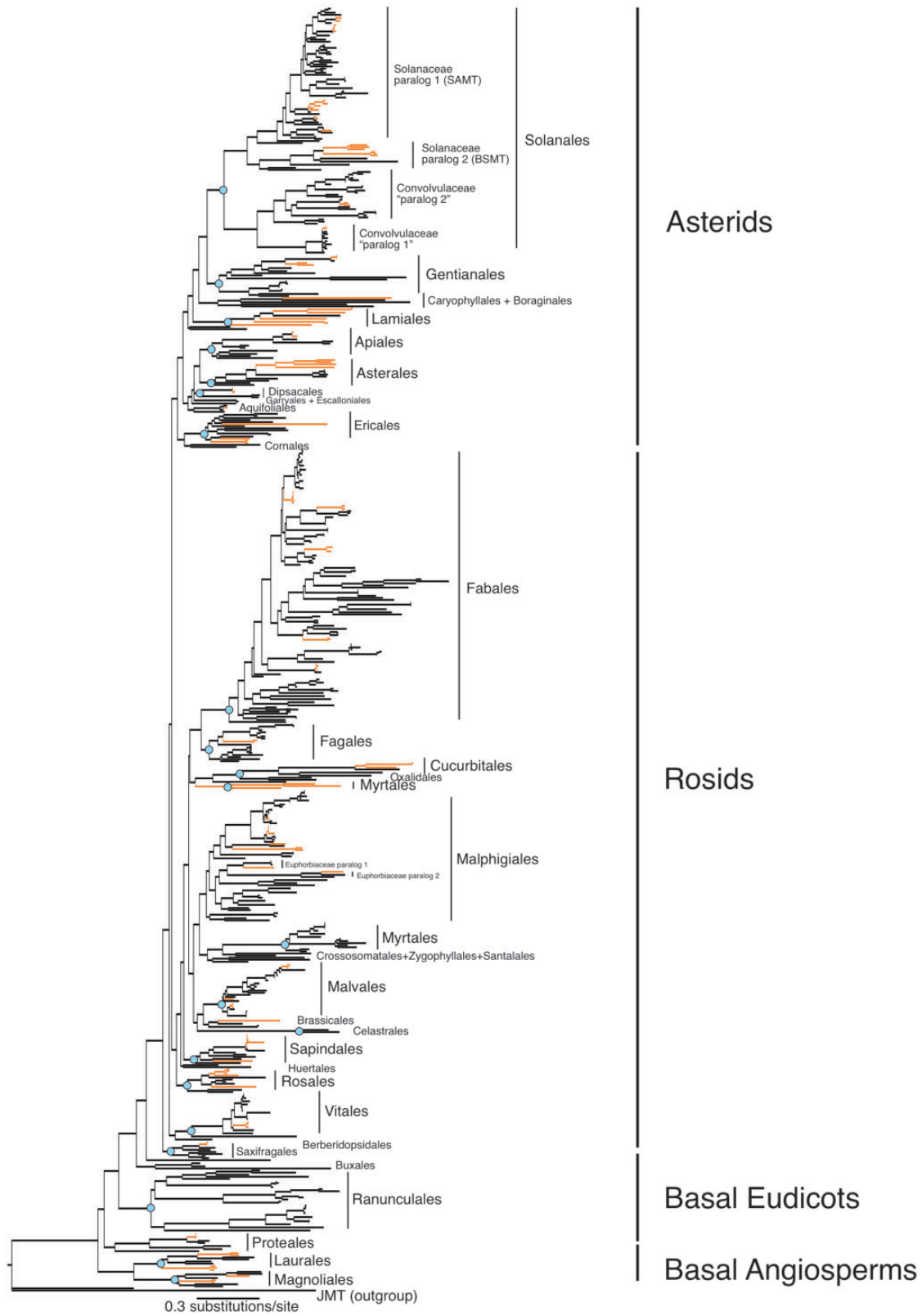
expresses numerous other SABATH sequences; yet, none are orthologous to SAMT.

Although some orders appear to have lost SAMT orthologs, within others it is apparent that recent family- or genus-specific duplications/proliferations have occurred. For instance, SAMT appears to have duplicated early in Solanaceae family history since most sampled species appear to possess at least two divergent copies (fig. 3), as has been reported previously (Barkman et al. 2007). Likewise, *Ipomoea* (Convolvulaceae), also in Solanales, appears to have independently experienced duplication of SAMT since it expresses at least two SAMT-type sequences (fig. 3). Recent SAMT-specific duplications are also inferred from the estimated branching patterns in *Ricinus* (Euphorbiaceae) in Malpighiales (fig. 3). These particular duplications are highlighted because we report on their functional assays below. Quantifying the precise number of recent duplicated genes in each lineage is challenging given the combination of both genomic and transcriptomic sequences in our analysis so, although there appear to be others within various families (e.g., Fabaceae and Vitaceae), we do not describe them further.

### Relative Enzyme Preference for SA over BA Appears to Be Anciently Conserved in Flowering Plant SAMT

Relative substrate preference for SA and BA has been characterized for 76 enzymes from the SAMT lineage that were isolated from 61 genera, 40 families, and 23 orders of angiosperms (fig. 4 and fig. 3, branches highlighted in orange). Fifty-six of these enzymes are newly reported (supplementary table S1, Supplementary Material online) and increase the number from six previously sampled families due to the upper division undergraduate/first-year graduate level, single-semester college course we implemented to study enzyme evolution. Previously unstudied plant species were chosen to clone genes from by extracting RNA from various tissues, using bioinformatics to design primers and performing reverse transcriptase-polymerase chain reaction (RT-PCR) for cDNA amplification (supplementary table S1, Supplementary Material online). Not all plant tissues yielded usable high-quality RNA or expressed the SAMT gene which necessitated switching to a different tissue/species when time permitted. In approximately 60% of cases, cDNA was obtained and successfully cloned and enzyme activity was detected from the heterologously expressed protein.

In the majority of assays, SA is the preferred substrate across the flowering plant lineage (fig. 4). The fact that most SAMT enzymes have preference for SA, including those from the early diverging basal angiosperms, *Persea*, *Sassafras*, *Calycanthus*, *Magnolia*, and *Annona*, as well as those from more recently diverged Asterids and Rosids, strongly suggests that this enzymatic trait has been conserved for at least 125 My in angiosperms. Such conservation of function indicates an important role for MeSA production in flowering plants, as has been demonstrated in the few plants for which its biosynthesis has been experimentally altered (Park et al. 2007; Chen et al. 2019). Nonetheless, some enzymes within the SAMT lineage deviate and exhibit a preference for BA over SA including relatively recent duplicates isolated from



**Fig. 3.** A phylogenetic estimate of SAMT sequences shows that most orders of angiosperms possess orthologs except for monocots. Because SAMT has been retained in several basal angiosperms, basal eudicots, rosids and asterids, orthologs are predicted to have been present in ancestral angiosperms. Lineages marked in orange include at least one enzyme for which functional analyses have been performed as shown in [figure 4](#). Nodes for orders supported by bootstrap proportions >90 are marked by filled blue circles.





*Nicotiana* (Hippauf et al. 2010), *Ipomoea*, and *Ricinus* (fig. 4). In these cases, each genus possesses one enzyme with substrate preference for SA and a second that shows higher preference for BA (fig. 4). Additionally, SAMT orthologs with preference for *O*-anisic acid or nicotinic acid (see fig. 1 for structures) have been reported in *Nicotiana* (Hippauf et al. 2010). Some degree of enzymatic neofunctionalization is not surprising in these cases given that one duplicate appears to be preserved for SA methylation thereby allowing others to evolve alternative substrate preferences. The only other exceptional SAMTs include one from *Nelumbo* that prefers to methylate BA over SA and one enzyme from *Medicago* which appears to have recently gained the preferential methylation of anthranilic acid over either BA or SA (Pollier et al. 2019). The genomes from both of these genera encode additional SAMT duplicates that remain uncharacterized.

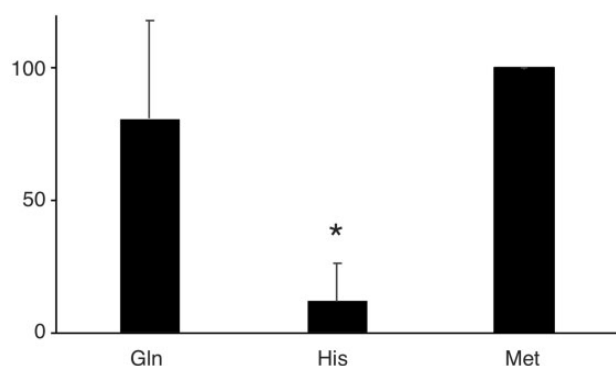
### SABATH Paralogs Appear to Functionally Compensate for Lack of SAMT Orthologs in at Least Three Angiosperm Lineages

Due to the apparent conserved preference for SA methylation across the angiosperm lineage, it is, perhaps, not surprising that loss of SAMT in specific lineages has been compensated for by SABATH paralogs. First, even though no SAMT ortholog is known from monocots, it has been previously shown that BSMT in *Oryza* exhibits preference for SA over BA (Koo et al. 2007; Zhao et al. 2010). Activity with SA and BA has likely been maintained since relatively early in the history of the monocot lineage because the BSMT orthologs we characterized from *Musa* and *Phalaenopsis* (Zingiberales and Asparagales, respectively) show activity with both substrates (fig. 4) as does a previously reported one from *Lilium* (Liliales) (Wang et al. 2015). Second, the phylogenetic analysis and characterization of a *Carica* SAMT sequence (fig. 4) indicates that orthologs were present in the ancestor of Brassicales and capable of SA methylation but were later lost some time after the divergence of *Arabidopsis* (and likely all of Brassicaceae since orthologs appear to be missing from all sequenced genera in the family). Nonetheless, the paralogous *Arabidopsis* BSMT appears to compensate for the loss of SAMT orthologs because, although it prefers to methylate BA (fig. 4), the enzyme demonstrates detectable *in vitro* and *in planta* activity with SA (Chen et al. 2003). Third, two members of Rosaceae, *Malus* and *Prunus*, are lacking SAMT coding sequences even though other members of Rosales (*Morus*, *Ficus*, and *Humulus*) have functional orthologs preferring to methylate SA (fig. 4). Instead, these two genera possess XMT-type enzymes that preferentially catalyze the formation of methyl salicylate (fig. 4). These enzymes are most closely related to those that are otherwise known for xanthine alkaloid methylation in *Coffea* and *Citrus* (McCarthy and McCarthy 2007; Huang et al. 2016). Finally, it appears that *Geranium* also expresses an XMT-type enzyme that shows maximal activity with SA (fig. 4). However, whether the XMT-type enzyme compensates for lack of SAMT is unclear since no genomic sequence is available for that genus or the order Geraniales in which it

belongs. Our findings of functional compensation by close paralogs to maintain MeSA production reinforce conclusions from *Arabidopsis* knock-out mutants (Hanada et al. 2011).

### Met150 Is a Reliable Predictor of Substrate Preference

Because previous studies have shown an important functional role for Met replacement of His150 of the ancestral SABATH enzyme (Node A in fig. 4), we predicted that SA substrate preference would be associated with Met150 (Barkman et al. 2007; Huang et al. 2012). Yet, to our surprise, Met was found in the active site of only 41 out of 66 of the SAMT-type enzymes showing SA methylation preference (fig. 4). The other 25 SA-preferring SAMT enzymes have Gln in the homologous position 150. Previous analyses of Huang et al. (2012) that inferred a mechanistic role of Met for the evolution of SA preference in the ancient enzyme of Node B (fig. 4) had fewer orthologs available for comparative analyses and only included *Clarkia* as a member of the Rosid clade, which has Met like most Asterids (fig. 4). Now, with this broad sampling of Rosid sequences, ancestral state estimates indicate that Gln, not Met, replaced His in the ancestor of core eudicots (Node B) and was retained by nearly all Rosids with subsequent parallel substitution to Met separately in Myrtales and Malpighiales (fig. 4). Even though Met150 may not be responsible for the evolution of SA preference in the progenitor of the SAMT lineage as hypothesized by Huang et al. (2012), it remains a robust predictor of substrate preference since every enzyme with Met in the active site prefers to methylate SA (fig. 4). Furthermore, a phylogenetic correlation test (Pagel 1994) revealed that a model assuming dependence of SA preference on evolving Met at active site position 150 was a significantly better fit to the data than the null model of no association ( $P = 0.05$ ). Under the model of trait dependence, the rate of gain of SA preference after evolving Met in the active site was 17 times higher than the rate estimated under the model of independent evolution of active site residue and substrate preference (supplementary fig. S1, Supplementary Material online). The higher estimated rate may be due to the two independent substitutions of Met150 in the paralogous BAMS/BASMT and XMT lineages that appear coincident with the evolution of SA preference (fig. 4). Specifically, although the monocot BSMT clade of enzymes appear to have ancestrally had His in active site position 150, at least one historical shift to Met is estimated such that the descendant enzymes in *Phalaenopsis*, *Musa*, and *Oryza* evolved preference for SA over BA (fig. 4). Likewise, in the XMT clade, Met apparently replaced His in the common ancestor of *Prunus* and *Malus* (both Rosaceae), as well as *Geranium*, and their enzymes prefer to methylate SA (fig. 4). For context, our survey of the 1,578 SABATH sequences of figure 2 shows that across land plants, 69.6% have His, 13.6% have Gln, 9.7% have Met, 3.1% have Asn, whereas 3.9% show other amino acids at active site position 150. Of the 9.7% sequences with Met, all are found in the SAMT/BSMT/XMT clades. Thus, the association of Met with these SA-preferring enzymes appears to be nonrandom. In fact, a phylogenetic ANOVA (Garland et al. 1993) found a significant difference in the level of SA preference in enzymes with Met



**Fig. 5.** Average level of SA preference for enzymes with different amino acid residues in active site position 150. Phylogenetic ANOVA results indicate that enzymes with Gln and Met are not significantly different from one another. However, enzymes with His in the active site have significantly lower SA methylation preference than those with either Gln or His ( $P = 0.003$  in both cases).

compared with His; however, the difference of SA preference between enzymes with Met versus Gln was not significant at  $P < 0.05$  (fig. 5 and supplementary table S2, Supplementary Material online).

The phylogenetic ANOVA also showed a significant difference in SA preference for enzymes with Gln in the active site relative to those with His (fig. 5 and supplementary table S2, Supplementary Material online). However, the association of Gln at position 150 with SA preference appears to be context-dependent. There are 32 enzymes with Gln in figure 4 but seven of those show preference for other substrates. For instance, the duplicate SAMT-type sequence with Gln150 from *Ipomoea* prefers to methylate BA as compared with SA (fig. 4). In addition, the *Medicago truncatula* enzyme orthologous to SAMT preferentially methylates anthranilic acid even though it exhibits Gln in the active site (Pollier et al. 2019). In contrast, of the 12 enzymes shown in figure 4 that have His in the active site position 150, none prefer to methylate SA. Yet, the SA-preferring enzymes from the liverwort, *Conocephalum* (Zhang et al. 2019) and the Gymnosperm, *Picea* (Chaiprasongsuk et al. 2018) both possess His at the homologous position in the active site. These results strongly imply that additional amino acid positions, perhaps near the active site, provide altered substrate preference. In order to dissect the evolutionary basis for substrate preference shifts in these SABATH enzymes, it will likely require ancestral sequence resurrection coupled with mutagenesis (Thornton 2004) and integration of structural, computational, and experimental data (Torrens-Spence et al. 2020).

It is clear that the increased number of enzymes assayed here has advanced our understanding of the SABATH family in terms of shifts of substrate preference and associations with particular active site residues. Further experimental study of many more SABATH family members will likely provide finer-scale resolution of the relationship of active site configuration and substrate preference and has the potential to yield additional insights into how functional diversification among paralogous enzymes evolves. Although computational approaches may assist protein functional annotation broadly

(Furnham et al. 2012), engagement of as many students as possible in the process of experimental investigations could prove to be a fruitful way to understand the evolution of enzymatic properties and empirically assay the voluminous sequences that remain uncharacterized. This study demonstrates one possible outcome of bringing hands-on scientific inquiry into the college classroom for the purpose of understanding the functional significance of sequence variation across the Tree of Life.

## Materials and Methods

### Implementation of Collaborative Classroom Investigations

The overall goal of each semester-long class was to provide an integrative scientific experience that started with a living plant and ended up with metabolomic data obtained from gas chromatography-mass spectrometry (GC-MS) analyses to test hypotheses about the degree to which historical amino acid substitutions in SABATH proteins might promote enzyme functional evolution. Three major evolutionary concepts were the focus for the course. First, natural selection may impose variable levels of constraint upon different parts of proteins. All target sequences were manually aligned to a data set with representatives from each lineage of the SABATH family of enzymes. From this, it was demonstrated that although proteins from different species exhibit some sites that were hard to align and may be neutral, others are highly conserved and align easily because they presumably are under strong purifying selection such that they rarely mutate. Second, gene duplication and phylogenetic divergence can generate protein diversity. Subsequent to alignments, phylogenetic analyses of candidate proteins were performed in order to assess orthology/paralogy and make predictions about enzyme activity based on relatedness to proteins with demonstrated functions. Third, protein functional changes arise from historical amino acid replacements. This was investigated by mapping enzyme relative substrate preferences from the classroom investigations and published literature onto the phylogenetic tree and assessing variation relative to active site amino acid changes that may be correlated.

In order to collect novel data, basic principles of bioinformatics, primer design, RNA extraction, RT-PCR, cloning, transformation, protein expression, enzyme assays, and GC-MS analysis of metabolites were introduced and protocols both from the literature and commercially available kits were provided and are described in relevant sections below. These protocols were implemented independently for different taxa and lab notebooks documented all results, as expected for any scientific study. In any given semester, lab reports were written based upon analyzed and pooled class data that were interpreted relative to published literature. Through several semesters of collaboration, the classroom investigations revealed novel results including previously unknown paralog compensation and the correlation of Met150 with SA preference in XMT orthologs and that genus-specific SAMT duplicates may partition substrate preferences. In addition, these studies pointed to the apparent equivalence of



Gln150 and Met150 for predicting SA preference of most SAMTs. The generation of the novel data reported here also provided the opportunity to participate in all steps of the peer-review publication process related to this manuscript.

### Phylogenetic Methods

Trees were estimated from amino acid translations from a combination of genome and transcriptome sequences obtained from GenBank, 1KP, and Phytozome databases using the *C. breweri* SAMT (Acc. No. AF133053.1) as a query. Additional genome sequences were obtained through a BLAST search of the *Ginkgo biloba* (Guan et al. 2016), *Picea abies* (Nystedt et al. 2013), *Azolla filiculoides*, and *Salvinia cucullata* (Li et al. 2018) genome repositories. Sequences were aligned using the alignment software MAFFT (Katoh and Standley 2013) using default parameters. Trees and branch support estimates were calculated using IQTree (Nguyen et al. 2015) via the CIPRES Science Gateway. The JTT model of amino acid substitution and number of free rate categories were determined by IQTree's ModelFinder algorithm (Kalyaanamoorthy et al. 2017), and branch support values were generated using IQTree's ultrafast bootstrapping method (Hoang et al. 2018). Ancestral states were estimated using the "-asr" function in IQTree.

Statistical analyses of the dependence of SA preference upon Met in active site position 150 were performed across the lineages of figure 4 according to the models of Pagel (1994) as implemented in Mesquite (Maddison and Maddison 2021). The dependent model with eight rate parameters was compared with the six parameter model in which  $q_{12}$  and  $q_{34}$  were constrained to be equal, as were those of  $q_{21}$  and  $q_{43}$  (supplementary fig. S1, Supplementary Material online). Significance of the test was assessed using 1000 simulations.

In order to test for significant differences in relative substrate preference for SA according to the active site residue at position 150 (Met, Gln, or His) shown for enzymes in figure 4, we used phylogenetically explicit ANOVA (Garland et al. 1993) implemented in the R package "phytools" using the function, phylANOVA (Revell 2012). For the phylogenetic ANOVA, we performed 1,000 simulations and the relative substrate preference values were  $\log_{10}$  transformed to reduce the effect of differences in magnitude.

### Gene Isolation and Amplification

All plant tissues were obtained from the Western Michigan University Finch greenhouse or campus grounds. Roughly 100 mg of fresh tissue was flash-frozen using liquid nitrogen and subsequently ground into a fine powder using a sterile pestle. The tissue extracted for each species is listed in supplementary table S1, Supplementary Material online. RNA extractions were then performed using the RNeasy Plant Mini Kit (Qiagen Inc., Valencia, CA) or the Spectrum Plant Total RNA kit (Sigma-Aldrich, St. Louis, MO), according to manufacturer protocols. RNA was eluted using 40  $\mu$ l sterile RNase-free water and stored at  $-80^{\circ}\text{C}$ . RNA extraction quality was confirmed via electrophoresis in a 1.5% agarose

gel and concentration was determined using a NanoDrop spectrophotometer (ThermoFisher Scientific, Waltham, MA).

To obtain primers for RT-PCR of putative SAMT orthologs, the full amino acid sequence from *C. breweri* SAMT (GenBank accession number AF133053.1) was used as a query for BLAST searches (TblastN) of the GenBank nonredundant and transcriptome shotgun assembly databases, as well as the OneKP data set from China National Gene Bank. Forward and reverse primers were subsequently designed from full-length sequences to amplify the cDNAs of the SAMT genes. If DNA sequence data were not available for the species chosen, primers were designed from sequences of congeners. All primer sequences are listed in supplementary table S1, Supplementary Material online.

To obtain cDNA, the Superscript III One-Step RT-PCR System with Platinum Taq (Invitrogen, Carlsbad, CA) was used for RT-PCR. Each reaction consisted of 0.5  $\mu$ l of each primer (10  $\mu$ M), 12.5  $\mu$ l Invitrogen 2 $\times$  reaction mix, 1  $\mu$ l SSIII Platinum Taq mix, 500 ng total RNA, and RNase-free water to bring the reaction to a final volume of 25  $\mu$ l. RT-PCR utilized the following conditions: reverse transcription at  $50^{\circ}\text{C}$  for 10 min, followed by 40 cycles of denaturation at  $94^{\circ}\text{C}$  for 30 s, annealing at  $50^{\circ}\text{C}$  for 30 s and extension at  $72^{\circ}\text{C}$  for 70 s. The final extension was held for 25 min at  $72^{\circ}\text{C}$ . Accession numbers for the cDNA sequences encoding each assayed enzyme are provided in supplementary table S1, Supplementary Material online.

### Transformation and Protein Expression in *Escherichia coli*

Following amplification, cDNA was integrated into a pTrcHIS vector using the pTrcHIS TOPO TA Expression Kit or the pBAD plasmid using the pBAD TOPO TA kit (both were from Invitrogen, Carlsbad, CA). The TOPO TA cloning kits were chosen because they allow for rapid cloning in a college classroom setting. Ligation of cDNA into the pTrcHIS and pBAD vectors and subsequent transformation into Top10 *E. coli* cells was carried out according to the manufacturer's protocol. About 100  $\mu$ l of the transformation mixture were pipetted onto LB plates containing 50  $\mu$ g/ml ampicillin and incubated overnight at  $37^{\circ}\text{C}$ .

Colonies were screened for full-length inserts by PCR, using the pTrcHIS or pBAD forward and reverse primers. Each PCR reaction contained 0.5  $\mu$ l of each primer (10  $\mu$ M), 12.5  $\mu$ l Qiagen Taq 2 $\times$  PCR master mix (Qiagen Inc., Germantown, MD), and water to bring the reaction volume up to a total of 24.5  $\mu$ l. Bacteria were transferred into the reaction tube using a sterile pipette tip (estimated volume 0.5  $\mu$ l). Products containing full-length inserts were purified using the QIAGEN QIAquick PCR Purification Kit (Qiagen Inc., Germantown, MD) according to the manufacturer's protocol. Purified product was submitted to Genewiz Corp. (South Plainfield, NJ) for Sanger sequencing to confirm sequence orientation and that full-length open reading frames were obtained.

Pilot bacterial cultures were prepared in 5 ml LB medium with 50  $\mu$ g/ml ampicillin from colonies containing full-length, sense-orientation sequences. These pilot cultures were incubated with shaking overnight at  $37^{\circ}\text{C}$ . The following day,

48 ml of LB-amp was inoculated using 2 ml of the pilot culture and incubated at 37 °C until reaching an optical density of 0.6 at 600 nm. Protein expression was then induced by addition of IPTG to a final concentration of 1 mM for cultures transformed with ptrCHIS or arabinose to a final concentration of 0.2% for pBAD-containing cultures.

### Enzyme Assay with SA and BA

In order to determine substrate preference for heterologously expressed enzymes, we utilized the approach of Ross et al. (1999) as extended by Barkman et al. (2007). This approach assays enzyme activity with SA and BA directly from 50 ml *E. coli* LB cultures and was used for several reasons. First, it provides a direct measure of substrate preference since the substrates are provided in equimolar concentrations and are saturating. Indeed, our comparisons of substrate preference using the Barkman et al. (2007) method appear to be qualitatively comparable with kinetic estimates obtained from purified enzyme preparations for the cases we have examined. For example, we determined the ratio of specificity constants for SA and BA of the *D. wrightii* SAMT as follows:  $K_{cat}/K_M(SA):K_{cat}/K_M(BA) \approx 100$ . Using the method of Barkman et al. (2007) with *D. wrightii* SAMT, we obtained the following ratio of methylated products: MeSA:MeBA  $\approx 20$ . Therefore, both of these approaches indicate a clear qualitative preference for SA methylation over BA methylation. Second, several of the expressed proteins were not soluble when extracted; therefore, they could not be assayed by other methods such as the radioactive method of Huang et al. (2012) which only measures activity with single substrates. Third, given that most of the enzymes were isolated and assayed in the context of a one semester college course, complete kinetic characterization using radioisotopes was impractical.

For GC-MS assays of the Barkman et al. (2007) method, equimolar concentrations of SA and BA (0.2 mM final concentration) were added to the 50 ml LB cultures with 50 µg/ml ampicillin immediately after the IPTG was added and were shaken gently overnight at room temperature. These were then centrifuged at 4 °C for 10 min at 3,000×g to pellet cells. The supernatant was then extracted with 5 ml hexane to remove methylated products. The contents of hexane extracts were analyzed using GC-MS. The GC-MS analyses were performed on an HP6890 GC System coupled to an HP5973 Mass Selective Detector using a DB-5 capillary column. The oven conditions were 40 °C for 2 min, ramping 20 °C/min to 300 °C with a 2 min hold. Relative activity was determined by peak integration and comparing peak areas for each methylated product. Because we tested hypotheses related to relative substrate preference, using the same protein concentration in each assay was not required; in this case, providing competing substrates in equimolar concentrations is the most critical aspect. Replicate transformations were performed and mean and standard deviation of the ratio of products were determined. Importantly, 50 ml cultures in which an antisense vector was used for transformations served as negative controls.

## Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

## Acknowledgments

Chris Jackson is thanked for growing plants in the greenhouse that were used for this work. The WMU Department of Chemistry is thanked for providing access to GC-MS for the duration of our study. Kevin Blair, John Spitsbergen, Yan Lu, Ricky Stull, and two anonymous reviewers provided technical support and/or feedback on the project and earlier drafts of the manuscript. GenBank accession numbers for the cDNA studied in this work are OL661269–OL661331. This work was supported by National Science Foundation grant (MCB-1120624 to T.J.B.).

## Data Availability

Individuals interested in the data matrices underlying this article may obtain them upon request of the corresponding author. The original data underlying this article are available in <https://www.ncbi.nlm.nih.gov/genbank/> as well as <https://db.cngb.org/onekp/> (both of which were last accessed May 01, 2021).

## References

- Barkman TJ, Martins TR, Sutton E, Stout JT. 2007. Positive selection for single amino acid change promotes substrate discrimination of a plant volatile-producing enzyme. *Mol Biol Evol.* 24(6):1320–1329.
- Bouvier JW, Emms DM, Rhodes T, Bolton JS, Brasnett A, Eddershaw A, Nielsen JR, Unitt A, Whitney SM, Kelly S. 2021. Rubisco adaptation is more limited by phylogenetic constraint than by catalytic trade-off. *Mol Biol Evol.* 38(7):2880–2896.
- Chaiprasongsuk M, Zhang C, Qian P, Chen XL, Li GL, Trigiano RN, Guo H, Chen F. 2018. Biochemical characterization in Norway spruce (*Picea abies*) of SABATH methyltransferases that methylate phytohormones. *Phytochemistry* 149:146–154.
- Chang CR, Bowman JL, Meyerowitz EM. 2016. Field guide to plant model systems. *Cell* 167(2):325–339.
- Chen F, D'Auria JC, Tholl D, Ross JR, Gershenzon J, Noel JP, Pichersky E. 2003. An *Arabidopsis thaliana* gene for methylsalicylate biosynthesis, identified by a biochemical genomics approach, has a role in defense. *Plant J.* 36(5):577–588.
- Chen L, Wang WS, Wang T, Meng XF, Chen TT, Huang XX, Li YJ, Hou BK. 2019. Methyl salicylate glucosylation regulates plant defense signaling and systemic acquired resistance. *Plant Physiol.* 180(4):2167–2181.
- D'Auria JC, Chen F, Pichersky E. 2003. The SABATH family of MTS in *Arabidopsis thaliana* and other plant species. In: Romeo JT, editor. Recent advances in phytochemistry. Oxford: Elsevier Science Ltd. p. 253–283.
- Diss G, Ascencio D, Deluna A, Landry C.R. 2014. Molecular mechanisms of paralogous compensation and the robustness of cellular networks. *J Exp Zool B Mol Dev Evol.* 322(7):488–499.
- Effmert U, Saschenbrecker S, Ross J, Negre F, Fraser CM, Noel JP, Dudareva N, Piechulla B. 2005. Floral benzenoid carboxyl methyltransferases: from in vitro to in planta function. *Phytochemistry* 66(11):1211–1230.
- Furnham N, de Beer TAP, Thornton JM. 2012. Current challenges in genome annotation through structural biology and bioinformatics. *Curr Opin Struct Biol.* 22(5):594–601.
- Garland TJ, Dickerman AW, Janis CM, Jones JA. 1993. Phylogenetic analysis of covariance by computer simulation. *Syst Biol.* 42(3):265–292.

- Guan R, Zhao YP, Zhang H, Fan GY, Liu X, Zhou WB, Shi CC, Wang JH, Liu WQ, Liang XM, et al. 2016. Draft genome of the living fossil *Ginkgo biloba*. *Gigascience* 5(1):49.
- Han XM, Yang Q, Liu YJ, Yang ZL, Wang XR, Zeng QY, Yang HL. 2018. Evolution and function of the *Populus* SABATH family reveal that a single amino acid change results in a substrate switch. *Plant Cell Physiol*. 59(2):392–403.
- Hanada K, Sawada Y, Kuromori T, Klausnitzer R, Saito K, Toyoda T, Shinozaki K, Li W-H, Hirai MY. 2011. Functional compensation of primary and secondary metabolites by duplicate genes in *Arabidopsis thaliana*. *Mol Biol Evol*. 28(1):377–382.
- Hippauf F, Michalsky E, Huang R, Preissner R, Barkman TJ, Piechulla B. 2010. Enzymatic, expression and structural divergences among carboxyl O-methyltransferases after gene duplication and speciation in *Nicotiana*. *Plant Mol Biol*. 72(3):311–330.
- Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. 2018. UFBoot2: improving the ultrafast bootstrap approximation. *Mol Biol Evol*. 35(2):518–522.
- Huang R, Hippauf F, Rohrbeck D, Hausteiner M, Wenke K, Feike J, Sorrelle N, Piechulla B, Barkman TJ. 2012. Enzyme functional evolution through improved catalysis of ancestrally nonpreferred substrates. *Proc Natl Acad Sci U S A*. 109(8):2966–2971.
- Huang R, O'Donnell AJ, Barboline JJ, Barkman TJ. 2016. Convergent evolution of caffeine in plants by co-option of exapted ancestral enzymes. *Proc Natl Acad Sci U S A*. 113(38):10613–10618.
- Kaltenegger E, Eich E, Ober D. 2013. Evolution of homospermidine synthase in the Convolvulaceae: a story of gene duplication, gene loss, and periods of various selection pressures. *Plant Cell* 25(4):1213–1227.
- Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermini LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods*. 14(6):587–589.
- Kapteyn J, Qualley AV, Xie ZZ, Fridman E, Dudareva N, Gang DR. 2007. Evolution of cinnamate/p-coumarate carboxyl methyltransferases and their role in the biosynthesis of methylcinnamate. *Plant Cell* 19(10):3212–3229.
- Kato M, Kanehara T, Shimizu H, Suzuki T, Gillies FM, Crozier A, Ashihara H. 1996. Caffeine biosynthesis in young leaves of *Camellia sinensis*: in vitro studies on N-methyltransferase activity involved in the conversion of xanthosine to caffeine. *Physiol Plant*. 98(3):629–636.
- Kato M, Mizuno K. 2004. Caffeine synthase and related methyltransferases in plants. *Front Biosci*. 9(1–3):1833–1842.
- Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 30(4):772–780.
- Knudsen JT, Eriksson R, Gershenzon J, Stahl B. 2006. Diversity and distribution of floral scent. *Bot Rev*. 72(1):1–121.
- Kollner TG, Lenk C, Zhao N, Seidl-Adams I, Gershenzon J, Chen F, Degenhardt J. 2010. Herbivore-induced SABATH methyltransferases of maize that methylate anthranilic acid using S-adenosyl-L-methionine. *Plant Physiol*. 153(4):1795–1807.
- Kolosova N, Sherman D, Karlson D, Dudareva N. 2001. Cellular and subcellular localization of S-adenosyl-L-methionine:benzoic acid carboxyl methyltransferase, the enzyme responsible for biosynthesis of the volatile ester methylbenzoate in snapdragon flowers. *Plant Physiol*. 126(3):956–964.
- Koo YJ, Kim MA, Kim EH, Song JT, Jung C, Moon JK, Kim JH, Seo HS, Song SI, Kim JK, et al. 2007. Overexpression of salicylic acid carboxyl methyltransferase reduces salicylic acid-mediated pathogen resistance in *Arabidopsis thaliana*. *Plant Mol Biol*. 64(1–2):1–15.
- Lebens-Mack JH, Barker MS, Carpenter EJ, Deyholos MK, Gitzendanner MA, Graham SW, Grosse I, Li Z, Melkonian M, Mirarab S, et al. 2019. One thousand plant transcriptomes and the phylogenomics of green plants. *Nature* 574:679–684.
- Li FW, Brouwer P, Carretero-Paulet L, Cheng SF, de Vries J, Delaux PM, Eily A, Koppers N, Kuo LY, Li Z, et al. 2018. Fern genomes elucidate land plant evolution and cyanobacterial symbioses. *Nat Plants*. 4(7):460–472.
- Lin JY, Mazarei M, Zhao N, Zhu JW, Zhuang XF, Liu WS, Pantalone VR, Arelli PR, Stewart CN, Chen F. 2013. Overexpression of a soybean salicylic acid methyltransferase gene confers resistance to soybean cyst nematode. *Plant Biotechnol J*. 11(9):1135–1145.
- Maddison WP, Maddison DR. 2021. Mesquite: a modular system for evolutionary analysis. Version 3.70. Available from: <http://www.mesquiteproject.org>.
- Mandoli DF, Olmstead R. 2000. The importance of emerging model systems in plant biology. *J Plant Growth Regul*. 19(3):249–252.
- McCarthy AA, McCarthy JG. 2007. The structure of two N-methyltransferases from the caffeine biosynthetic pathway. *Plant Physiol*. 144(2):879–889.
- Murata J, Roepke J, Gordon H, De Luca V. 2008. The leaf epidermome of *Catharanthus roseus* reveals its biochemical specialization. *Plant Cell* 20(3):524–542.
- Murfitt LM, Kolosova N, Mann CJ, Dudareva N. 2000. Purification and characterization of S-adenosyl-L-methionine:benzoic acid carboxyl methyltransferase, the enzyme responsible for biosynthesis of the volatile ester methyl benzoate in flowers of *Antirrhinum majus*. *Arch Biochem Biophys*. 382(1):145–151.
- Negre F, Kolosova N, Knoll J, Kish CM, Dudareva N. 2002. Novel S-adenosyl-L-methionine:salicylic acid carboxyl methyltransferase, an enzyme responsible for biosynthesis of methyl salicylate and methyl benzoate, is not involved in floral scent production in snapdragon flowers. *Arch Biochem Biophys*. 406(2):261–270.
- Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol*. 32(1):268–274.
- Noel JP, Dixon RA, Pichersky E, Zubieta C, Ferrer JL. 2003. Structural, functional, and evolutionary basis for methylation of plant small molecules. *Recent Adv Phytochem*. 37:37–58.
- Nystedt B, Street NR, Wetterbom A, Zuccolo A, Lin YC, Scofield DG, Vezzi F, Delhomme N, Giacomello S, Alexeyenko A, et al. 2013. The Norway spruce genome sequence and conifer genome evolution. *Nature* 497(7451):579–584.
- Pagel M. 1994. Detecting correlated evolution on phylogenies: a general method for the comparative analysis of discrete characters. *Proc R Soc B Biol Sci*. 255:37–45.
- Park SW, Kaimoyo E, Kumar D, Mosher S, Klessig DF. 2007. Methyl salicylate is a critical mobile signal for plant systemic acquired resistance. *Science* 318(5847):113–116.
- Petronikolou N, Hollatz AJ, Schuler MA, Nair SK. 2018. Loganic acid methyltransferase: insights into the specificity of methylation on an iridoid glycoside. *Chembiochem*. 19(8):784–788.
- Pichersky E, Lewinsohn E. 2011. Convergent evolution in plant specialized metabolism. *Annu Rev Plant Biol*. 62:549–566.
- Pollier J, De Geyter N, Moses T, Boachon B, Franco-Zorrilla JM, Bai YC, Lacchini E, Gholami A, Vanden Bossche R, Werck-Reichhart D, et al. 2019. The MYB transcription factor emission of methyl anthranilate 1 stimulates emission of methyl anthranilate from *Medicago truncatula* hairy roots. *Plant J*. 99(4):637–654.
- Pott MB, Hippauf F, Saschenbrecker S, Chen F, Ross J, Kiefer I, Slusarenko A, Noel JP, Pichersky E, Effmert U, et al. 2004. Biochemical and structural characterization of benzenoid carboxyl methyltransferases involved in floral scent production in *Stephanotis floribunda* and *Nicotiana suaveolens*. *Plant Physiol*. 135(4):1946–1955.
- Revell LJ. 2012. phytools: an R package for phylogenetic comparative biology (and other things). *Methods Ecol Evol*. 3(2):217–223.
- Ross JR, Nam KH, D'Auria JC, Pichersky E. 1999. S-Adenosyl-L-methionine:salicylic acid carboxyl methyltransferase, an enzyme involved in floral scent production and plant defense, represents a new class of plant methyltransferases. *Arch Biochem Biophys*. 367(1):9–16.
- Seo HS, Song JT, Cheong JJ, Lee YH, Lee YW, Hwang I, Lee JS, Choi YD. 2001. Jasmonic acid carboxyl methyltransferase: a key enzyme for jasmonate-regulated plant responses. *Proc Natl Acad Sci U S A*. 98(8):4788–4793.
- Smith SD, Wang S, Rausher MD. 2013. Functional evolution of an anthocyanin pathway enzyme during a flower color transition. *Mol Biol Evol*. 30(3):602–612.



- Thornton JW. 2004. Resurrecting ancient genes: experimental analysis of extinct molecules. *Nat Rev Genet.* 5(5):366–375.
- Tieman D, Zeigler M, Schmelz E, Taylor MG, Rushing S, Jones JB, Klee HJ. 2010. Functional analysis of a tomato salicylic acid methyl transferase and its role in synthesis of the flavor volatile methyl salicylate. *Plant J.* 62(1):113–123.
- Torrens-Spence MP, Chiang Y-C, Smith T, Vicent MA, Wang Y, Weng J-K. 2020. Structural basis for divergent and convergent evolution of catalytic machineries in plant aromatic amino acid decarboxylase proteins. *Proc Natl Acad Sci U S A.* 117(20):10806–10817.
- Uefuji H, Ogita S, Yamaguchi Y, Koizumi N, Sano H. 2003. Molecular cloning and functional characterization of three distinct N-methyltransferases involved in the caffeine biosynthetic pathway in coffee plants. *Plant Physiol.* 132(1):372–380.
- Varbanova M, Yamaguchi S, Yang Y, McKelvey K, Hanada A, Borochoy R, Yu F, Jikumaru Y, Ross J, Cortes D, et al. 2007. Methylation of gibberellins by *Arabidopsis* GAMT1 and GAMT2. *Plant Cell* 19(1):32–45.
- Wang H, Sun M, Li LL, Xie XH, Zhang QX. 2015. Cloning and characterization of a benzoic acid/salicylic acid carboxyl methyltransferase gene involved in floral scent production from lily (*Lilium* ‘Yelloween’). *Genet Mol Res.* 14(4):14510–14521.
- Wu RR, Zhang FX, Liu LY, Li W, Pichersky E, Wang GD. 2018. MeNA, controlled by reversible methylation of nicotinate, is an NAD precursor that undergoes long-distance transport in *Arabidopsis*. *Mol Plant.* 11(10):1264–1277.
- Yang Y, Yuan JS, Ross J, Noel JP, Pichersky E, Chen F. 2006. An *Arabidopsis thaliana* methyltransferase capable of methylating farnesoic acid. *Arch Biochem Biophys.* 448(1–2):123–132.
- Zhang C, Chairprasongsuk M, Chanderbali AS, Chen X, Fu J, Soltis DE, Chen F. 2020. Origin and evolution of a gibberellin-deactivating enzyme GAMT. *Plant Direct.* 4(12):e00287.
- Zhang C, Chen XL, Crandall-Stotler B, Qian P, Kollner TG, Guo H, Chen F. 2019. Biosynthesis of methyl (E)-cinnamate in the liverwort *Conocephalum salebrosum* and evolution of cinnamic acid methyltransferase. *Phytochemistry* 164:150–159.
- Zhao N, Ferrer JL, Moon HS, Kapteyn J, Zhuang XF, Hasebe M, Stewart CN, Gang DR, Chen F. 2012. A SABATH methyltransferase from the moss *Physcomitrella patens* catalyzes S-methylation of thiols and has a role in detoxification. *Phytochemistry* 81:31–41.
- Zhao N, Ferrer JL, Ross J, Guan J, Yang Y, Pichersky E, Noel JP, Chen F. 2008. Structural, biochemical, and phylogenetic analyses suggest that indole-3-acetic acid methyltransferase is an evolutionarily ancient member of the SABATH family. *Plant Physiol.* 146(2):455–467.
- Zhao N, Guan J, Ferrer JL, Engle N, Chern M, Ronald P, Tschaplinski TJ, Chen F. 2010. Biosynthesis and emission of insect-induced methyl salicylate and methyl benzoate from rice. *Plant Physiol Biochem.* 48(4):279–287.
- Zubieta C, Ross JR, Koscheski P, Yang Y, Pichersky E, Noel JP. 2003. Structural basis for substrate recognition in the salicylic acid carboxyl methyltransferase family. *Plant Cell* 15(8):1704–1716.