Enzyme functional evolution through improved catalysis of ancestrally nonpreferred substrates

Ruiqi Huang[a,1], Frank Hippauf[b,1], Diana Rohrbeck[b], Maria Haustein[b], Katrin Wenke[b], Janie Feike[b], Noah Sorrelle[a], Birgit Piechulla[b], and Todd J. Barkman[a,2]

[a]Department of Biological Sciences, Western Michigan University, Kalamazoo, MI 49008; and [b]Institute of Biological Sciences and Biochemistry, University of Rostock, 18059 Rostock, Germany

Edited by Michael Lynch, Indiana University, Bloomington, IN, and approved January 10, 2012 (received for review December 29, 2010)

In this study, we investigated the role for ancestral functional variation that may be selected upon to generate protein functional shifts. We used ancestral protein resurrection, statistical tests for positive selection, forward and reverse evolutionary genetics, and enzyme functional assays. Data are presented for three instances of protein functional change in the salicylic acid/benzoic acid/theobromine (SABATH) lineage of plant secondary metabolite-producing enzymes. In each case, we demonstrate that ancestral nonpreferred activities were improved upon in a daughter enzyme after gene duplication, and that these functional shifts were likely coincident with positive selection. Both forward and reverse mutagenesis studies validate the impact of one or a few sites toward increasing activity with ancestrally nonpreferred substrates. In one case, we document the occurrence of an evolutionary reversal of an active site residue that reversed enzyme properties. Furthermore, these studies show that functionally important amino acid replacements result in substrate discrimination as reflected in evolutionary changes in the specificity constant (k_{cat}/K_M) for competing substrates, even though adaptive substitutions may affect K_M and k_{cat} separately. In total, these results indicate that nonpreferred, or even latent, ancestral protein activities may be coopted at later times to become the primary or preferred protein activities.

carboxyl methyltransferase | paleomolecular biology

The evolution of proteins is responsible in large part for generating the diversity of organismal form and function observed today. As such, understanding the nature of protein functional diversification is a primary goal of evolutionary biologists. In 1970, Ohno's landmark model of protein evolution explained how descendants of single function ancestral proteins may acquire novel functions after gene duplication (1). Subsequent models have explicitly required ancestral proteins to have multiple functions that then are partitioned and/or improved upon after gene duplication. In the cases of duplication–degeneration–complementation (DDC) (2) and escape from adaptive conflict (EAC) (3–6), multiple ancestral functions are selectively maintained in the single progenitor protein but for innovation–amplification–divergence (IAD) (7, 8), one ancestral function is under selection, whereas others are neutral. After gene duplication, loci may evolve neutrally such that ancestral functions are partitioned between daughters (DDC) or selection may improve both ancestral functions (EAC) or only a previously neutral one (IAD). The IAD model of protein functional change is congruent with related ideas arising from protein engineering studies (9–11) in which neutral, promiscuous protein functions can arise under purifying selection for maintenance of primary protein function and then be selected for (10, 12–15). Although a proposed mechanism for functional change has been demonstrated by engineering experiments, the importance of multiple ancestral activities for protein evolution in lineages of naturally occurring enzymes remains unclear.

Discerning the fate of ancestral activities during protein functional shifts is hampered by the fact that ancient proteins are extinct. Ancestral state estimates based on modern-day protein functions (16) can provide insight into ancestral conditions; however, a particularly promising strategy is to resurrect ancestral proteins and directly determine their activities. This paleomolecular approach has uncovered ancestral protein properties and indicated the structural bases of functional evolution in several studies (17–23). To investigate the importance of ancestral protein functions for enzyme functional divergence, we have resurrected and biochemically characterized ancestral enzymes for a group of plant methyltransferases from the salicylic acid/benzoic acid/theobromine (SABATH) gene family that are important for floral fragrance production, pathogen and herbivore defense, and plant development (24–29). The enzymes focused on in this study, salicylic acid carboxyl methyltransferase (SAMT), benzoic/salicylic acid carboxyl methyltransferase (BSMT), and nicotinic acid carboxyl methyltransferase (NAMT), display considerable functional variation even though they are also capable of discrimination among even structurally similar substrates (30). To investigate functional change in the SABATH gene family, we used a combination of paleomolecular biology, evolutionary statistics, and forward and reverse evolutionary genetic techniques. Together, these complementary approaches allow insight into ancestral conditions and provide experimental evolutionary tests to understand protein functional diversification.

Results

Modern-Day SABATH Enzymes Have High Activity with Few Substrates and Lesser Activity with Many. As part of our on-going studies of carboxyl methyltransferases, we determined enzyme activities of SAMT or BSMT from *Hoya carnosa* (Apocynaceae), *Nicotiana suaveolens*, *Datura wrightii*, and *Cestrum nocturnum* (all Solanaceae) against 18 substrates. All four of these species emit the products of these enzymes, methyl salicylate and/or methyl benzoate, from their flowers (31–33). Fig. 1 shows a comparison of the relative activities of these four enzymes with nine others from Apocynaceae and Solanaceae we have previously characterized (30, 33) (Table S1). It is clear that the substrate profiles of nearly all SAMTs are highly similar to each other and show highest activity with salicylic acid (SA), which they prefer over benzoic acid (BA) by 3- to 10-fold or greater (Fig. 1). Otherwise only 2,3-dihydroxyBA and 2,5-dihydroxyBA are methylated at appreciable levels (>15%). SAMT from *Stephanotis floribunda*, a close relative of *H. carnosa*, differs in that it has high relative activity with a large number of different substrates, including SA (33). The SAMT substrate profiles are markedly different from those shown for BSMTs from Solanaceae, which have highest

Author contributions: F.H., B.P., and T.J.B. designed research; R.H., F.H., D.R., M.H., K.W., J.F., N.S., and T.J.B. performed research; F.H. and T.J.B. analyzed data; and B.P. and T.J.B. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]R.H. and F.H. contributed equally to this work.

[2]To whom correspondence should be addressed. E-mail: todd.barkman@wmich.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1019605109/-/DCSupplemental.

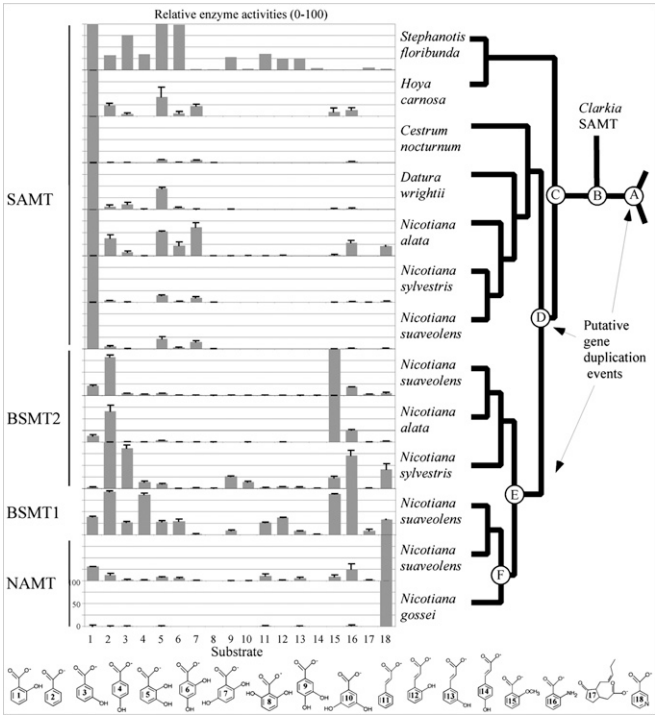

Fig. 1. Relative enzyme activities of 13 SAMT, BSMT, and NAMT with 18 substrates arranged by evolutionary relationships. *Hoya* and *Stephanotis* are from Apocynaceae, whereas all other species are members of Solanaceae. Activity with the favored substrate was used to normalize all other activities for each enzyme, which range from 0 to 100. Most SAMT show high relative preference for salicylic acid (SA) relative to benzoic acid (BA) and all other substrates. Most BSMT show higher relative preference for BA than SA, although high activities are shown with *o*-anisic acid and anthranilic acid as well. NAMT exhibits highest preference for nicotinic acid (NA). The gene duplication event that gave rise to the entire lineage of enzymes shown is hypothesized to have occurred at node A (31) (Fig. S1). A gene duplication event at node D occurred early in Solanaceae evolutionary history and resulted in the SAMT and BSMT lineages of that family. A later duplication event is thought to have occurred at node E only within *Nicotiana* (30). Data for *S. floribunda* are from ref. 33. Numbered structures are as follow: 1, salicylic acid; 2, benzoic acid (BA); 3, 3-hydroxyBA; 4, 4-hydroxyBA; 5, 2,3-dihydroxyBA; 6, 2,4-dihydroxyBA; 7, 2,5-dihydroxyBA; 8, 2,6-dihydroxyBA; 9, 3,4-dihydroxyBA; 10, 3,5-dihydroxyBA; 11, cinnamic acid; 12, *o*-coumaric acid; 13, *m*-coumaric acid; 14, *p*-coumaric acid; 15, *o*-anisic acid; 16, anthranilic acid; 17, jasmonic acid; and 18, nicotinic acid.

activities with BA, 3- or 4-hydroxyBA, *o*-anisic acid, or anthranilic acid (Fig. 1). Nearly all BSMTs show much less activity with SA, preferring BA by 2- to 10-fold or greater. Remarkably, the Solanaceae SAMT and BSMT differences have largely been maintained since the lineages diverged from each other at least 50 Mya when the family originated (34). The most enzymatically divergent enzyme shown is NAMT, known only from *N. gossei* and *N. suaveolens*, which prefers methylate nicotinic acid (NA) above all other substrates (Fig. 1 and Table S1) (30). Whereas these enzymes show differing levels of activity with these substrates, assays of their ancestors are necessary to understand their functional evolution.

Assays of Ancestral Enzymes Indicate Three Evolutionary Shifts in Substrate Preference After Gene Duplication. We investigated the extent of ancestral enzyme divergence after gene duplication with respect to the three substrates used by modern-day enzymes (Fig. 1) for floral scent production and pathogen/herbivore defense: SA, BA, and NA. For each of three lineages, we used three lines of evidence to understand the role of ancestral functional variation for protein divergence. First, in each case, ancestral enzyme activity with a nonpreferred substrate was shown to have become the primary activity in a descendant enzyme after gene duplication. Second, statistical tests indicate historical episodes of positive selection that were concomitant with changes in substrate preference. Third, mutagenesis experiments verified the role of putatively adaptive sites for the functional changes.

Evolution of Functional Change Between Nodes A and B. First, the lineage of SAMT, BSMT, and NAMT enzymes shown in Fig. 1 is thought to have arisen from the gene duplication event at node A, which separates them from the other functionally diverse members of the family including IAMT, GAMT, FAMT, JMT, and caffeine synthase (Fig. 2 and Fig. S1) (31, 35). Thus, we investigated evolution of substrate preference between node A and B. To determine ancestral activities at node A, we experimentally resurrected six alleles of an ancestral enzyme, ancMT-A, that were estimated from contemporary protein sequences (36), including all functionally characterized SABATH enzymes (Fig. S1). Despite the high levels of sequence divergence within this family of proteins, confidence was high for the reconstructed ancestral amino acid sequence of ancMT-A (mean posterior probability (PP = 0.92; Figs. S2 and S3). Fig. 2 shows that ancMT-A activity with BA was maximal, whereas activity with SA was 4- to 10-fold lower (Fig. 2*A*). Resurrection of three alleles of ancestral enzyme ancMT-0, for an even more ancient node in the SABATH gene tree (node 0; Figs. S1–S3) also indicates that BA was ancestrally preferred over SA. Assuming this ancestral condition at node A, it is apparent from experimental analysis of six alleles of the resurrected enzyme at node B (ancMT-B) (Figs. S1–S3) that it evolved to have highest activity with the formerly nonpreferred substrate, SA (Fig. 2*B*).

Second, one of the predictions of most models of protein functional divergence after gene duplication is that positive selection will promote change. As shown between nodes A and B (Fig. 2), a branch-sites statistical analysis (37, 38) indicates that although there likely were >50 changes along this branch, positive selection was associated with substitution at a single active site residue and concomitant with the change in preference toward SA along this lineage [$d_N/d_S = 35$; $P < 0.05$; position 201: His to Met (PP = 0.998)]. This branch has been previously reported to have experienced positive selection (31) and even with the addition of several newly characterized sequences, the statistical signature of adaptive evolution remains.

Third, to experimentally verify the predictions of the statistical analyses of selection between nodes A and B, we forward mutated ancMT-A_1 and ancMT-0_3. In each case, replacement of His201 by Met recapitulated the inferred evolutionary change along branch A–B: both ancestral enzymes changed from preferring BA >4-fold over SA (Fig. 2*A*) to preferring SA over BA by at least 1.6-fold (Fig. 2*B*).

Although it could be argued that the change in relative enzyme preference was due to a decrease in activity with BA, rather than an increase with SA, this does not appear to be the case. An investigation of enzyme kinetics of a site-directed reverse mutant M201H *Hoya* SAMT enzyme revealed increased relative activity with BA compared with wild type (Table S2). Replacement of Met201 by His in *H. carnosa* SAMT affected the catalytic efficiency of this enzyme for SA because k_{cat}/K_M decreased in the mutant by more than fourfold (Table S3). In contrast, k_{cat}/K_M for BA is only twofold lower in M201H compared with wild type. Whereas wild type has a k_{cat}/K_M for SA that is nearly eight times higher than that for BA, M201H is only approximately three times higher (Table S3). Thus, it appears that one change along lineage A–B promoted increased activity with the ancestrally nonpreferred substrate, SA, and this change was due mostly to an increased k_{cat}.

Evolution of Functional Change Between Nodes D and E. First, the major enzymatic divergence between Solanaceae SAMT and BSMT appears to have coincided with gene duplication at node

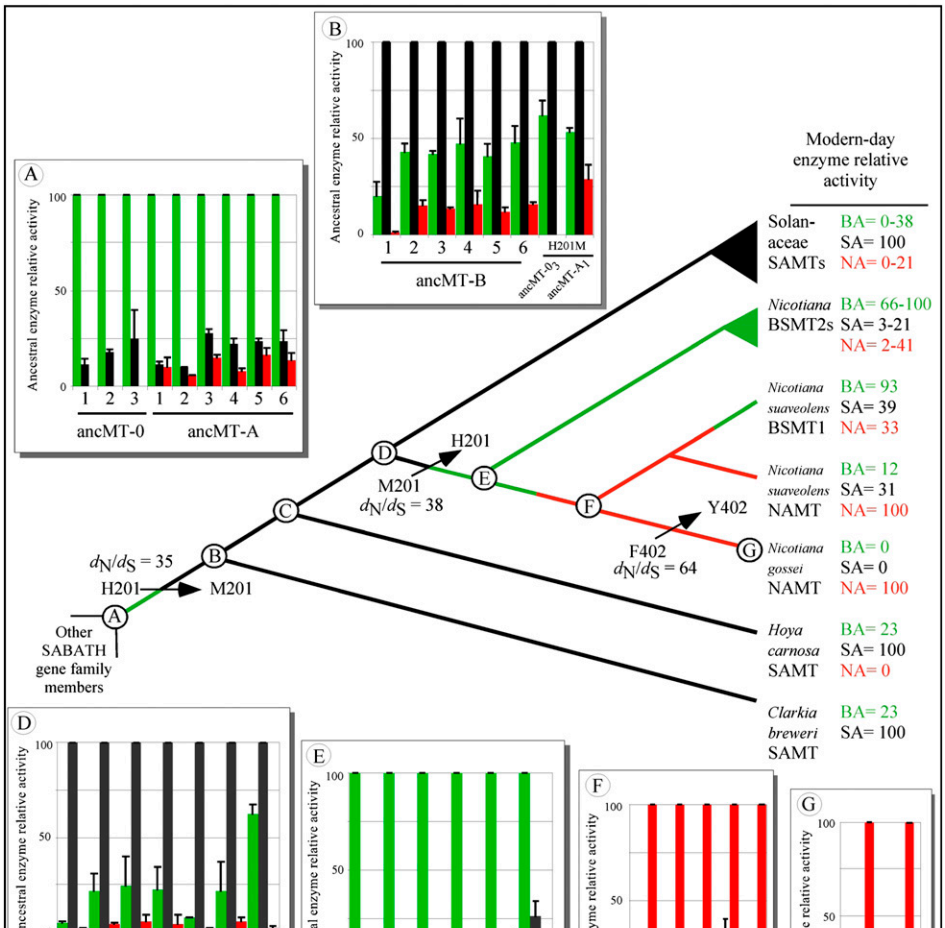

Fig. 2. Experimental results used to investigate the fate of ancestral nonpreferred enzyme activities after functional shifts. The tree shown is simplified from that shown in Fig. 1 with the same node labels. (A–G) Nodes for which relative enzyme preference for BA (green), SA (black), and NA (red) for either resurrected enzymes and their forward mutants or reverse mutants of modern-day enzymes (bold) was determined. Mean and SD are shown on the basis of at least two replicate assays. Color shown for lineages indicates the highest ancestral relative activity with a particular substrate. The ancestral enzyme at node A appears to have preferred BA but later evolved to prefer SA as shown for node B. This change in preference was concomitant with positive selection for the replacement of His by Met at position 201 ($P < 0.05$). The ancestral enzyme at node D had a 20-fold higher relative activity for SA over BA, but after gene duplication, the descendant enzyme at node E evolved a >20-fold higher relative activity for BA over SA. This evolutionary reversal appears to have been concomitant with positive selection for the reverse replacement of Met by His at position 201 ($P < 0.05$). The ancestral enzyme, ancMT-F, evolved high relative preference for NA from an ancestor that preferred BA (ancMT-E). Subsequently, preference for NA increased even more in NAMT, probably by the replacement of Phe420 by Tyr, although it is not clear whether positive selection was concomitant with this change ($P > 0.05$).

D (Fig. 1) (30, 31). Thus, protein functional evolution was investigated between nodes D and E (Fig. 2). At node D, SA appears to have been the preferred substrate, on the basis of assays of six alleles of the resurrected ancestral enzyme, ancMT-D (Figs. S1–S3), whereas in comparison, activity with BA was fourfold less (Fig. 2*D*). Subsequent to gene duplication at node D, activity with BA evolved to become maximal, whereas activity with SA diminished as indicated by assays of five alleles of the resurrected enzyme ancMT-E (Fig. 2*E* and Figs. S1–S3). This evolutionary change of relative enzyme preference is the reverse of that observed between nodes A and B described above (Fig. 2 *A* and *B*).

Second, a statistical signature of positive selection (d_N/d_S = 38; $P < 0.05$) is associated with the increased relative activity with BA that evolved between nodes D and E after gene duplication. Of the >50 likely changes along this branch, the only predicted adaptive site was position 201, which appears to have reversed this active site residue from Met back to the ancestral His (PP = 0.99) (compare with amino acid replacement between nodes A and B).

Third, to test for functional relevance of the positively selected site, we recapitulated the evolutionary change of M201 to His in ancMT-D_1. This mutant enzyme changed from showing a 10-fold higher activity with SA relative to BA to showing 3-fold higher activity with previously nonpreferred substrate, BA (Fig. 2*E*). The importance of this site for activity with BA is strengthened by mutagenesis of His201 back to Met in *N. suaveolens* BSMT1. The reverse evolutionary mutation resulted in a switch from

preference for BA over SA in wild-type BSMT1 to a 1.6-fold preference for SA over BA (Fig. 2*D* and Table S2).

Finally, it appears that the functional shift in relative enzyme preference between nodes D and E was due to an increase in activity with the previously nonpreferred substrate BA rather than merely decreasing activity with SA, on the basis of kinetic measurements of both forward and reverse mutated enzymes. For ancMT-D, it appears that positive selection promoted increased substrate preference for BA because K_M of the mutant M201H for BA decreased nearly 4-fold and k_{cat} increased 5-fold, whereas for SA, K_M and k_{cat} decreased 2-fold (Table S3). As a result, k_{cat}/K_M for BA increased 20-fold, whereas it did not change for SA. It should be noted that k_{cat}/K_M is still higher for SA. For the *N. suaveolens* BSMT1 reverse mutant enzyme H201M, k_{cat}/K_M for BA decreased 14-fold compared with wild type, whereas k_{cat}/K_M for SA was essentially the same as wild type (Table S3). The H201M mutant enzyme had a >10-fold higher K_M with BA compared with wild type, even though k_{cat} was largely unchanged. Mutant K_M with SA was 6-fold higher than wild type but k_{cat} actually increased toward SA (Table S3). These data indicate that changes along lineage D–E resulted in an increased affinity toward the ancestrally nonpreferred substrate, BA, apparently due to selection for decreased K_M and increased k_{cat}.

Evolution of Functional Change Between Nodes E and F and NAMT. First, node E is postulated to have undergone gene duplication within *Nicotiana* (Fig. S2) (30). Thus, enzyme activity evolution was investigated between nodes E and F. Assays of five resurrected alleles of ancMT-F (Figs. S1–S3) indicate that activity with NA was highest for this ancestral enzyme, whereas activity with the ancestrally preferred substrate, BA, was 7- to 30-fold less (Fig. 2*F*). This high level of activity with NA evolved from an ancestral condition at node E in which BA was preferred and activity with NA was only minimal (Fig. 2*E*). From the condition at node F, it appears that NAMT evolved to have negligible activity with BA and SA relative to NA (Fig. 2*G*). Surprisingly, given the ancestral condition at node F, BSMT1 evolved to have higher activity with BA relative to NA, an evolutionary reversal from that observed between nodes E and F (Fig. 2).

Second, although no positive selection was detected during the divergence of NAMT from ancestor F when analyzing all SABATH family members together (Fig. S1) ($P > 0.05$), an analysis that involved only SAMT lineage members, including a dense sampling from Solanaceae (Fig. S4), did result in an optimal estimate of $d_N/d_S = 64$ (Fig. 2), although this was not statistically distinguishable from the null ($P > 0.05$). Sequence divergence is very low along this branch and thus the test may lack sufficient power to detect statistical significance in this case (39). Nonetheless, in the second statistical analysis assuming the tree in Fig. S4, of approximately eight sites that have changed along this branch, two sites (141 and 402) in the NAMT lineage were predicted to be under positive selection having switched from previously being under purifying selection. No significant signature of positive selection was detected between nodes E and F or from node F to BSMT1.

Third, to test the functional importance of one of the putatively selected sites, forward mutagenesis was performed. Introduction of the F402Y active site mutation into ancMT-F resulted in a further enzyme preference for NA (Fig. 2*G*) compared with node F (Fig. 2*F*). However, it is not yet known whether evolution proceeded by increasing enzyme relative preference for NA, decreasing it for BA and SA, or a combination of both. Experiments to ascertain the role of particular amino acid replacements contributing to the enzymatic divergences between nodes E and F and node F and BSMT1 are currently underway.

Discussion

In this family of enzymes, it is clear that ancestral functional variation was refined to generate activity shifts in multiple lineages. These evolutionary patterns of enzyme activity change were uncovered by the resurrection and functional characterization of ancestral proteins in combination with statistical analyses that implicated specific amino acid residues. Subsequent experimental manipulation of these sites demonstrated their respective contributions to functional change. Because SABATH family enzyme functional evolution appears to depend upon the existence of multiple ancestral activities, we attempted to reconcile the theoretical predictions of DDC, EAC, and IAD (2–4, 40, 41) with the results we have reported.

Conflict Appears to Exist Among Some Ancestral Activities. It appears that in this family of enzymes, neither ancestral nor modern-day enzymes can be equally effective with the physiologically relevant substrates, BA and SA (Figs. 1 and 2). The changes, introduced to the active site residue 201 that improved preference for one substrate, came at the expense of reduced relative preference for the other (Fig. 2, lineages A–B and D–E, and Tables S2 and S3). In fact, relative activities with BA and SA in extant enzymes are negatively related (covariance = −0.331; phylogenetic covariance estimate = −0.168). Assuming the efficient production of MeBA and MeSA is advantageous, adaptive conflict appears to exist between specialization for SA or BA, making the enzyme functional shifts of lineage A–B and D–E potentially best described by the EAC model of protein evolution. After gene duplication, the ancestral enzymes at nodes B and E evolved improved activity with SA and BA, respectively, consistent with EAC. The other daughter enzymes descending from the gene duplication events may also have evolved improvement with an ancestral function because the SAMT enzymes, descended from node D, appear to have evolved increased relative activity for SA to some extent; however, it is currently unknown what activities the other descendant of node A evolved to specialize upon due to uncertainty in the gene tree (Fig. S1). Also consistent with EAC is the finding that a single positively selected codon (amino acid position 201) largely accounts for the change in preference between nodes A and B and D and E; still, it is not clear to what extent positive selection may have acted on the other branches descending from duplication events at nodes A and D because statistical significance was observed. The finding that a single residue governs enzyme substrate preference is particularly important because it identifies an obvious mechanism by which adaptive conflict could arise and has been implicated for a laboratory study of bacterial protein evolution (42). Because other studies have reported single amino acid switches that interconvert modern-day phenylpropanoid enzyme substrate preferences (43) and product outcome of diterpene- and carotenoid-producing enzymes (44, 45), adaptive conflict may be pervasive, especially for proteins involved in specialized metabolite production. Due to the uncertainty associated with inferring positive selection and determining whether both descendant enzymes improved upon ancestral activities after the gene duplication events at nodes A and D, it remains possible that IAD also describes these divergences, despite the apparent conflicting ancestral activities.

Latent, Low Level Activities Provide Raw Material for Evolutionary Change. Our results show that low ancestral activity with nonpreferred substrates may be latent for long periods and persist through multiple gene duplication and speciation events, but ultimately these low activities can evolve to become high. Specifically, NA activity was minor in the ancestor of the entire lineage at node A (*ca.* 100 Mya) and through nodes B–E, yet at node F (<7.5 Mya) (46), a shift occurred such that the ancestral enzyme evolved to prefer NA above all others. So, whereas the ancestral enzymes shown in Fig. 2 vary between having preference for SA or BA, activity with NA does not change appreciably until node F and thus may have been historically neutral with

respect to enzyme preference for the other substrates. If so, the functional shifts between node E and F and NAMT in *N. gossei* likely represent instances of evolution by IAD. In this case, the previously minor activity with NA would serve as an innovation. Amplification appears to have occurred in *N. gossei* and *N. suaveolens* after duplications at nodes D and E and because they are allopolyploids providing them with no less than four SAMT/BSMT-like genes that could have potentially provided considerable NA methylation ability in combination (30, 47). Divergence appears to have occurred over multiple time frames after duplication at node E, whereby node F evolved highest activity with NA, and this activity was further evolved in the lineage leading to NAMT at node G (Fig. 2). As predicted by IAD, whereas activity with NA was increased in one daughter after duplication at node E, relative activity with BA does not seem to have increased after duplication in descendants of the other daughter of ancMT-E, BSMT2. Although we have identified sites that contributed to increased activity with NA, positive selection was not statistically significant, making DDC potentially explanatory for the results as well.

Limited Constraint on Evolutionary Reversals of Enzyme Activity. In addition to demonstrating the importance of ancestral promiscuous activities for protein functional evolution, we have documented a clear case of an evolutionary reversal of an active site residue that resulted in a reversal of associated enzyme activity. Initially, after gene duplication resulted in formation of the SAMT lineage at node A, the active site His201 was replaced by Met (Fig. 2 between nodes A and B), resulting in the evolution of enzyme preference for SA (Fig. 2*B*). Remarkably, after gene duplication at node D, this homologous site in one of the daughter lineages then underwent an evolutionary reversal from Met back to the ancestral His (Fig. 2, between nodes D and E), which resulted in a reversal of enzymatic properties including loss of preference for SA and a return to higher preference for BA (Fig. 2*E*). Furthermore, whereas the causative amino acid replacements have not yet been identified, it is also clear that a second reversal occurred: between nodes E and F, high activity with NA evolved from an ancestor that preferred BA. Subsequently, the ancestral NA-preferring enzyme at node F gave rise to BSMT1, which prefers BA. Previous studies have clearly shown that restrictive epistatic interactions among amino acid sites can constrain evolutionary reversals (19, 48). Because the original change from His201 to Met likely predated the divergence of rosids and asterids (*ca*. 100–125 Mya) (30) and the reversal occurred after the origin of Solanaceae (*ca.* 50 Mya), it is surprising that mutations did not accumulate during the *ca.* 50–75 million-year interval that would constrain this active site reversal and associated functional change. If generalizable, these results indicate that epistasis may not constrain evolutionary changes of enzymes involved in specialized metabolism to the extent it does with other proteins (19, 48).

Our mutagenesis studies of this family of methyltransferases validate the impact of historical changes on shaping enzyme preference. Specifically, selection appears to have promoted directional changes in catalytic efficiency (k_{cat}/K_M) for the two competing substrates, salicylic and benzoic acid, even though the adaptive substitutions may affect K_M and k_{cat} separately (Table S3). It seems probable that the ability to discriminate between SA and BA is valuable, given that recent studies have shown a central role of MeSA in pathogen (29) and herbivore defense (35); therefore, its efficient production by SAMT is likely advantageous. Likewise, selection for efficient MeBA and MeNA production in *Nicotiana* is likely related to effective pollinator attraction as both BSMT1 and NAMT are expressed primarily in petal tissue from which the corresponding volatiles are emitted (30). It should be noted that measured substrate preferences of these enzymes is not necessarily indicative of the role in planta, because the SAMT orthologous enzymes in *Petunia* and *Stephanotis*, which prefer SA in vitro, are used to methylate BA in petals (33, 49). In this case, substrate pool is an important determining factor for phenotype rather than enzyme kinetic properties alone (33, 49). Although it may be adaptive for an enzyme to distinguish among competing substrates as shown for SAMT, BSMT, and NAMT, it seems that variable levels of relative activity with many substrates appear nearly universal in ancestral and extant enzymes and have likely served to facilitate functional diversification throughout the history of this family. The importance of ancestral functional variation for protein evolution may be much broader than previously appreciated. Beyond its obvious role in the evolution of novel activity via protein engineering (9, 11, 50), ligand-binding promiscuity has been reported in ancestral vertebrate steroid receptors, which served as the basis for subsequent natural evolution of novel receptor–ligand interactions (20, 51). Likewise, *jingwei*, a chimeric alcohol dehydrogenase from *Drosophila*, evolved preference for substrates that were nonpreferred in the progenitor enzyme (52, 53).

Materials and Methods

Heterologous Expression and Purification of Enzymes. The basic protocols used for gene cloning and protein overexpression were performed as previously described (30, 33). Briefly, genes were cloned into expression vectors and overexpression of His_6 protein was achieved in BL-21 cells. Purification of the His_6-tagged protein was purified by Ni-NTA affinity chromatatography (Qiagen) according to the manufacturer's instructions. To determine protein concentration, a standard Bradford assay was used. For mutagenesis, the QuikChange Site-Directed Mutagenesis kit (Stratagene) was used according to the manufacturer's instructions.

Enzyme Assays. The purified enzymes were tested for activity with the 18 substrates shown in Fig. 1. Radiochemical assays were performed as previously described (30). The highest enzyme activity reached with a specific substrate was set to 100 and relative activities with remaining substrates were calculated. Each assay was run in duplicate and mean, plus SD, was calculated for Fig. 1. For the ancestral enzyme functional assays, we used the same radiochemical assays described above except that we did not add HCL before extraction of reaction products. All assays shown in Fig. 2 were performed on total protein because relative activity levels were similar between total and purified protein for the two enzymes we compared (ancMT-D and ancMT-D M201H).

Estimation of Michaelis–Menten Parameters. For kinetic measurements, enzyme assays were performed by varying SA and BA concentrations, whereas SAM and enzyme concentrations were held at saturating levels. All kinetic studies were performed in two independent experiments with incubation times chosen so that reaction velocity was linear. Lineweaver-Burk plots were performed to determine the K_M and k_{cat} values as in our previous studies (33).

Statistical and Molecular Evolutionary Analyses. DNA sequences from all enzymatically characterized SABATH gene family members were obtained from GenBank. Maximum likelihood phylogenetic analyses were performed with PAUP* (54) as previously described (30) assuming the HKY+I+G model of nucleotide substitution as chosen by Modeltest (55) using 10 random addition sequences and TBR swapping. Bootstrapping was performed using 100 replicates. PAML ver. 4.2 (36) was used to test the hypothesis of positive selection in the SABATH gene family using the branch-sites test because it is expected that positive selection should act only on a subset of sites and branches of a gene tree as functional divergence occurs. This test has recently been shown to be robust under a wide range of conditions and the most powerful test available (38, 56). Analyses were performed multiple times using different starting values of ω and assumed the HKY model of nucleotide substitution. Codeml was used to estimate ancestral enzyme sequences for the SABATH gene family with the HKY model with κ estimated. The γ and invariant parameters cannot be implemented in Codeml under the branch-sites model. Additional details about ancestral sequence estimation and how alternative sites were chosen to assess uncertainty are provided in *SI Materials and Methods*. The optimal sequences were subsequently synthesized by Genescript with codons chosen for optimal protein expression in *Escherichia coli*.

ACKNOWLEDGMENTS. We thank Doug Coulter, Pam Hoppe, Talline Martins, Julie Ryan, John Spitsbergen, and James Kiddle for helpful discussions and assistance; and the Chemistry Department at Western Michigan University for facilitating our GC/MS analyses. This study was supported by National Science Foundation Grants DEB 0344496 and MCB-1120624 and a FRACAA award from Western Michigan University (to T.J.B.) and by Deutsche Forschungsgemeinschaft Pi 153/22 (to B.P.).

1. Ohno S (1970) *Evolution by Gene Duplication* (Springer, New York).
2. Force A, et al. (1999) Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* 151:1531–1545.
3. Hughes AL (1994) The evolution of functionally novel proteins after gene duplication. *Proc Biol Sci* 256:119–124.
4. Des Marais DL, Rausher MD (2008) Escape from adaptive conflict after duplication in an anthocyanin pathway gene. *Nature* 454:762–765.
5. Hittinger CT, Carroll SB (2007) Gene duplication and the adaptive evolution of a classic genetic switch. *Nature* 449:677–681.
6. Deng C, Cheng C-HC, Ye H, He X, Chen L (2010) Evolution of an antifreeze protein by neofunctionalization under escape from adaptive conflict. *Proc Natl Acad Sci USA* 107:21593–21598.
7. Bergthorsson U, Andersson DI, Roth JR (2007) Ohno's dilemma: Evolution of new genes under continuous selection. *Proc Natl Acad Sci USA* 104:17004–17009.
8. Soo VWC, Hanson-Manful P, Patrick WM (2011) Artificial gene amplification reveals an abundance of promiscuous resistance determinants in *Escherichia coli*. *Proc Natl Acad Sci USA* 108:1484–1489.
9. Aharoni A, et al. (2005) The 'evolvability' of promiscuous protein functions. *Nat Genet* 37:73–76.
10. Khersonsky O, Roodveldt C, Tawfik DS (2006) Enzyme promiscuity: Evolutionary and mechanistic aspects. *Curr Opin Chem Biol* 10:498–508.
11. Bloom JD, Arnold FH (2009) In the light of directed evolution: Pathways of adaptive protein evolution. *Proc Natl Acad Sci USA* 106(Suppl 1):9995–10000.
12. Tracewell CA, Arnold FH (2009) Directed enzyme evolution: Climbing fitness peaks one amino acid at a time. *Curr Opin Chem Biol* 13:3–9.
13. Bershtein S, Tawfik DS (2008) Ohno's model revisited: Measuring the frequency of potentially adaptive mutations under various mutational drifts. *Mol Biol Evol* 25: 2311–2318.
14. Bloom JD, Romero PA, Lu Z, Arnold FH (2007) Neutral genetic drift can alter promiscuous protein functions, potentially aiding functional evolution. *Biol Direct* 2:17.
15. Peisajovich SG, Tawfik DS (2007) Protein engineers turned evolutionists. *Nat Methods* 4:991–994.
16. Pagel M (1999) Inferring the historical patterns of biological evolution. *Nature* 401: 877–884.
17. Yokoyama S, Tada T, Zhang H, Britt L (2008) Elucidation of phenotypic adaptations: Molecular analyses of dim-light vision proteins in vertebrates. *Proc Natl Acad Sci USA* 105:13480–13485.
18. Thomson JM, et al. (2005) Resurrecting ancestral alcohol dehydrogenases from yeast. *Nat Genet* 37:630–635.
19. Harms MJ, Thornton JW (2010) Analyzing protein structure and function using ancestral gene reconstruction. *Curr Opin Struct Biol* 20:360–366.
20. Bridgham JT, Carroll SM, Thornton JW (2006) Evolution of hormone-receptor complexity by molecular exploitation. *Science* 312:97–101.
21. Zhang J (2006) Parallel adaptive origins of digestive RNases in Asian and African leaf monkeys. *Nat Genet* 38:819–823.
22. Ortlund EA, Bridgham JT, Redinbo MR, Thornton JW (2007) Crystal structure of an ancient protein: Evolution by conformational epistasis. *Science* 317:1544–1548.
23. Field SF, Matz MV (2010) Retracing evolution of red fluorescence in GFP-like proteins from Faviina corals. *Mol Biol Evol* 27:225–233.
24. D'Auria JC, Chen F, Pichersky E (2003) *Recent Advances in Phytochemistry*, ed Romeo JT (Elsevier Science, Oxford), Vol 37, pp 253–283.
25. Murata J, Roepke J, Gordon H, De Luca V (2008) The leaf epidermome of *Catharanthus roseus* reveals its biochemical specialization. *Plant Cell* 20:524–542.
26. 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* 367:9–16.
27. Seo HS, et al. (2001) Jasmonic acid carboxyl methyltransferase: A key enzyme for jasmonate-regulated plant responses. *Proc Natl Acad Sci USA* 98:4788–4793.
28. Varbanova M, et al. (2007) Methylation of gibberellins by *Arabidopsis* GAMT1 and GAMT2. *Plant Cell* 19:32–45.
29. Park S-W, Kaimoyo E, Kumar D, Mosher S, Klessig DF (2007) Methyl salicylate is a critical mobile signal for plant systemic acquired resistance. *Science* 318:113–116.
30. Hippauf F, et al. (2010) Enzymatic, expression and structural divergences among carboxyl O-methyltransferases after gene duplication and speciation in Nicotiana. *Plant Mol Biol* 72:311–330.
31. 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* 24:1320–1329.
32. Altenburger R, Matile P (1988) Circadian rhythmicity of fragrance emission in flowers of *Hoya carnosa* R Br. *Planta* 174:248–252.
33. Pott MB, et al. (2004) Biochemical and structural characterization of benzenoid carboxyl methyltransferases involved in floral scent production in *Stephanotis floribunda* and *Nicotiana suaveolens*. *Plant Physiol* 135:1946–1955.
34. Paape T, et al. (2008) A 15-Myr-old genetic bottleneck. *Mol Biol Evol* 25:655–663.
35. Ament K, et al. (2010) Methyl salicylate production in tomato affects biotic interactions. *Plant J* 62:124–134.
36. Yang ZH (2007) PAML 4: Phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 24:1586–1591.
37. Yang ZH, Nielsen R (2002) Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Mol Biol Evol* 19:908–917.
38. Yang Z, dos Reis M (2011) Statistical properties of the branch-site test of positive selection. *Mol Biol Evol* 28:1217–1228.
39. Nozawa M, Suzuki Y, Nei M (2009) Reliabilities of identifying positive selection by the branch-site and the site-prediction methods. *Proc Natl Acad Sci USA* 106:6700–6705.
40. Zhang J (2003) Evolution by gene duplication: An update. *Trends Ecol Evol* 18: 292–298.
41. Barkman TJ, Zhang J (2009) Evidence for escape from adaptive conflict? *Nature* 462: E1.E2, discussion E2–E3.
42. McLoughlin SY, Copley SD (2008) A compromise required by gene sharing enables survival: Implications for evolution of new enzyme activities. *Proc Natl Acad Sci USA* 105:13497–13502.
43. Gang DR, et al. (2002) Characterization of phenylpropene *O*-methyltransferases from sweet basil: Facile change of substrate specificity and convergent evolution within a plant *O*-methyltransferase family. *Plant Cell* 14:505–519.
44. Cunningham FX, Jr., Gantt E (2001) One ring or two? Determination of ring number in carotenoids by lycopene epsilon-cyclases. *Proc Natl Acad Sci USA* 98:2905–2910.
45. Xu M, Wilderman PR, Peters RJ (2007) Following evolution's lead to a single residue switch for diterpene synthase product outcome. *Proc Natl Acad Sci USA* 104: 7397–7401.
46. Wu F, et al. (2010) COSII genetic maps of two diploid *Nicotiana* species provide a detailed picture of synteny with tomato and insights into chromosome evolution in tetraploid *N. tabacum*. *Theor Appl Genet* 120:809–827.
47. Chase MW, et al. (2003) Molecular systematics, GISH and the origin of hybrid taxa in *Nicotiana* (Solanaceae). *Ann Bot (Lond)* 92:107–127.
48. Bridgham JT, Ortlund EA, Thornton JW (2009) An epistatic ratchet constrains the direction of glucocorticoid receptor evolution. *Nature* 461:515–519.
49. Negre F, et al. (2003) Regulation of methylbenzoate emission after pollination in snapdragon and petunia flowers. *Plant Cell* 15:2992–3006.
50. Arnold FH (2009) *Cold Spring Harbor Symposia on Quantitative Biology* (Cold Spring Harbor Lab Press, Cold Spring Harbor, NY), Vol LXXIV, pp 41–46.
51. Carroll SM, Bridgham JT, Thornton JW (2008) Evolution of hormone signaling in elasmobranchs by exploitation of promiscuous receptors. *Mol Biol Evol* 25:2643–2652.
52. Zhang J, Dean AM, Brunet F, Long M (2004) Evolving protein functional diversity in new genes of *Drosophila*. *Proc Natl Acad Sci USA* 101:16246–16250.
53. Zhang J, Yang H, Long M, Li L, Dean AM (2010) Evolution of enzymatic activities of testis-specific short-chain dehydrogenase/reductase in *Drosophila*. *J Mol Evol* 71: 241–249.
54. Swofford DL (2003) *PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods) Version 4* (Sinauer, Sunderland, MA).
55. Posada D, Crandall KA (1998) MODELTEST: Testing the model of DNA substitution. *Bioinformatics* 14:817–818.
56. Zhang JZ, Nielsen R, Yang ZH (2005) Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol Biol Evol* 22: 2472–2479.

AUTHOR QUERIES

AUTHOR PLEASE ANSWER ALL QUERIES 1

Q: 1_Please contact PNAS_Specialist.djs@sheridan.com if you have questions about the editorial changes, this list of queries, or the figures in your article. Please include your manuscript number in the subject line of all e-mail correspondence; your manuscript number is 201019605. Please (i) review the author affiliation and footnote symbols carefully, (ii) check the order of the author names, and (iii) check the spelling of all author names and affiliations. Please indicate that the author and affiliation lines are correct by adding the comment "OK" next to the author line. Please note that this is your opportunity to correct errors in your article prior to publication. Corrections requested after online publication will be considered and processed as errata.

Q: 2_If your article contains links to Web sites (other than the SI links for your article), please verify that the links are valid and will direct readers to the proper Web page.

Q: 3_Author names may have been edited to match those provided during article submission; please check carefully and note your approval in the margin. (Your article cannot be published until your approval has been received.)

Q: 4_There are two institutes listed in the original transmittal information: Leibniz Institute for Farm Animal Biology for author D.R. and Leibniz Institute for Baltic Sea Research for author J.F. but they do not appear in the author listing. Please check and amend as necessary.

Q: 5_Please verify authors match their affiliations as listed.

Q: 6_PNAS allows up to five key terms that (i) do not repeat terms present in the TITLE OR ABSTRACT (which are searchable online) and (ii) do not include nonstandard abbreviations. You may add 3 terms. Also, please check the order of your key terms and approve or reorder them as necessary.

Q: 7_Please review the information in the author contribution footnote carefully. Please make sure that the information is correct and that the correct author initials are listed. Note that the order of author initials matches the order of the author line per journal style. You may add contributions to the list in the footnote; however, funding should not be an author's only contribution to the work.

Q: 8_You will receive a notification from the PNAS eBill system in 1-2 days. Each corresponding author is required to log in to the system and provide payment information for applicable publication charges (purchase order number or credit card information) upon receipt of the notification. You will have the opportunity to order reprints through the eBill system if desired, as well. Failure to log in and provide the required information may result in publication delays.

Q: 9_Reminder: You have chosen not to pay an additional **$1300** (or **$975** if your institution has a site license) for the PNAS Open Access option.

Q: 10_Please verify that all supporting information (SI) citations are correct. Note, however, that the hyperlinks for SI citations will not work until the article is published online. In addition, SI that is not composed in the main SI PDF (appendices, datasets, movies, and "Other Supporting

AUTHOR QUERIES

Information Files") have not been changed from your originally submitted file and so are not included in this set of proofs. The proofs for any composed portion of your SI are included in this proof as subsequent pages following the last page of the main text. If you did not receive the proofs for your SI, please contact **PNAS_Specialist.djs@sheridan.com**.

Q: 11_The callout for "Table S2-5" in the sentence beginning "The changes, introduced . . ." has been changed to "Tables S2 and S3" because there are only three (not five) SI tables. Please check and amend as necessary. If there should be two additional tables, please supply them.

Q: 12_Please add the expansion of TBR in "TBR swapping" if possible in the sentence beginning "Maximum likelihood phylogenetic . . ."

Q: 13_Please replace FRACAA in the acknowledgments with its expansion.

Q: 14_It has been assumed that the callout for Pott et al, 2004 in the legend to Fig. 1 is ref. 33. Okay?

Supporting Information

Huang et al. 10.1073/pnas.1019605109

SI Materials and Methods

Phylogenetic covariances were obtained from modern-day enzyme relative activities using BayesTraits (1). The normalized enzyme preference data from modern-day salicylic acid carboxyl methyltransferase (SAMT), benzoic/salicylic acid carboxyl methyltransferase (BSMT), and nicotinic acid carboxyl methyltransferase (NAMT) for BA, SA, and NA (Fig. 1) were log transformed before ancestral state estimation. A posterior distribution of input trees was obtained using Bayesphylogenies (2) in which a Markov chain was run for one million generations and sampled every 10,000 generations after the burn-in period. The Markov chain for ancestral state estimation was run according to our previous analyses (3). A strong prior of zero for the lower bound of ancestral state estimates was set because negative proportions do not have biological meaning.

Different methods of ancestral sequence reconstruction have strengths and weaknesses. For this study, we relied on maximum likelihood methods although the sequences obtained from parsimony reconstructions were similar particularly for sites at which high posterior probabilities were obtained. For sites at which confidence is low, all methods we used were uncertain. To decide which sites to mutate in our ancMT ancestral enzymes to assess the robustness of our results to ancestral sequence uncertainty, we compared the sequences obtained from our codon-based estimates for the tree in Fig. S1 to those obtained using the denser sampling of sequences shown in Fig. S4. For ancMT-A and ancMT-0, we evaluated the different amino acid estimates using the same set of sequences but assuming different roots. We provide posterior probabilities for original and mutated sites for all mutations in the following list:

ancMT-02 198 Y (0.981) to F (0.755)
ancMT-03 398 C (0.497) to S (0.712)
ancMT-A2 78 K (0.526) to N (0.987)
ancMT-A3 134 L (0.998) to F (0.898)
ancMT-A4 252 S (0.79) to D (0.951)
ancMT-A5 327 S (0.592) to G (0.872)
ancMT-A6 476 E (0.996) to L (0.902)
ancMT-B2 78 N (0.676) to S (0.835)
ancMT-B3 158 E (0.145) to deletion
ancMT-B4 188 Q (0.498) to E (0.496)
ancMT-B5 389 D (0.622) to N (0.460)
ancMT-B6 471 E (0.665) to D (0.534)
ancMT-D2 71 T (0.482) to I (0.981)
ancMT-D3 158 E (0.139) to deletion
ancMT-D4 211 N (1.0) to deletion
ancMT-D5 315 F (0.682) to L (0.822)
ancMT-D6 454 S (0.853) to A (0.952)
ancMT-E2 148 Y (0.454) to D (0.613)
ancMT-E3 303 K (0.724) to E (0.915)
ancMT-E4 330 F (0.495) to C (0.942)
ancMT-E5 472 A (0.643) to S (0.984)
ancMT-F2 266 K (1.0) to I (0.973)
ancMT-F3 339 L (1.0) to I (0.912)
ancMT-F4 348 N (1.0) to K (0.894)
ancMT-F5 441 I (1.0) to T (0.749)

1. Pagel M, Meade A, Barker D (2004) Bayesian estimation of ancestral character states on phylogenies. *Syst Biol* 53:673–684.
2. Pagel M, Meade A (2004) A phylogenetic mixture model for detecting pattern-heterogeneity in gene sequence or character-state data. *Syst Biol* 53:571–581.
3. Hippauf F, et al. (2010) Enzymatic, expression and structural divergences among carboxyl O-methyltransferases after gene duplication and speciation in Nicotiana. *Plant Mol Biol* 72:311–330.

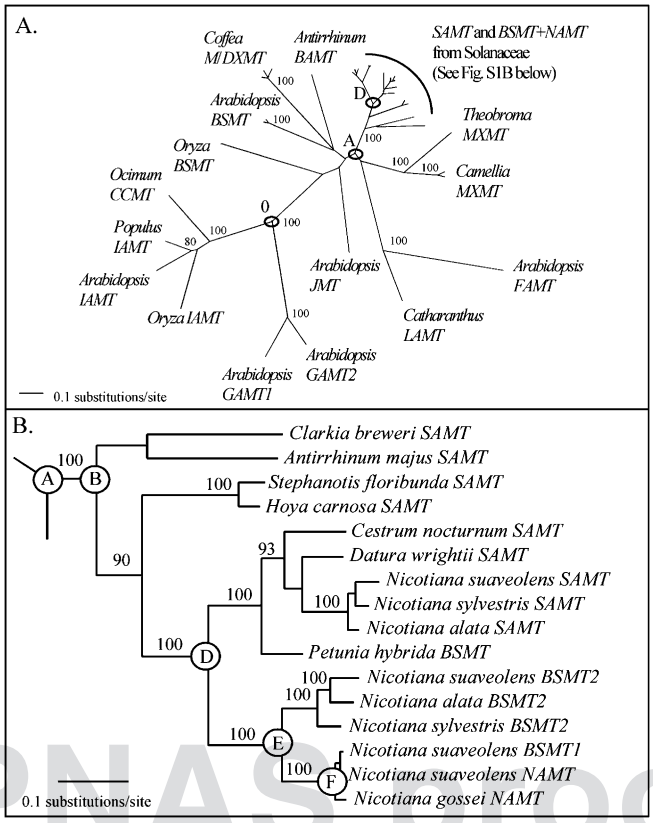

Fig. S1. (*A*) Estimated salicylic acid/benzoic acid/theobromine (SABATH) gene tree showing relationships among functionally characterized members. The *SAMT+BSMT+NAMT* lineage contains all *SAMT*, and *BSMT*, and *NAMT* from Solanaceae. This gene tree was used for statistical analyses of positive selection and ancestral sequence estimation. (*B*) Relationships among *SAMT* and Solanaceae *BSMT+NAMT*. Bootstrap percentages >80 are shown next to corresponding nodes. Nodes A, D, and E represent putative gene duplication events. Ancestral proteins were resurrected and biochemically assayed from nodes 0, A, B, D, E, and F. *Nicotiana suaveolens BSMT1* was isolated from an individual grown in Germany, whereas *NAMT* was obtained from an individual grown in Michigan.

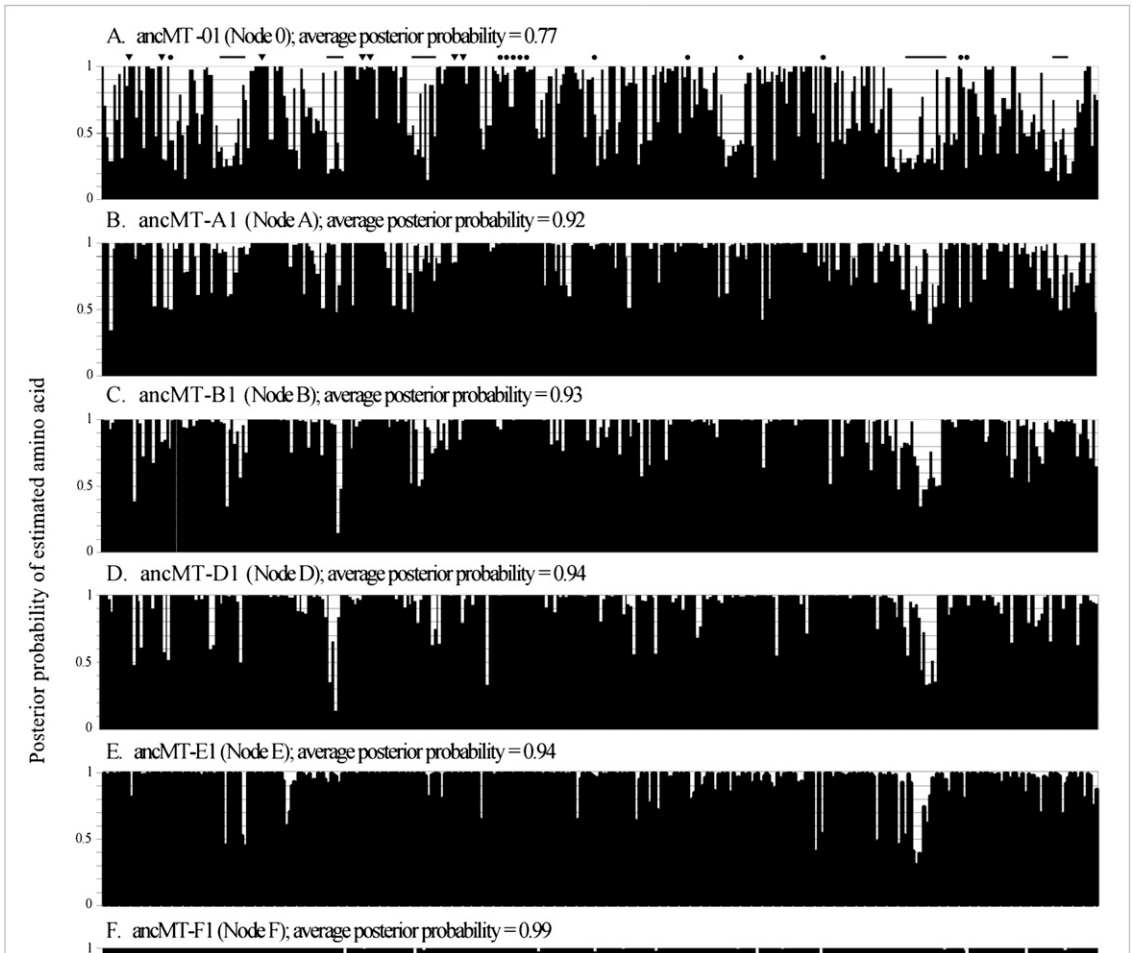

Fig. S2. Site-specific posterior probabilities for ancestral amino acid estimates for the ancestral enzymes shown in Figs. S1 and S3 and functionally characterized in Fig. 2. (*A*) Site-specific posterior probabilities for node 0. (*B*) Site-specific posterior probabilities for node A. (*C*) Site-specific posterior probabilities for node B. (*D*) Site-specific posterior probabilities for node D. (*E*) Site-specific posterior probabilities for node E. (*F*) Site-specific posterior probabilities for node F. All positions known to interact with the methyl donor SAM in the active site have very high posterior probabilities as shown by arrowheads. Active site residues known to interact with the methyl acceptor substrates are shown by filled circles. Nearly all of these have high posterior probabilities (>0.8). Many sites that are on the surface of the protein or form parts of loops (shown by horizontal lines) have lower posterior probabilities. The functional effects of alternative amino acids at sites for which confidence is lower has been investigated for all ancMTs (Fig. 2) (Fig S3 shows specific replacements).

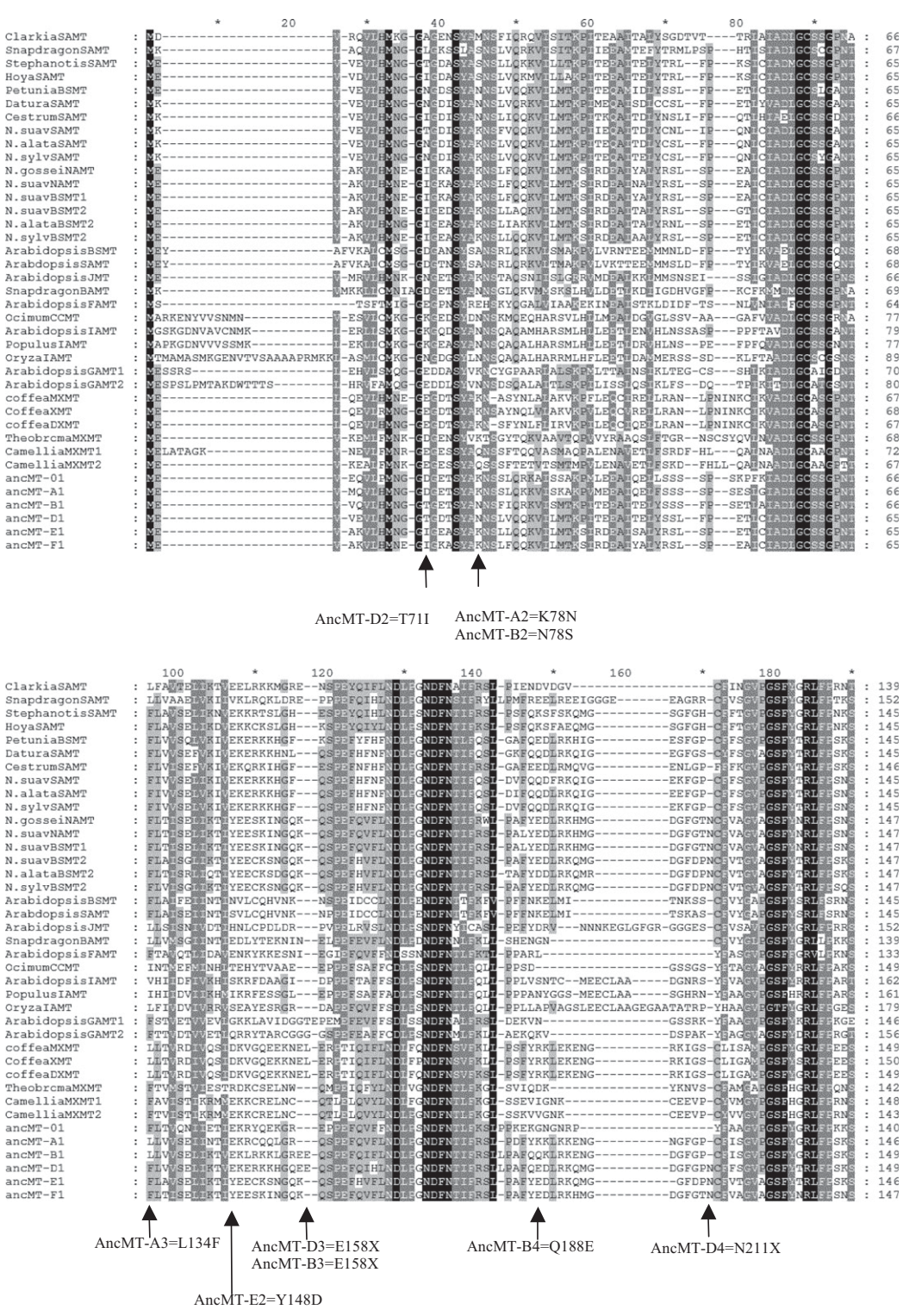

Fig. S3. (Continued)

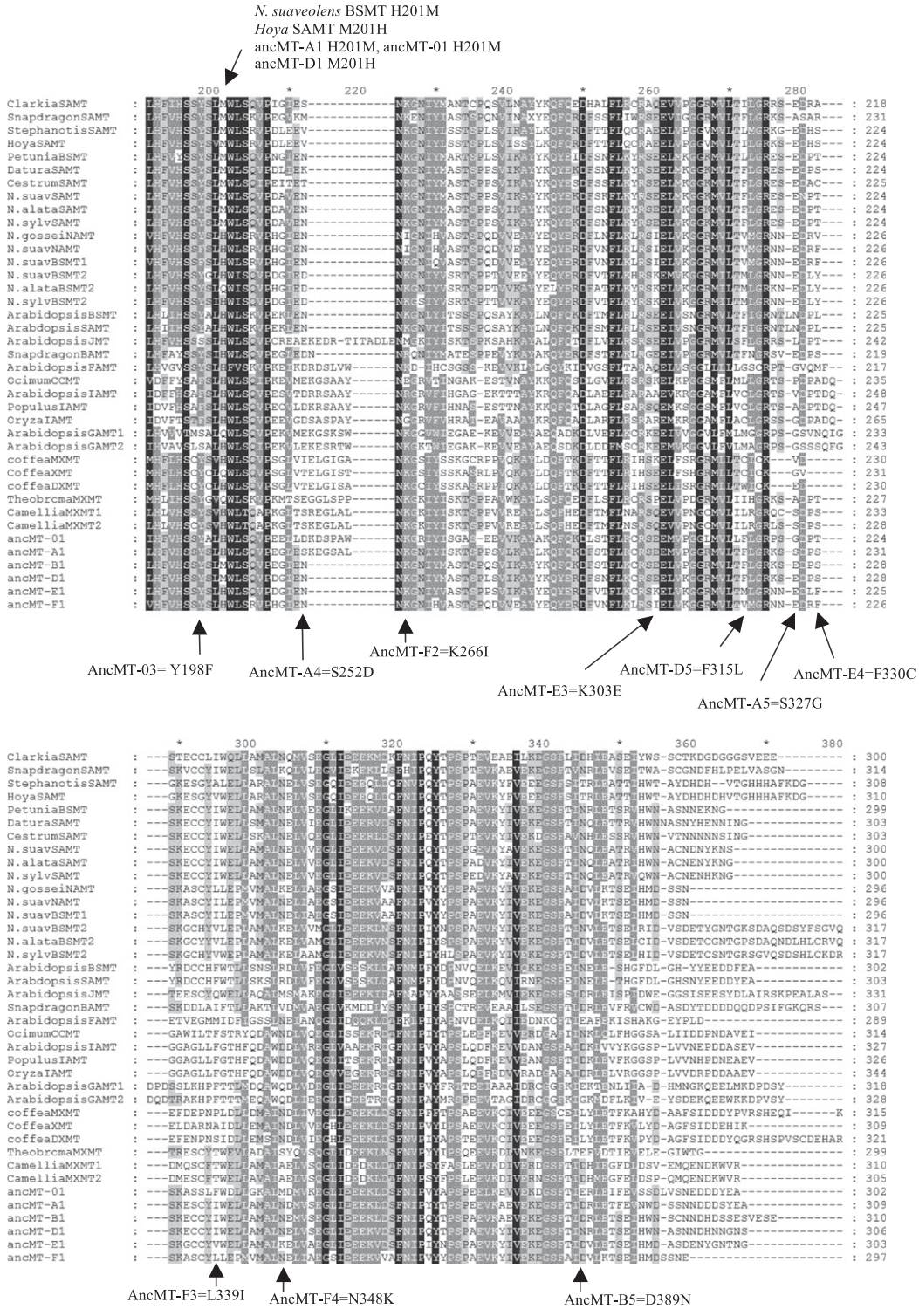

Fig. S3. (Continued)

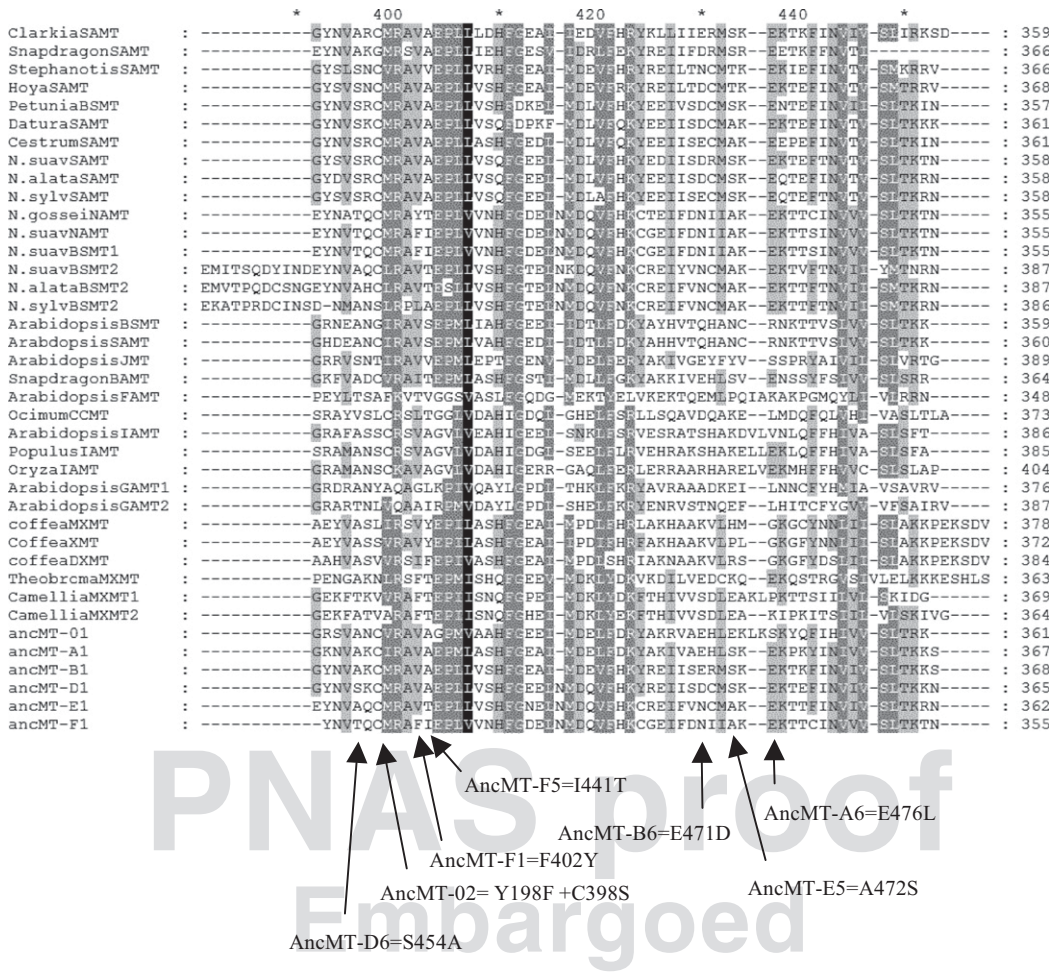

Fig. S3. Aligned amino acid sequences from functionally characterized SABATH family members including ancestral sequences from nodes 0, A, B, D, E, and F. Arrows show positions of mutations characterized in this study.

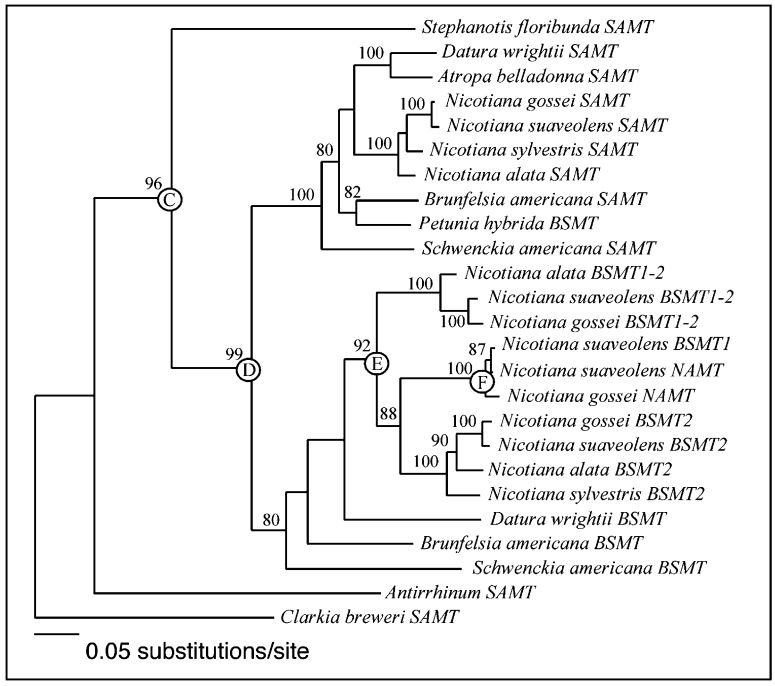

Fig. S4. Estimated relationships among *SAMT* and Solanaceae *BSMT* and *NAMT*. This gene tree was used for statistical analyses of positive selection and contained a denser sampling of Solanaceae gene sequences than shown in Fig. S2. Bootstrap values >80 are shown for corresponding nodes. *Nicotiana suaveolens BSMT1* was isolated from floral tissue of an individual grown in Germany, whereas *N. suaveolens NAMT* was obtained from floral tissue of an individual grown in Michigan. *BSMT1-2* have yet to be functionally characterized and are therefore tentatively named.

Table S1. Relative activities of four newly characterized SABATH family enzymes

	Hoya carnosa SAMT	*Datura wrightii* SAMT	*Cestrum nocturnum* SAMT	*Nicotiana suaveolens* NAMT
Salicylic acid	**100**	**100**	**100**	30.9
Benzoic acid	23.4	5.74	0.28	12.3
3-Hydroxybenzoic acid	4.6	10.79	0.75	2.5
4-Hydroxybenzoic acid	0	0.44	0	1.8
2,3-Dihydroxybenzoic acid	41.5	44.4	6.8	8
2,4-Dihydroxybenzoic acid	6.1	3.73	1.4	6
2,5-Dihydroxybenzoic acid	21.8	33.05	5.8	0.73
2,6-Dihydroxybenzoic acid	0	0	0.22	0.04
3,4-Dihydroxybenzoic acid	0	0.25	0	0.14
3,5-Dihydroxybenzoic acid	0	0	0	0.07
Cinnamic acid	0	0	0	10.3
o-Coumaric acid	0	0	0	1.27
m-Coumaric acid	0	0	0	5.25
p-Coumaric acid	0	0	0	0.03
2-Methoxy-BA (*o*-Anisic acid)	9.07	1.16	0	8.5
Anthranilic acid	12.9	2.32	2.86	25.2
Jasmonic acid	0	0	0	1.85
Nicotinic acid	0	0	0.12	**100**
Highest enzyme activity with salicylic acid (pkat/mg enzyme)	**257**	**211**	**38.9**	**58.32**

Q:2 Q:3 For each reaction, 1 mM substrate was added to Ni-NTA purified enzymes. Values are derived from specific activities measured in duplicate (n = 2). The highest activity with a given substrate was set to 100%.

Table S2. *H. carnosa* and *N. suaveolens* wild-type and mutant relative enzyme activities with various substrates

	H. carnosa SAMT	M201H	*N. suaveolens* BSMT1-TOPO	H201M-TOPO	*N. suaveolens* BSMT1-SUMO
Salicylic acid	**100**	**100**	11.83	**100**	38.75
Benzoic acid	23.4	82.98	**100**	61.98	93.03
3-Hydroxybenzoic acid	4.6	46.16	16.96	2.48	26.79
4-Hydroxybenzoic acid	0	1.57	72.98	40.5	86.89
2,3-Dihydroxybenzoic acid	41.5	45.5	13.61	69.01	27.7
2,4-Dihydroxybenzoic acid	6.1	2.71	8.88	63.22	29.16
2,5-Dihydroxybenzoic acid	21.8	12.13	0.02	0.83	1.39
2,6-Dihydroxybenzoic acid	0	0	0	0	0
3,4-Dihydroxybenzoic acid	0	0.33	3.75	0.5	8.92
3,5-Dihydroxybenzoic acid	0	1.67	0.04	0	0.02
Cinnamic acid	0	0.55	14	0.58	26.34
o-Coumaric acid	0	0.14	19.13	0.08	37.14
m-Coumaric acid	0	0	3.16	0.12	8.12
p-Coumaric acid	0	0.41	0.41	0.29	0.89
o-Anisic acid	9.07	35.77	71.01	5.79	88.24
Anthranilic acid	12.9	45.48	92.11	41.74	**100**
Jasmonic acid	0	0	3.75	0	8.65
Nicotinic acid	0	ND	ND	ND	33
Highest enzyme activity with favored substrate (pkat/mg enzyme)	**257**	**197**	**5.07**	**2.42**	**32.89**

For each reaction, 1 mM substrate was added to Ni-NTA purified enzymes. Values are derived from specific activities measured in duplicate (n = 2). The highest activity with a given substrate was set to 100%. ND, not determined.

Table S3. Michaelis–Menten kinetic parameters for *Hoya carnosa* SAMT, *Nicotiana suaveolens* BSMT1, and ancMT-D and mutants

	H. carnosa SAMT	M201H	*N. suaveolens* BSMT1-TOPO	H201M-TOPO	ancMT-D	ancMT-D M201H
K_M (SA), μM	61.5	79.84	162.2	975.29	15.77	6.9
K_M (BA), μM	156.5	548.7	148.55	1,989.74	295.3	78.1
k_{cat} (SA), s-1	17.1×10^{-3}	4.75×10^{-3}	2.59×10^{-5}	10.24×10^{-5}	2.31×10^{-3}	1.14×10^{-3}
k_{cat} (BA), s-1	5.41×10^{-3}	9.77×10^{-3}	9.28×10^{-5}	8.35×10^{-5}	9.78×10^{-4}	4.94×10^{-3}
k_{cat}/K_M (SA), $s^{-1} \times M^{-1}$	278.1	59.48	0.16	0.105	146.5	165
k_{cat}/K_M (BA), $s^{-1} \times M^{-1}$	34.6	17.8	0.625	0.042	3.31	63.3

AUTHOR QUERIES

AUTHOR PLEASE ANSWER ALL QUERIES

Q: 1_It has been assumed that the callout for ref. 30 in the sentence beginning "The Markov chain . . ." is ref. 30 from the main text, Hippauf et al. As such, it has been renumbered as SI ref. 3 and added to the SI reference list. Please check.

Q: 2_Please add the explanation of the use of boldface in Table S1 and Table S2 or delete the bold from the tables.

Q: 3_Please add the measure (%?) if needed to interpret/understand the numbers in Tables S1 through S3.