Plant Mol Biol (2010) 72:311–330
DOI 10.1007/s11103-009-9572-0

Enzymatic, expression and structural divergences among carboxyl *O*-methyltransferases after gene duplication and speciation in *Nicotiana*

Frank Hippauf · Elke Michalsky · Ruiqi Huang · Robert Preissner · Todd J. Barkman · Birgit Piechulla

Received: 21 July 2009 / Accepted: 4 November 2009 / Published online: 21 November 2009

Abstract Methyl salicylate and methyl benzoate have important roles in a variety of processes including pollinator attraction and plant defence. These compounds are synthesized by salicylic acid, benzoic acid and benzoic acid/salicylic acid carboxyl methyltransferases (SAMT, BAMT and BSMT) which are members of the SABATH gene family. Both *SAMT* and *BSMT* were isolated from *Nicotiana suaveolens*, *Nicotiana alata*, and *Nicotiana sylvestris* allowing us to discern levels of enzyme divergence resulting from gene duplication in addition to species divergence. Phylogenetic analyses showed that *Nicotiana SAMT*s and *BSMT*s evolved in separate clades and the latter can be differentiated into the *BSMT1* and the newly established *BSMT2* branch. Although *SAMT* and *BSMT* orthologs showed minimal change coincident with species divergences, substantial evolutionary change of enzyme activity and expression patterns occurred following gene duplication. After duplication, the BSMT enzymes evolved higher preference for benzoic acid (BA) than salicylic acid (SA) whereas SAMTs maintained ancestral enzymatic preference for SA over BA. Expression patterns are largely complementary in that *BSMT* transcripts primarily accumulate in flowers, leaves and stems whereas *SAMT* is expressed mostly in roots. A novel enzyme, nicotinic acid carboxyl methyltransferase (NAMT), which displays a high degree of activity with nicotinic acid was discovered to have evolved in *N. gossei* from an ancestral BSMT. Furthermore a SAM-dependent synthesis of methyl anthranilate via BSMT2 is reported and contrasts with alternative biosynthetic routes previously proposed. While BSMT in flowers is clearly involved in methyl benzoate synthesis to attract pollinators, its function in other organs and tissues remains obscure.

Keywords Carboxyl methyltransferase · SAMT · BSMT · NAMT · Benzoic acid · Salicylic acid · Nicotinic acid · Anthranilic acid · Methyl benzoate · Methyl salicylate · Methyl anthranilate · Methyl nicotinate · *Nicotiana*

Electronic supplementary material The online version of this article (doi:10.1007/s11103-009-9572-0) contains supplementary material, which is available to authorized users.

F. Hippauf · B. Piechulla (✉)
Institute of Biological Sciences, Biochemistry, University of Rostock, Albert-Einstein-Straße 3, 18059 Rostock, Germany
e-mail: birgit.piechulla@uni-rostock.de

F. Hippauf
e-mail: frank.hippauf@uni-rostock.de

R. Huang · T. J. Barkman
Department of Biological Sciences, Western Michigan University, Kalamazoo, MI 49008, USA
e-mail: ruiqi.huang@wmich.edu

T. J. Barkman
e-mail: todd.barkman@wmich.edu

E. Michalsky · R. Preissner
Institute for Physiology, Structural Bioinformatics Group, Charité Berlin, Arnimallee 22, 14195 Berlin, Germany
e-mail: elke.michalsky@charite.de

R. Preissner
e-mail: robert.preissner@charite.de

Introduction

Plant primary metabolism is conserved throughout land plants and is responsible for the production of compounds that are required for basic growth and development. By contrast, secondary (or specialized) metabolism is often variable among taxonomic groups and results in the

production of certain, often unique chemicals. These specialized metabolites may be necessary for survival but additionally improve individual fitness (Berenbaum 1995; Firn and Jones 2000; Nugroho and Verpoorte 2002). Altogether, more than 100,000 secondary metabolites of various structural classes have been isolated from plants (Nugroho and Verpoorte 2002; Noel et al. 2003). Many specialized metabolites are non-volatile, but a large proportion is volatile and plays diverse physiological and ecological roles. The large diversity of volatile secondary metabolites is generated by many different derivatisations and modifications of basic phenylpropanoid, terpenoid, and fatty acid structures including hydroxylation, acetylation, and methylation, all of which may alter the activity of the molecule and enhance volatility from tissues.

One common enzymatic modification of plant secondary metabolites is *O*-methylation, which results in the formation of ethers and esters (D'Auria et al. 2003). *O*-methyltransferases (*O*-MTs) that catalyse a methyl transfer reaction are grouped into three classes: (1) type I *O*-MTs exclusively methylate oxygen atoms of hydroxyl moieties of phenylpropanoid-based compounds, (2) type II are specific for phenylpropanoid esters of coenzyme A, and (3) type III methylate carboxyl groups of small molecules and also nitrogen atoms of certain alkaloids (Noel et al. 2003). Type III enzymes belong to the SABATH family which was described and named after the first three identified enzymes: salicylic acid carboxyl methyltransferase (SAMT), benzoic acid carboxyl methyltransferase (BAMT), and theobromine synthase (D'Auria et al. 2003). A total of 24 and 41 ORFs of the SABATH family have been identified from *Arabidopsis thaliana* and *Oryza sativa*, respectively (D'Auria et al. 2003; Zhao et al. 2008; Xu et al. 2006). The carboxyl MT members of this family transfer the activated methyl group from the ubiquitous methyl group donor *S*-adenosyl-L-methionine (SAM) to carboxyl groups of small molecules such as salicylic acid (SA), benzoic acid (BA), jasmonic acid, farnesoic acid, cinnamic/coumaric acid, indole-3-acetic acid and gibberellic acid (Ross et al. 1999; Murfitt et al. 2000; Seo et al. 2001; D'Auria et al. 2003; Effmert et al. 2005; Qin et al. 2005; Yang et al. 2006; Kapteyn et al. 2007; Varbanova et al. 2007; Zhao et al. 2008). Most of the enzymes encoded by this gene family in *A. thaliana* and *O. sativa* remain uncharacterized with respect to preferred substrates and in planta function.

The compounds synthesized by SABATH enzymes have various functions in plants. Methylated gibberellins and methyl-IAA have roles in plant development (Qin et al. 2005; Varbanova et al. 2007; Zhao et al. 2008; Yang et al. 2008). Methyl jasmonate is a well-known plant hormone involved in signal transduction cascades induced by biotic and abiotic stresses (Seo et al. 2001, Wasternack 2007). Caffeine and its precursors likely have a role in plant defense (Kim et al. 2006) but the role of methyl farnesoate is unclear in planta (Yang et al. 2006). Some of the most well-studied compounds produced by this family of enzymes include methyl salicylate (MeSA) and methyl benzoate (MeBA). Methyl salicylate was shown to act as a plant-plant communication signal and its unmethylated form (SA) was thought for a long time to be required to develop systemic acquired resistance (SAR; Shulaev et al. 1997; Seskar et al. 1998). Only recently it was shown that MeSA is the mobile signal leading to the development of SAR (Park et al. 2007). MeSA has also been shown to be emitted from herbivore-damaged leaf tissues (van Poecke et al. 2001; Van den Boom et al. 2004). MeSA and MeBA are often found in floral scents likely playing roles in pollinator attraction because insects can detect the molecules and show behavioural responses to them (Raguso et al. 1996; Fraser et al. 2003; Hoballah et al. 2005; Knudsen et al. 2006).

The enzymes that catalyze the formation of MeSA and MeBA are well studied and have been divided into two categories according to their methyl acceptor preferences: the SAMT-type and the BAMT-type (Effmert et al. 2005). The primary substrates for the two enzyme types are structurally similar yet the enzymes have evolved distinct preferences. SAMTs possess a lower Km value and higher catalytic efficiency for SA than for BA. The enzymes of the BAMT-type can be divided into BAMTs and BSMTs (benzoic acid/salicylic acid carboxyl methyltransferase). The BAMT is highly specific to BA, whereas BSMTs often possess similar Km values for both substrates but have a higher catalytic efficiency for BA. So far only one BAMT isolated from *Antirrhinum majus* was described (Murfitt et al. 2000, Effmert et al. 2005). The overall amino acid sequence identities between the SAMT- and BAMT-type enzymes range from 35 to 45%, and several differences are found in the active pockets. One conspicuous structural difference between both enzyme types is the presence of a Met (position 150 in *C. breweri* SAMT) residue in the SAMT-type that is replaced by a His residue in the BAMT-type enzymes (Effmert et al. 2005; Barkman et al. 2007). The *SAMT* genes found in *Clarkia breweri*, snapdragon, and various Solanaceae and Apocynaceae are expressed in flowers, roots and leaves (Ross et al. 1999; Negre et al. 2002; Fukami et al. 2002; Pott et al. 2002). Members of the BAMT-type include *BAMT* of *Antirrhinum majus* and *BSMT* of *A. thaliana*, *A. lyrata* and *Nicotiana suaveolens* (Murfitt et al. 2000; Chen et al. 2003; Pott et al. 2004). Whereas, *BAMT* of *A. majus* and *BSMT* of *N. suaveolens* are mainly expressed in flowers, *Arabidopsis BSMT*s are expressed in leaves, stems and flowers (summarized in Effmert et al. 2005).

SAMT and *BSMT* both occur in Solanaceae and appear to have resulted from a gene duplication event early in the history of the family (Barkman et al. 2007; Martins et al.

2007). Members of the SAMT-type are present in all sampled members of the family but only those of *Atropa belladonna*, *Datura wrightii*, and *Petunia hybrida* have been functionally characterized (Fukami et al. 2002, Negre et al. 2003, Barkman et al. 2007). Partial *BSMT* sequences have been isolated from a few members of the family but only in *N. suaveolens* has the enzyme been characterized (Pott et al. 2004). The presence of duplicated genes encoding functionally similar enzymes in the Solanaceae provides an opportunity to investigate their potential evolutionary fates. Although most duplicated genes are predicted to become pseudogenes, at least three other outcomes are possible (Zhang 2003; Moore and Purugganan 2005). Complete conservation of expression patterns and enzyme function may occur in both duplicates, although this is likely a rare outcome. More commonly, subfunctionalization of duplicated genes (and the enzymes they encode) results in the evolution of tissue specific expression for one or both duplicates to collectively carry out the ancestral functions. Subfunctionalization may also result in evolution of the coding sequences to partition the ancestral functions that the single progenitor performed. Finally, neofunctionalization may also occur in which case novel functions evolve in one duplicate gene while the other maintains ancestral function. The presence of the duplicated *SAMT* and *BSMT* in Solanaceae provides an excellent opportunity to examine enzyme evolution in terms of expression pattern and enzymatic activity.

The family Solanaceae is distributed worldwide and contains many taxa of agronomic (potato, tomato, and pepper) and medicinal (mandrake, tobacco, deadly nightshade and henbane) importance. *Nicotiana* is the fifth largest genus in the family (75 species in 13 sections), with species distributed primarily in America and Australia (Goodspeed 1954; Knapp et al. 2004). The phylogeny of *Nicotiana* is well understood (Chase et al. 2003; Clarkson et al. 2004) and numerous species have been studied in terms of floral scent and pollination (Loughrin et al. 1990; Raguso et al. 2003; Raguso et al. 2006). Within the genus, it was shown that *N. suaveolens* emits high levels of MeBA and little MeSA, whereas *N. alata* emits little MeBA and traces of MeSA. For *N. sylvestris* only MeBA emission could be detected (Raguso et al. 2003). This scent variation could be due to differences in substrate availability, expression levels of SAMT and BSMT as well as alterations of enzyme activity as a result of structural differences of the amino acid sequences of the active site. The different floral emission profiles of MeSA and/or MeBA by *Nicotiana alata*, *N. sylvestris*, and *N. suaveolens*, provide an opportunity to investigate the divergence of these chemical phenotypes at the molecular level. Because both *SAMT* and *BSMT* were sampled from the same three species we had an opportunity to examine enzymatic and expression evolution through two gene duplication events and two species divergences. Enzyme divergence arising from a third speciation event was studied for *N. gossei*, a close relative of *N. suaveolens* that, as shown below, differs markedly in floral scent composition.

Materials and methods

Plant material and growth conditions

Seeds from *Nicotiana alata* TW7 and *Nicotiana sylvestris* Speg. & Comes were obtained from Dr. Robert A. Raguso (Cornell University). *Nicotiana suaveolens* Lehm., *N. alata* and *N. sylvestris* plants were grown in growth chambers on vermiculite under long day conditions (16 h illumination at 160 μmol m^{-2} s^{-1} and 22°C, 8 h darkness at 18°C) as described in Roeder et al. (2007). All plants were watered with Hoagland solution (Hoagland and Aronon 1938).

RNA isolation

Plant tissue was harvested from ~3-month-old plants that just began flowering. For determination of gene expression levels, tissue of stems, leaves, roots and 1 day old flowers that opened the night before were harvested and pooled at 06.00 am and 06.00 pm. Plant tissue was immediately frozen in liquid nitrogen and stored at −70°C. RNA was isolated from 0.5 to 1 g frozen plant material according to Chang et al. (1993) and RNA was stored at −70°C. To isolate *SAMT*, a leaf disk (1 cm^2) was incubated in 5 mM SA (pH 6.5) for 24 h to induce gene expression prior to RNA extraction (Martins and Barkman 2005). For determination of expression levels of *Nicotiana SAMT* and *BSMT* genes after induction with SA or BA in leaves, leaf disks (1 cm^2) were incubated in 5 mM SA or BA (pH 6.5) for 24 h and RNA was isolated as described above.

Isolation of *BSMT* and *SAMT* by RT-PCR

For isolation of *BSMT* and *SAMT* sequences, RNA was isolated from leaves of at least two different plants of each species, as described above. Prior to RT-PCR, a DNaseI (Fermentas) digestion was performed at 37°C for 30 min to avoid genomic contamination. All RT-reactions were performed with 2 μg of total RNA and SuperScript III reverse transcriptase (Invitrogen) to amplify *SAMT* and *BSMT* cDNA sequences according to the manufacturer's instructions. Five microliters of the cDNA synthesis mix was added to the Qiagen Mastermix (Qiagen) and PCR reactions were run under the following conditions: 90 s at 94°C for an initial denaturation, followed by 35 cycles of 30 s at 94°C for denaturation, 30 s at 50–60°C for annealing

(depending on the primer used), 1 min at 72°C for extension and a final extension at 72°C for 5 min. For isolation of sequences, primers were derived from known *SAMT* and *BSMT* sequences. It is possible that additional *SAMT* and *BSMT* sequences are present in the investigated *Nicotiana* species, which could be isolated with other primer pairs. All primer sequences are shown in Supplemental Table S1. The PCR products were analyzed by gel electrophoresis and were recovered from the agarose gel using a gel extraction kit (Qiagen). The purified fragments were cloned using the pGEM-T Cloning kit (Promega) and between 10 and 15 clones per PCR product were sequenced using the SequiTherm Excel II DNA Sequencing kit on a LI-COR automated sequencer (MWG-Biotech). For sequencing, IRD-800 labeled T7 and SP6 promoter primers were used. For the newly isolated *BSMTs* and *SAMTs* one sequence was obtained. Nucleotide alterations, which appeared rarely and randomized and not consistently were considered as artifacts (e.g. *Taq*-polymerase or sequencing errors). However, it is possible that there could have been multiple alleles in the individuals sampled (that they may have been heterozygotes) but we did not detect both alleles. The resulting amino acid sequences encoded by these fragments were compared to known protein sequences of databases using BLAST (National Center for Biotechnology Information [NCBI]).

Sequence completion of SAMT and BSMT sequences

To isolate full-length cDNA, 5′ and 3′ RACE was performed. The primer sequences and amplification conditions for all reactions are shown in Supplemental Table S2 and S3. All RT-PCR and RACE reactions were performed after DNaseI (Fermentas) digestion of 2 µg total RNA at 37°C for 1 h. 5′UTR-regions of all *SAMT*s were isolated by RT-PCR using a primer derived from the 5′UTR of *SAMT* from *N. tabacum* (Martins and Barkman 2005). To obtain the 5′UTR sequences of the *BSMT* genes, ThermoScript RT-PCR system (Invitrogen) was used. The cDNA synthesis was carried out at 54°C for 1 h. The reactions were purified with the Millipore Montage Kit (Millipore) to remove all nucleotides according to the manufacturer's protocol. For adding a polyadenosine sequence to the cDNAs, terminal deoxynucleotidyl transferase (15 u/µl; Invitrogen) was used following the manufacturer's protocol. Five microliters of tailing reaction was used for a 25 µl PCR. In contrast to the procedure described above, *N.sua.BSMT2* was completed by using the start-primer from *N.sua.BSMT1-1* (5′-ATGGAAGTTGCCAAAGTTCT-3′). All amplified fragments were recovered from an agarose gel, cloned into pGEM-T vector (Promega) and sequenced with IRD-800 labeled primer as described above.

To identify the 3′UTR of *SAMT* from *N. alata* (*N.ala.SAMT*) and *N. suaveolens* (*N.sua.SAMT*) as well as *BSMT* of *N. alata* (*N.ala.BSMT2*) and *N. sylvestris (N.syl.BSMT2*) 3′-RACE was performed. RT-reactions were carried out with a temperature program as described by Pott et al. (2004). Isolation of the 3′UTR of *BSMT* from *N. suaveolens* (*N.sua.BSMT2*) and *SAMT* from *N. sylvestris* (*N.syl.SAMT*) used the ThermoScript RT-PCR system (Invitrogen) and primers derived from the 3′UTRs of the isolated *Nicotiana BSMT* and *SAMT* sequences. The *N.sua.BSMT2* cDNA synthesis was carried out at 50°C. The temperature program for cDNA synthesis of *N.syl.SAMT* followed a gradual decrease of temperature from 65 to 50°C to ensure the optimal primer annealing (Supplemental Table S3). The amplified fragments were recovered from an agarose gel, cloned into pGEM-T vector (Promega) and sequenced with IRD-800 labeled primer as described above.

Cloning into expression vectors

The full-length *N.sua.BSMT2* and *N.ala.BSMT2* as well as the *N.sua.SAMT*, *N.syl.SAMT* and *N.gos.NAMT* were cloned into the expression vector using the pET SUMO Expression kit (Invitrogen) according to the manufacturer's instructions. The full-length *N.syl.BSMT2* and *N.ala.SAMT* were cloned into the expression vector using the pET101 Directional TOPO Expression kit (Invitrogen). Two micrograms of total RNA was digested with DNaseI at 37°C for 1 h as described above. The RT reaction was carried out at 50°C for 1 h using SuperScript III reverse transcriptase (Invitrogen). Five microliters of the RT reaction was used for a 25 µl PCR. The primer sequences and amplification conditions for all reactions are shown in Supplemental Table S4. All plasmids were transformed into TOP10 cells (Invitrogen). To ensure the right orientation of sequences and detect possible errors resulting from *Taq*-polymerase amplification, the fragments were sequenced as described above.

Floral scent sampling

SPME headspace sampling was performed for 1 h using airtight vials. The portable SPME field sampler was composed of a PDMS stationary phase with a film thickness of 100 µm (Supelco). SPME fibers were conditioned using split mode for 15 min at 250°C prior to use. Fibers were exposed to the floral headspace of *N. gossei* flowers for 1 h at night (8.00 pm). Compounds were desorbed in the injector port for 1 min using the splitless mode. GC–MS analyses were performed on an HP6890 GC System equipped with a DB-5 capillary column coupled to an HP5973 Mass Selective Detector. The oven conditions

were 40°C for 2 min, ramping 20°C/min to 300°C with a 2 min hold.

GC–MS analysis of enzyme products

Prior to purification, activity of the enzymes was tested in 50 ml LB cultures. A 50 ml cell culture of HMS174 (DE3) expressing *BSMT*, *SAMT* or *NAMT* was induced with 1 mM isopropyl thiogalactoside after reaching an OD_{600} of 0.6 and was then further incubated at 20°C. Thirty minutes after induction, 1 mM (final concentration) BA, SA or NA was added and incubation was continued for an additional 20 h. After removing the cells by centrifugation, the remaining supernatant (~40 ml) was extracted with 3 ml of hexane. Samples were analyzed on a DB5-MS column (60 m × 0.25 mm × 0,25 µm; J&W Scientific) in a GC–MS-QP5000 (Shimadzu) with helium as the carrier gas at a flow rate of 1.1 ml min^{-1}. One µl of hexane was injected into the splitless injector port which was held at 200°C. The temperature program started at 35°C, with a 2 min hold, and temperature ramping to 280°C at a rate of 10°C min^{-1}, and a final 15 min hold. Products were identified via comparison of mass spectra and retention times with those of available standards and with spectra in the library of National Institute of Standards and Technology (NIST 147).

Heterologous expression and purification of recombinant protein

Escherichia coli strain HMS174 (DE3) was used for overexpression of His_6-tagged genes. Overexpressed proteins were obtained after preincubation of cells at 37°C until OD_{600} of 0.6 was reached. Cells were induced with 1 mM isopropyl thiogalactoside and incubation continued for 20 h at 20°C. The cells were harvested and centrifuged at 4°C at 6,000*g* for 10 min, resuspended in 5 ml of lysis buffer (50 mM NaH_2PO_4, pH 8.0; 300 mM NaCl; 10 mM imidazol; 10% [w/v] glycerol; 10 mM ß-mercaptoethanol) and sonicated for 10 s, ten times on ice. The soluble extract was centrifuged at 12,000*g*. The overexpressed protein was purified by Ni–NTA affinity chromatatography (Qiagen) according to the manufacturer′s instructions. After two washing steps, the recombinant protein was eluted with 500 µl extraction buffer containing 250 mM imidazol. Protein concentrations were measured using the standard Bradford assay (Bradford 1976). Protein purification was checked on 12.5% SDS polyacrylamide gels.

Enzyme assays

The purified and His-tagged BSMT and SAMT enzymes were tested for enzyme activity (Wang et al. 1997). All substrates shown in Table 2 were added at a final concentration of 1 mM to each assay. The 50 µl assays contained 10 µl of purified protein, 10 µl of assay buffer (250 mM Tris–HCl, pH 7.0; 25 mM KCl), 1 µl of 50 mM unmethylated substrate, 1 µl S[methyl-^{14}C]adenosyl-L-Met (58 mCi $mmol^{-1}$; Hartmann Analytics), and 28 µl H_2O. As a control reaction, 1 µl of pure ethanol was added instead of the unmethylated substrate. The samples were incubated at 25°C for 40 min. The reaction was stopped by adding 3 µl concentrated HCl followed by the addition of 100 µl ethyl acetate for extracting the labeled methylated product. The samples were mixed and centrifuged for 1 min at 10,000 rpm. 30 µl of the upper organic phase was transferred to a scintillation vial, mixed with 2 ml scintillation fluid (Perkin-Elmer) and counted in a scintillation counter (Tri-Carb 2100 TR; Canberra Packard). Relative enzyme activity with each substrate was calculated and the product which reached the highest dpm value (counts per minute) per time unit was set to 100%.

Determination of gene expression patterns by RT-PCR

An RT-PCR approach was used to examine in which plant organ and at which time point the *SAMT* and *BSMT* genes were expressed in the *Nicotiana* species. This qualitative method will highlight different expression patterns occuring between the different species but was not used to quantitate RNA accumulation levels. Total RNA was isolated from pooled leaves and flowers of three different plants per species and from stems and roots of two different plants per species at 06.00 am and 06.00 pm as described above. RNA concentration was determined photometrically and checked via gel electrophoresis. Prior to RT-PCR, a DNaseI digestion of 2 µg total RNA at 37°C for 60 min was performed. The RT-reactions were carried out with SuperScript III reverse transcriptase (Invitrogen) for 1 h at 52°C (for primers, see Supplemental Table S5). PCR was carried out using *Taq* PCR Master Mix Kit (Qiagen). To each reaction 12.5 µl Master mix, 5 µl cDNA, 1 µl (10 µM) of each primer and 5.5 µl RNase-free water was added to reach a final volume of 25 µl. Cycling conditions were as follows: denaturation at 94°C for 90 s, annealing at 54°C for 30 s and extension for 30–60 s (depending on the expected length of the amplification products) at 72°C. At the end of the cycling there was a 5 min final extension step at 72°C. Expression of the plant translation elongation factor 1α (*EF-1α*) gene was used as an external control. RT reactions were done using Moloney murine leukemia virus (MMLV) reverse transcriptase (Invitrogen), at an annealing temperature of 42°C for 1 h. Five microliters of the cDNA synthesis mix was added to the Qiagen Mastermix (Qiagen) and PCR reactions were run under the following conditions: 90 s at 94°C for an initial denaturation, followed by

30 s at 94°C for denaturation, 30 s at 50°C for annealing, 30 s at 72°C for extension and a final extension at 72°C for 5 min. The PCRs were carried out with following cycle numbers: N.sua.BSMT1-1: 30×, N.sua.BSMT1-2: 40×, N.sua.BSMT2: 40×, N.sua.SAMT: 40×, N.ala.BSMT1: 35×, N.ala.BSMT2: 30×, N.ala.SAMT: 40×, N.syl.BSMT2: 30×, N.syl.SAMT: 35×, EF1: 25×.

Phylogenetic tree construction

DNA sequences from all enzymatically characterized SABATH gene family members were obtained from GenBank or were generated as part of this study. Other uncharacterized EST sequences from several rosid species were obtained by BLAST analysis to assess relationships with characterized sequences. All partial sequences had missing sequence coded as "?" which is interpreted as missing data. DNA sequences were aligned with ClustalX (Thompson et al. 1997) with subsequent minor adjustments to preserve codon structure. Alignment ambiguous regions were excluded from analyses because homology among such sites could not be confidently determined. Maximum likelihood analyses, assuming the GTR+I+G model of nucleotide substitution as chosen by Modeltest (Posada and Crandall 1998), were performed with PAUP* (Swofford 2003). Maximum likelihood bootstrapping was performed using 100 replicates using GARLI (Zwickl 2006). Phylogenetic tree estimation was also performed using Bayesian analyses using MrBayes v3.1.2 assuming the best-fit model of nucleotide substitution (Huelsenbeck and Ronquist 2001). Four chains were simultaneously run for one million generations and these were sampled every 100 generations. The first 10,000 generations were discarded as the "burn-in" period based on inspection of the scores obtained and posterior probabilities (PP) for individual clades were then obtained from the remaining samples.

Ancestral state estimates of the ratio of MeSA/MeBA were obtained using BayesTraits (Pagel et al. 2004). For the analyses of this continuous variable, a posterior distribution of alpha (the estimate of the ancestral state at the root of the tree) was obtained. This distribution was then used for ancestral state estimation using a MCMC chain that was run for 1 million iterations that was sampled every 100 generations with a burn-in of 50,000. A uniform prior was assumed. Histograms of ancestral states shown in Fig. 6 were generated by plotting the estimates from each sampled iteration of the MCMC chain. Estimates of the posterior probability of ancestral amino acids and tissue-specific gene expression patterns were obtained using the reversible-jump hyperprior approach assuming an exponential distribution. For all chains, the RateDev parameter was set to achieve a 20–40% acceptance rate.

Modeling of N.sua.BSMT2, N.sua.SAMT and N.gos.NAMT structures

Enzyme models were built via homology modeling using the crystal structure of the *Clarkia breweri* SAMT as a template (Zubieta et al. 2003). Modeling, energy optimization and assignment of the secondary structures were performed with the Swiss-PdbViewer software (Guex and Peitsch 1997). Missing loops were modeled using the tool SuperLooper (Hildebrand et al. 2009). It was not possible to model the region 305–329 of N.sua.BSMT2, because the insertion of about 20 amino acids is too long to obtain a reasonable structure prediction. Docking of the substrates was done with the GOLD software (Verdonk et al. 2003). An analysis of intermolecular interactions was also performed using InsightII (Accelrys Inc.). Docking was achieved using Monte Carlo simulations and simulated annealing in which the ligand and residues within 6 Å (angstroms) of it were defined as flexible. Total energy, interaction energy between the ligand and protein, and LUDI 3 scores were calculated and compared among the models.

Sequence data from this article have been deposited in GenBank under following accession numbers: GU014480 for *N. suaveolens BSMT2*; GU014481 for *N. suaveolens BSMT1-2*-like cDNA sequence; GU014479 for *N. suaveolens SAMT*; GU014483 for *N. alata BSMT2*; GU014484 for *N. alata BSMT1*-like cDNA sequence; GU014482 for *N. alata SAMT*; GU014486 for *N. sylvestris BSMT2*; GU014485 for *N. sylvestris SAMT*; GU169286 for *N. gossei NAMT*; GU169289 for *N. gossei SAMT*-like cDNA sequence; GU169288 for *N. gossei BSMT2*-like cDNA sequence; GU169287 for *N. gossei BSMT1-2*-like cDNA sequence.

Results

GC–MS headspace analysis of *N. gossei*

We sampled the headspace of *N. gossei* flowers using solid phase microextraction (SPME). One of the most abundant compounds detected in the headspace was methyl nicotinate (MeNA; peak 3; Fig. 1a). Although *N. gossei* is closely related to *N. suaveolens*, the floral scent of these species is particularly different in that *N. suaveolens* produces predominantly MeBA and only small amounts of MeNA, whereas *N. gossei* does not produce detectable quantities of MeBA. MeNA emission is rare in *Nicotiana* (Raguso et al. 2003; Raguso et al. 2006) but it has been reported from the headspace of at least six other angiosperm families, many of which exhibit a moth pollination syndrome, like *N. gossei* (Knudsen et al. 3). In addition to MeNA, we found numerous sesquiterpenes, like alpha-farnesene, that were particularly abundant and together

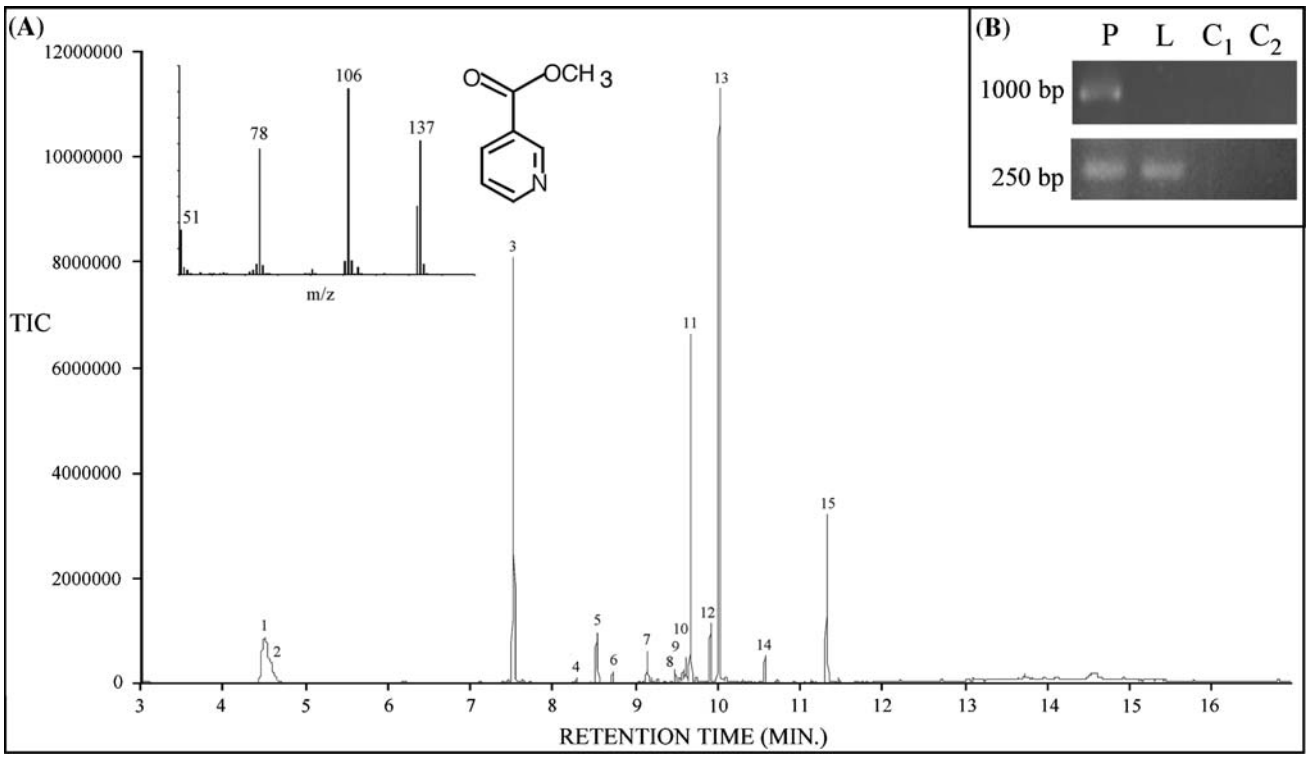

Fig. 1 *N.gos.NAMT* is highly expressed in petals of *N. gossei* where MeNA emission was detected. **a** Total ion chromatogram for SPME sampled headspace of *Nicotiana gossei* flowers. Inset mass spectrum was obtained from peak 3 and is diagnostic for methyl nicotinate which is drawn above the peak. Numbered peaks refer to top 15 most abundant compounds in headspace. Compound identifications are based on comparisons with NIST library spectra. Tentative names are provided only if mass spectra matched >90% with the library. *1* 2-methyl butyl aldoxime (syn; nitrogenous compound), *2* 2-methyl butyl aldoxime (anti; nitrogenous compound), *3* methyl nicotinate (nitrogenous compound), *4* geraniol (oxygenated monoterpene), *5* contaminant, *6* unknown, *7* 3-(1-methyl-2-pyrrolidinyl)-pyridine (nitrogenous compound), *8* 2,6-dimethyl-6-(4-methyl-3-pentenyl)-bicyclohept-2-ene (sesquiterpene hydrocarbon), *9* unknown, *10* 2,6-dimethyl-6-(4-methyl-3-pentenyl)-bicyclohept-2-ene (sesquiterpene hydrocarbon), *11* 7, 11-dimethyl-3-methylene-1, 6, 10-dodecatriene (sesquiterpene hydrocarbon), *12* 3, 7, 11-trimethyl-1, 3, 6, 10-dodecatetraene (sesquiterpene hydrocarbon), *13* alpha-farnesene, *14* unknown, *15* 3, 7, 11-trimethyl-2, 6, 10-dodecatrien-1-ol (oxygenated sesquiterpene). **b** RT-PCR results showing floral specific expression of a BSMT-like sequence in petals of *N. gossei*. 1,000 bp band corresponds to near-full length BSMT-like cDNA. 250 bp band corresponds to actin cDNA. *P* petal tissue, *L* leaf tissue, C_1 negative (-RNA) control, C_2 negative (-RT step) control

these accounted for ca. 50% of the headspace volatiles. Within *Nicotiana*, only *N. sylvestris* appears to have floral scent that is also rich in sesquiterpenes with caryophyllene accounting for up to 48% of its headspace (Loughrin et al. 1990; Raguso et al. 2003). Methyl butyl aldoximes were also detected in *N. gossei* and these compounds are found in other members of *Nicotiana* and appear to be found in the headspace of many moth-pollinated plant species (Raguso et al. 2003).

Isolation and sequence characterization of *SAMT*s and *BSMT*s from *N. suaveolens*, *N. sylvestris*, *N. alata* and *N. gossei*

Using RT-PCR, we successfully isolated full length *SAMT* and *BSMT*-like sequences from leaves of *N. suaveolens*, *N. alata* and *N. sylvestris* (Table 1). The complete coding sequence of all putative *SAMT*s comprised 1,074 nucleotides, which is consistent with the length known from other plant species. The predicted SAMT protein consists of 358 aa with an estimated molecular weight of 40.7 kD. The isolated 5′UTRs were fairly similar ranging from 43 to 78 nucleotides, while the length of the 3′-UTRs varied significantly from 75 to 323 nucleotides. The complete open reading frames of the putative *BSMT* sequences included 1,161 nucleotides in *N. suaveolens* and *N. alata*, and 1,158 nucleotides in *N. sylvestris* with predicted protein lengths of 387 amino acids (386 in *N. sylvestris*) and calculated molecular masses of 43.4–43.7 kD (Table 1). The length of the 5′UTRs of *BSMT*s ranged between 43 and 53 nucleotides, while the isolated 3′UTRs varied between 68 and 135 nucleotides. A *BSMT*-like sequence was isolated from *N. gossei* floral tissue and is 1,065 bp, which is the same length as the previously isolated floral *BSMT* of *N. suaveolens* (Pott et al. 2004; Table 1). This sequence was expressed in petal tissue but no expression was

Table 1 Newly isolated *Nicotiana* carboxyl methyltransferases

Species	Enzyme nomination	ORF length (bp)	Protein length (aa)	Estimated molecular mass of the protein (kD)	Expression
N. suaveolens	N.sua.BSMT1-2	833	277 partial	–	L[a]; S[b]; F[c]
N. suaveolens	N.sua.BSMT2	1,161	387	43.7	L; R[d]; S; F
N. suaveolens	N.sua.SAMT	1,074	358	40.7	R
N. alata	N.ala.BSMT1	421	140 partial	–	L; R; S; F
N. alata	N.ala.BSMT2	1,161	387	43.5	L; S; F
N. alata	N.ala.SAMT	1,074	358	40.7	L; R; S; F
N. sylvestris	N.syl.BSMT2	1,158	386	43.4	L; S; F
N. sylvestris	N.syl.SAMT	1,074	358	40.7	R
N. gossei	N.gos.NAMT	1,065	355	39.9	F
N. gossei	N.gos.BSMT1-2	450	150 partial	–	L
N. gossei	N.gos.BSMT2	722	240 partial	–	L
N. gossei	N.gos.SAMT	376	125 partial	–	L

The N.sua.BSMT1-1 (not shown in Table 1) was previously isolated (Pott et al. 2004)

[a] Leaf

[b] Stem

[c] Flower

[d] Root

detected in leaves (Fig. 1b). The newly isolated *BSMT* genes from *N. suaveolens*, *N. alata* and *N. sylvestris* encode 31 (32) amino acids more than the previously isolated *BSMT* from *N. suaveolens* (hereafter referred to as *N.sua.BSMT1-1*) due to an insertion near the C-terminal end of the protein (Supplemental Fig. S1). To distinguish the new sequences from the floral *N.sua.BSMT1-1* sequence, they are hereafter referred to as *BSMT2*. Since two *BSMT* sequences were obtained from *N. suaveolens* (*N.sua.BSMT2* and floral *N.sua.BSMT1-1* from Pott et al. 2004), we attempted to isolate additional genes via RT-PCR. Partial *SAMT* and *BSMT* sequences were obtained from *N. alata*, *N. gossei* and *N. suaveolens* (named as *N.ala.BSMT1*, *N.gos.SAMT*, *N.gos.BSMT1-2*, *N.gos.BSMT2* and *N.sua.BSMT1-2*, respectively) and indicate that further genes of this family exist and are expressed (Table 1). Although the latter sequences are partial, we used them in phylogenetic analyses to provide a clearer picture of *SAMT*/*BSMT* gene family evolution within Solanaceae.

Phylogenetic relationships of *Nicotiana SAMT*s and *BSMT*s

The SAMT predicted protein sequences were 93.3–96.4% identical to each other (Supplemental Table S6). Identities of the SAMTs relative to the isolated BSMT protein sequences ranged from 58.6 to 61.4%. The newly isolated BSMT2 sequences were ~86.0% identical to each other and 74.1–75.8% identical to the floral N.sua.BSMT1-1 from *N. suaveolens* (Pott et al. 2004). The BSMT-like sequence from *N. gossei* was very similar to the floral N.sua.BSMT1-1 (95.8%) differing by only 15 amino acid changes, which is not surprising because of the close phylogenetic relationship of these species (Chase et al. 2003).

A phylogenetic analysis including all enzymatically characterized members of the SABATH gene family indicates that the functionally distinct members are highly divergent from each other (Fig. 2a). All *SAMT* and *BSMT* from Solanaceae form part of a strongly supported monophyletic lineage that is likely 125 million years old because sequences have been isolated from both rosid (*Clarkia* and others) and asterid (*Nicotiana*, *Stephanotis*) species, the two major lineages of eudicots (Fig. 2b). Because this eudicot tree is similar to angiosperm phylogeny, the apparently non-duplicated sequences shown in Fig. 2b are more likely orthologs rather than paralogs (Fitch 2000). Additional sequences from other diverse angiosperms will be necessary to increase the resolution of this gene tree and provide confirmation of the orthology of these sequences. The fact that *Carica*, a member of the same order as *Arabidopsis* (Brassicales), has an *SAMT* ortholog indicates that the absence of an orthologous sequence in *Arabidopsis* is due to a loss at some point since the origin of their common ancestor (Fig. 2b). This phylogenetic analysis also indicates that there may have been four independent origins of genes that encode enzymes with SA/BA carboxyl methyltransferase activity in flowering plant history, once in *Arabidopsis BSMT*, once in *Antirrhinum BAMT*, once in rice *BSMT*

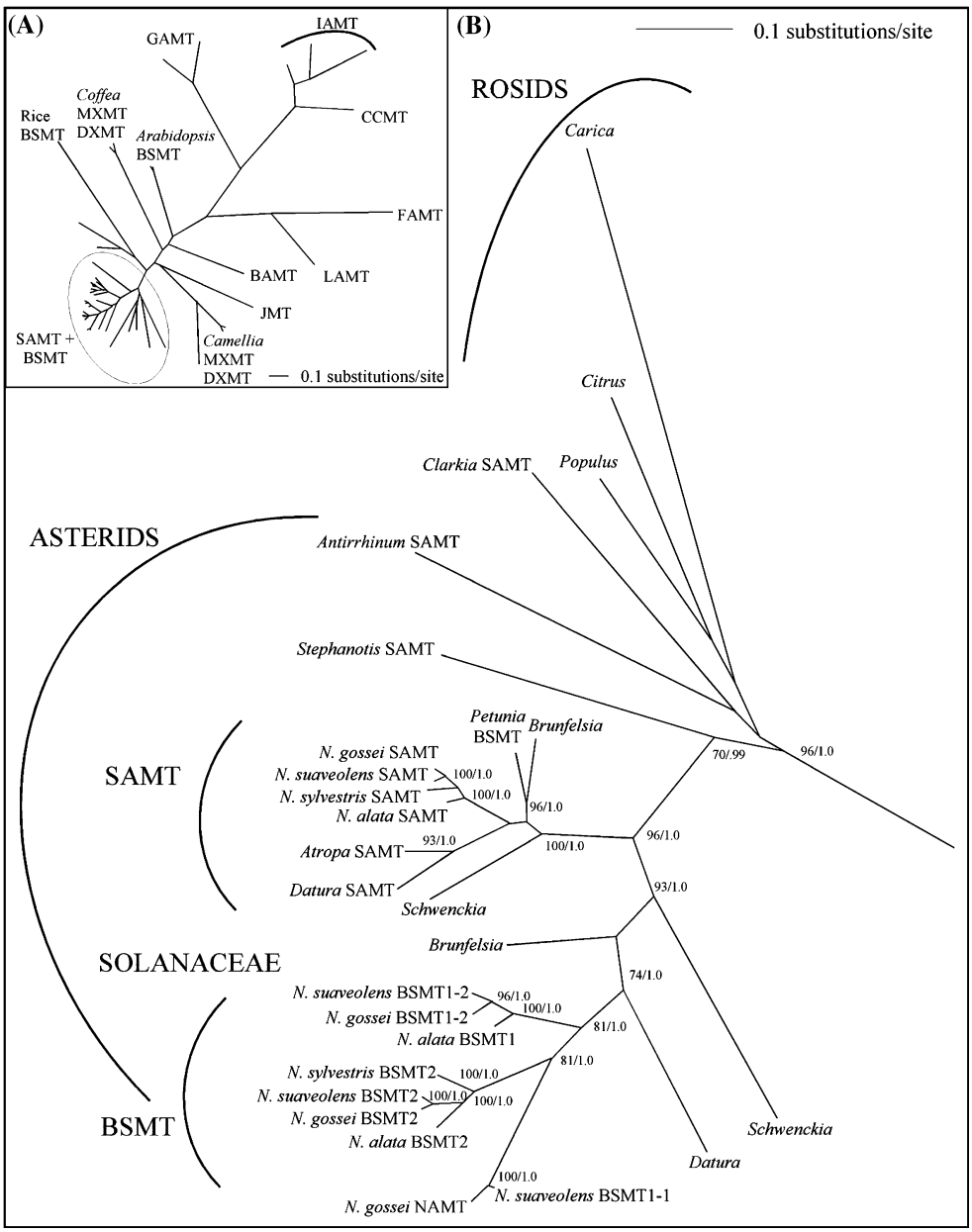

Fig. 2 Phylogenetic relationships among SABATH gene family members. **a** Unrooted phylogenetic tree of enzymatically characterized carboxyl methyltransferases. All *SAMT* and *BSMT* from Solanaceae appear to be monophyletic (shown by *ellipse*) and are evolutionarily divergent from all other members of the gene family. The isolated sequences from *Nicotiana* species characterized in this paper are all members of this lineage. Accession numbers are shown in Supplemental Table S7. **b** Detailed phylogenetic analysis of the circled SAMT/BSMT lineage in angiosperms. This lineage of enzymes appears to be ancient because they are found in both rosid and asterid species. The phylogeny indicates that a duplication early in the history of the Solanaceae resulted in separate SAMT and BSMT lineages of enzymes so that all species appear to have at least one of each. Within the BSMT lineage, a subsequent gene duplication event appears to have given rise to two copies of *BSMT* in all *Nicotiana* species, *BSMT 1* and *BSMT 2*. All sequences that have been functionally characterized have been labeled by enzyme name. Unlabeled sequences are enzymatically uncharacterized cDNAs or ESTs. Bootstrap proportions of 70 or greater and posterior probabilities >0.95 are shown for each node

(Koo et al. 2007), and once in the circled lineage of SAMT and Solanaceae BSMT (Fig. 2a). Alternatively, it may be that SA and BA methylation is ancestral in angiosperms and that it only arose once, with specialization to other substrates occurring later in other gene family members. It should be noted that bootstrap support for the separation of the *Arabidopsis BSMT*, *Antirrhinum BAMT*, and rice *BSMT* lineages is not high, so their positions could change somewhat relative to each other with further study of additional sequences from a diversity of angiosperms.

A more detailed view of *SAMT*/*BSMT* phylogeny within Solanaceae reveals at least two duplication events in the history of the gene family (Fig. 2b). There appears to have been one duplication in the ancestor of the family such that all descendants now possess at least one copy of *SAMT* and one of *BSMT*. Within the *BSMT* lineage, a second more recent duplication event appears to have occurred only within *Nicotiana* because two *BSMT* sequences are found in multiple species (Fig. 2b). Although it is expected that allopolyploid species like *N. suaveolens* and *N. gossei* would have two homologous *BSMT* sequences, one from each parental genome involved in its hybrid origin, the presence of two loci in the diploid taxon, *N. alata*, suggests instead that a duplication event occurred early in the history of the genus. It should be noted that although the *Petunia hybrida* sequence is named *BSMT* (Negre et al. 2003), it is clearly orthologous and functionally similar to the *SAMT* sequences found throughout Solanaceae.

Because gene duplications provide opportunities for sub- and neofunctionalization, we investigated the enzymatic properties and expression patterns of these enzymes in detail.

Biochemical characterization of *Nicotiana* SAMT, BSMT and NAMT

To elucidate the biochemical features of the newly isolated carboxyl methyltransferases from *Nicotiana*, the coding sequences were cloned into the pET 101/D-TOPO and pET SUMO expression vectors. In preliminary analyses of enzyme activity, we supplied BA, SA or NA as substrates to the *E. coli* cultures as in Ross et al. (1999). GC–MS analysis of the hexane extracts showed distinct production of MeSA and MeBA for the *N. suaveolens*, *N. alata*, and *N. sylvestris* enzymes, and MeNA in the case of the *N. gossei* enzyme (Supplemental Fig. S2). Subsequently, we over-expressed the His-tagged proteins in *E. coli* HMS174 (DE3), purified them by Ni–NTA affinity chromatography, and analysed the purifications by SDS–PAGE (Fig. S3). Enriched preparations of proteins with apparent molecular masses ranging from 40 to 60 kD were obtained. The differences in protein size were a result of the different expression vectors used as well as the inherent variability of the coding sequences.

The purified *Nicotiana* carboxyl methyltransferases were tested with eighteen substrates, including several BA and cinnamic acid derivatives and jasmonic acid. The relative enzyme activities are summarized in Table 2. The maximum activities of the SAMTs ranged from 2 to 116 pkat/mg protein and showed highest relative methylation activity with SA (100%) and much less activity with BA. All three SAMT enzymes possess greater activity with the doubly hydroxylated substrates 2,3-dihydroxy BA and 2,5-dihydroxy BA than with BA. The N.ala.SAMT enzyme is somewhat different from SAMT of the other two species because of its higher relative enzymatic activities with BA and other ortho-hydroxylated BA derivatives (20–60% relative activity). For these enzymes the 3- and 4-hydroxylated BA derivatives were not effectively converted substrates.

The isolated BSMT enzymes preferred BA over SA as a substrate and are therefore at the biochemical level

Table 2 Relative *Nicotiana* SAMT, BSMT and NAMT enzyme activities with various substrates

	N.sua.SAMT	N.ala.SAMT	N.syl.SAMT	N.sua.BSMT2	N.ala.BSMT2	N.syl.BSMT2	N.gos.NAMT
Salicylic acid	**100**	**100**	**100**	20.65	13.41	3.25	20.7
Benzoic acid	5.26	37.82	3.87	81.89	66.18	**100**	0.9
3-Hydroxybenzoic acid	0.63	7.95	0.54	5.0	0.91	86.51	0.5
4-Hydroxybenzoic acid	0.07	0.17	0.04	2.91	0.24	13.66	0
2,3-Dihydroxybenzoic acid	21.67	52.5	14.4	4.97	3.44	10.29	0.6
2,4-Dihydroxybenzoic acid	3.15	21.94	2.27	0.9	0.45	0.43	0
2,5-Dihydroxybenzoic acid	14.7	61.27	9.38	0.65	0.57	1.67	0
2,6-Dihydroxybenzoic acid	0.31	0.31	0.34	0.2	0.21	0.49	0
3,4-Dihydroxybenzoic acid	0.02	0.22	0.05	0.7	0.06	26.36	0
3,5-Dihydroxybenzoic acid	0.08	0.67	0.05	0.26	0.21	13.7	0
Cinnamic acid	0.02	0.33	0.03	0.48	0.07	2.28	0.8
o-Coumaric acid	0.07	0.76	0.13	0.32	0.32	4.0	0
m-Coumaric acid	0	0.03	0.05	0.22	0.11	3.26	0.5
p-Coumaric acid	0.02	0.04	0.1	0.23	0.1	0.94	0
o-Anisic acid	0.18	1.64	0.12	**100**	**100**	22.93	0
Anthranilic acid	1.55	8.48	1.55	18.27	26.21	ND	1.6
Jasmonic acid	0.04	0	0.3	2.85	0.36	1.0	0
Nicotinic acid	ND	ND	ND	*	*	*	**100**
Highest enzyme activity with favoured substrate (pkat/mg enzyme)	**116**	**2.3**	**19.7**	**7.5**	**3.5**	**1.6**	**0.6**

To Ni–NTA purified enzymes 1 mM substrate were added. 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

* No methylation of NA in *E. coli* extracts (Supplemental Fig. S2); The N.sua.BSMT1-1(not shown) was previously characterised by Pott et al. (2004)

Bold indicates substrate revealing 100% relative enzyme activity, and highest specific enzyme activity in pkat/mg

substantially different from the SAMT enzymes. The enzyme activities of the BSMTs range from 1.6 to 7.5 pkat/mg protein. Interestingly, the N.sua.- and N.ala.BSMT2 showed the highest activity with 2-methoxy BA (o-anisic acid), which was also an effective substrate for the N.sua.BSMT1-1 (Pott et al. 2004). The N.syl.BSMT2 has a different pattern because its activity with *o*-anisic acid is only ~20% of that of BA; however, it exhibits relatively high activity with all 3-hydroxy BA substrates tested. All BSMT2s possessed relatively high enzyme activites with anthranilic acid. Anthranilic acid is also a very good substrate of the floral N.sua.BSMT1-1 with 92% relative activity (Feike and Piechulla unpublished). Overall, these enzymes had low activity with 2-hydroxylated substrates.

The activity profile of the BSMT-like enzyme from *N. gossei* was notably different from all other enzymes, in spite of its high degree of sequence identity to the florally-expressed N.sua.BSMT1-1 (Fig. 2b; Table S1). The *N. gossei* enzyme was highly specific for nicotinic acid, and only SA was otherwise methylated at an appreciable level (20.7% relative activity; Table 2). In contrast, N.sua.BSMT1-1 showed only 1.8% relative methylation activity with NA (Feike and Piechulla, unpublished). From a biochemical point of view, the N.gos.NAMT is highly divergent from SAMTs or BSMTs and it was therefore named Nicotinic acid carboxyl methyltransferase (NAMT) to indicate its specificity for NA and the fact that it was isolated from tissues that emit MeNA (Fig. 1). This enzymatic result adds another function for methyltransferases on this branch of the SABATH family of enzymes and further demonstrates that sequence comparison alone is not sufficient to delineate the function and role of many enzymes involved in plant specialized metabolism.

Expression analysis of *Nicotiana SAMT* and *BSMT* genes

To examine expression patterns of the *SAMT* and *BSMT* genes and to document further divergence between members of the carboxyl methyltransferase gene family in *Nicotiana* species, qualitative RT-PCR reactions were carried out with RNA extracts from whole flowers, leaves, stems and roots harvested at different time points during the day. The newly isolated *Nicotiana BSMT*s and *SAMT*s showed distinct expression patterns (Fig. 3). The *SAMT*s from *N. suaveolens* and *N. sylvestris* were only expressed in roots, while the *N.ala.SAMT* transcripts were detected in all organs. *Nicotiana BSMT2* transcripts were expressed in leaves, stems and flowers, but at lower or undetectable levels in roots. Interestingly, expression of the *N.sua. BSMT1-2* exhibits the same pattern as *N.ala.BSMT2* and *N.syl.BSMT2*, indicating that these paralogous BSMT enzymes may have similar functions. The *N.sua.BSMT1-1*

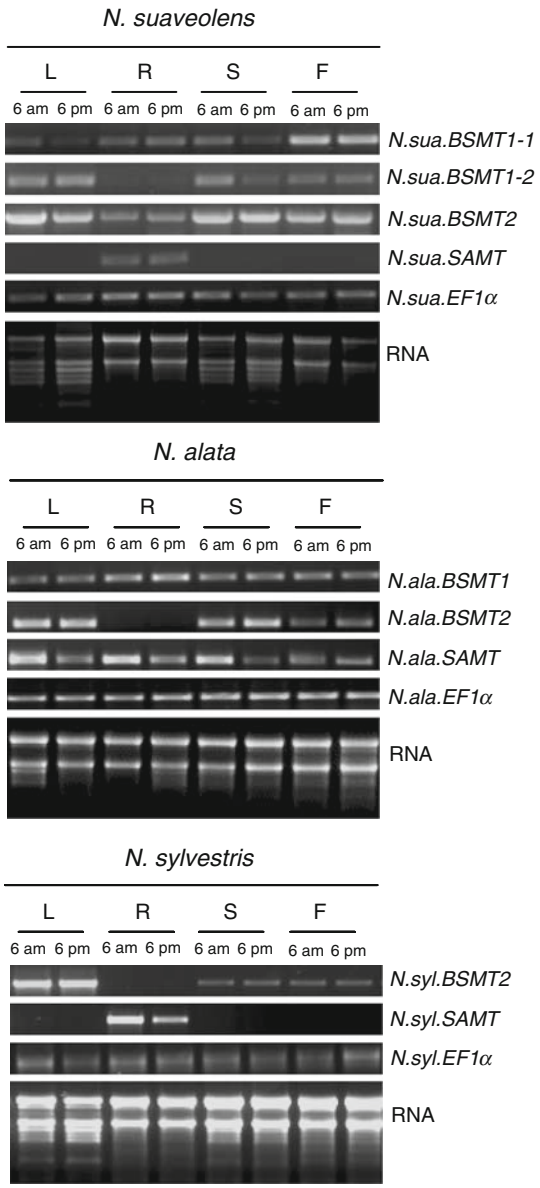

Fig. 3 Determination of expression of *Nicotiana SAMT* and *BSMT* genes via qualitative RT-PCR. Plant material was harvested from leaf (*L*), root (*R*), stem (*S*) and flowers (*F*) at 6 am and 6 pm. 2 µg of total RNA was used for RT-PCR reactions. Translation elongation factor 1α (*EF1α*) was used as an external control

and *N.ala.BSMT1* genes are expressed in all organs, but the former shows highest expression in flower tissue. This result is consistent with the original isolation of this gene from floral tissue and expression patterns documented by Northern blot analysis previously (Pott et al. 2004). The *N.ala.BSMT1* gene seems to be constitutively expressed and may have a general role in the plant tissues.

SAMT from *N. alata* and *N. sylvestris* was inducible in leaves by SA treatment as compared to controls (Fig. 4). BA treatment appeared to have no effect on *SAMT* expression, and *BSMT2* expression was not induced by SA

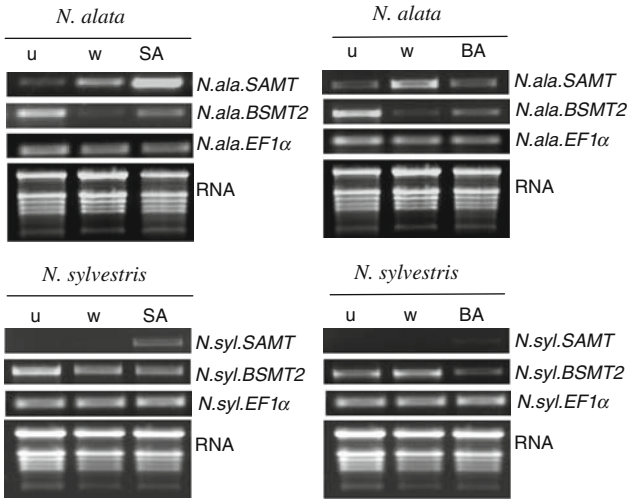

Fig. 4 Determination of expression levels of *Nicotiana SAMT* and *BSMT* genes after induction with SA or BA. Leaf discs were incubated for 24 h in 5 mM salicylic acid (*SA*), 5 mM benzoic acid (*BA*) and pure water (*w*), respectively prior to RNA extraction. As a control, untreated leaves were utilized (*u*). Two µg of total RNA was used for RT-PCR reactions. Translation elongation factor 1α (*EF1α*) served as an external control

or BA treatment in these two species. Overall, these results give a first hint in which plant organ *SAMT* and *BSMT* genes are expressed and how their expression is affected by various factors. Although similar amounts of RNAs were used for RT-PCR, as indicated by the internal control of EF1α, small differences in expression intensities should not be over-interpreted.

In silico modelling of the substrate binding sites of N.sua.SAMT, N.sua.BSMT2 and N.gos.NAMT

The three dimensional structures of N.sua.SAMT and N.sua.BSMT2 as well as N.gos.NAMT were elucidated by in silico modelling (Fig. 5). The overall structures of these enzymes are similar to the structure of *C. breweri* SAMT (Zubieta et al. 2003). All enzyme monomers investigated here possess a globular domain containing various ß-sheets and α-helices as well as an α-helical cap. The globular domains of N.sua.SAMT, N.sua.BSMT2 and N.gos.NAMT interact with the methyl donor and show overall structural similarity (Fig. 5a–c). In contrast, the protein domains composing the α-helical cap of the enzymes exhibit more substantial structural differences. The in silico modelling of N.sua.BSMT2 shows that the 32 amino acid C-terminal insertion starting at Thr-298 (according to N.sua.BSMT1-1) is located within the α-helical cap between ß-fold 6 and helix 8 (Fig. 5a and Fig. S1). In silico modelling gave no reliable structure for that region and therefore it is shown as a loop (Fig. 5a). The functional significance of these structural divergences is unknown.

Substrate acceptance by an enzyme is an intrinsic feature due to the amino acid sequence of the protein, particularly in the active pocket. The carboxyl methyltransferases possess two binding sites, one for the methyl donor *S*-adenosyl-L-methionine and the other for the methyl acceptor molecule. The amino acids of the SAM binding site are highly conserved in the SAMTs and BSMTs from *Nicotiana* (Table 3A). All putative SAM binding residues are identical to those determined from SAMT isolated from *C. breweri*, except for Lys-10 which is replaced by Asn in all *Nicotiana* carboxyl methyltransferases (Table 3A). A comparison of the SA binding sites of SAMT shows that despite 125 million years of divergence, SA binding sites from *Nicotiana* and *C. breweri* are identical. In contrast, the substrate binding pocket of BSMT2 and N.gos.NAMT are more variable and divergent as compared to SAMT (Table 3B). Within the active site of BSMT2, Tyr-147, Trp-151, Leu-210, Tyr-255 and Phe-347 were conserved while variation is exhibited at positions 25, 150, 225, 226, 308 and 311 relative to SAMT (amino acids according to C.b.SAMT sequence). Because only positions 150, 225 and 308 are substituted in all BSMT2, it is likely that much of the shared biochemical divergence noted in Table 2 is explained by these replacements. In particular, Met-150 and Met-308 of SAMT that keep SA in a favourable position for methylation (Zubieta et al. 2003) are replaced in BSMT2s by His or Gln at position 150 and Leu at position 308. Ile at position 225 is replaced by the smaller, nonpolar amino acid Val in all BSMT2 sequences but it is unclear what role this residue plays in substrate binding or catalysis. One apparent collective impact of the substitutions of the smaller amino acids Val-233, Leu-234 and Leu-336 in BSMT2 relative to SAMT is to provide a larger active pocket volume.

In silico modelling showed that the radical replacement of the nonpolar Met-156 (Met150 in C.b.SAMT) by the basic His and the Met-308 by Leu may prevent the formation of a molecular (Met-Met) clamp important for tight binding of the substrate in SAMTs as already described by Zubieta et al. 2003 (Fig. 5d–f). While SA is tightly surrounded by the amino acids of the active pocket of the N.sua.SAMT, the amino acids in the N.sua.BSMT2 enzyme are not in close vicinity to the substrate; the effect of this appears to account for the reduced specificity for SA observed (Fig. 5d and e; Table 3). Compared to the N.sua.BSMT1-1, the N.sua.BSMT2 possesses a lower substrate spectrum. But while *o*-anisic acid is the third best used substrate from N.sua.BSMT1-1, it is the favoured component of N.sua.BSMT2. It is thought that the ring nitrogen of His158 could form a hydrogen bond with the 2-methoxy group of *o*-anisic acid as already described by Pott et al. 2004. A similar role is conceivable for the Gln in the corresponding position of N.ala.BSMT2. Whereas substrate specificity of N.sua.BSMT2 is very similar to

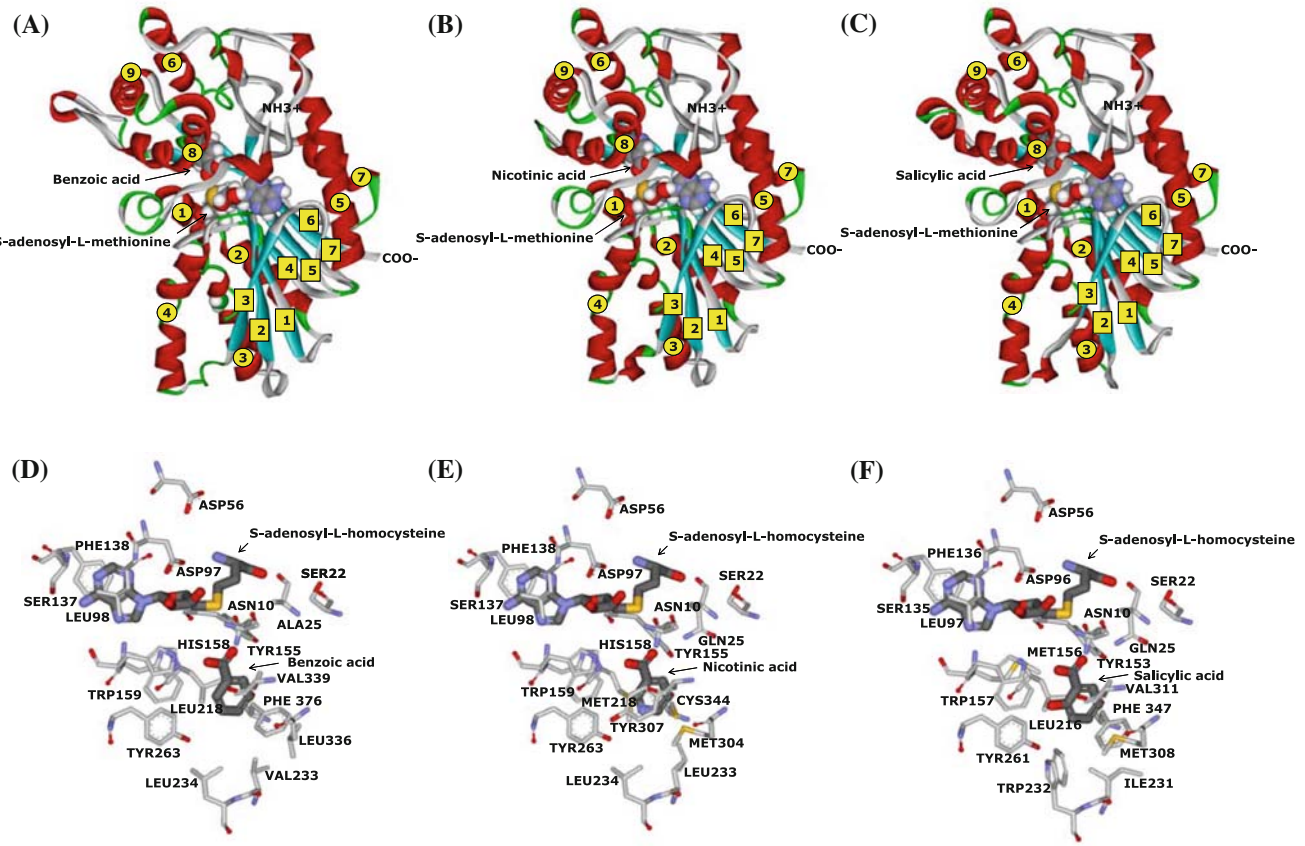

Fig. 5 Structure models of N.sua.BSMT2, N.gos.NAMT and N.sua.-SAMT protein monomers and active sites. Complete protein monomer of N.sua.BSMT2 (**a**), N.gos.NAMT (**b**) and N.sua.SAMT (**c**), respectively. Helices and ß-strands are numbered. *Helices* are shown in red and are indicated with circles. *Folds* are shown in *blue* and are indicated with *squares*. Three dimensional view of active sites from N.sua.BSMT2 (**d**), N.gos.NAMT (**e**) and N.sua.SAMT (**f**). The *blue colour* of the *sticks* indicate: nitrogen atoms; *red* oxygen atoms and *yellow* sulphur. For all models their favoured substrates are indicated

N.ala.BSMT2, it markedly differs from N.syl.BSMT2. In addition to BA, N.syl.BSMT2 prefers 3-hydroxy BA as a substrate likely caused by a Trp at position 234 that potentially hydrogen bonds with the 3-hydroxy group. Compared to BSMT2s, the BSMT1-1 from *N. suaveolens* differs from those enzymes at six of the active site residues. Some of these changes probably account for the lower substrate specificity of N.sua.BSMT1-1.

Structural modeling of NA in the active site of N.gos.-NAMT reveals an orientation and set of interactions that are similar to those of *C. breweri* SAMT and SA (Zubieta et al. 2003). Hydrogen bonding occurs between Gln25 and Trp159 and the carboxylate moiety of NA to form a tether that positions it for transmethylation. Additional hydrogen bonding interactions occur between His158 and the carboxyl group. However, the substituted Tyr307 seems to be particularly important in forming hydrogen bonds with the nitrogen of nicotinic acid due to its proximity to the substrate. N.gos.NAMT is the only enzyme in the SABATH family with Tyr in this position and thus it seems likely that this replacement is important for the specialization to NA. A comparison of N.gos.NAMT to its close relative N.sua.BSMT1-1 reveals that only four active site residues differ between them which may account for their enzymatic divergence (Table 2; Table 3B). While a substitution of Leu for Ile at position 233 is unlikely to account for the divergence, the charge-changing replacements at positions 307 and 344 (311 and 347 in C.b.SAMT, respectively) are more likely candidates (Fig. 5f).

Ancestral state estimation

Figure 6 shows ancestral state estimates for three nodes in the phylogeny of Solanaceae SAMT and BSMT based on experimentally determined enzyme activity data from Table 2. At node A, the ancestor of all SAMT and BSMT likely exhibited a fivefold higher preference for methylation of SA over BA as indicated by the estimated ancestral MeSA:MeBA. This preference for SA did not change significantly along the branch between node A and C because the estimated ratio of MeSA:MeBA is similar; however, a nearly fivefold reduction in preference for SA

Table 3 Amino acids of substrate binding sites of *Nicotiana* carboxyl methyltransferases

C.b.SAMT	N.sua.SAMT	N.ala.SAMT	N.syl.SAMT	N.sua.BSMT2	N.ala.BSMT2	N.syl.BSMT2	N.sua.BSMT1-1[a]	N.gos.NAMT
(A)								
Lys 10	**Asn 10**	**Asn**	**Asn**	**Asn 10**	**Asn**	**Asn**	**Asn**	**Asn 10**
Ser 22	Ser 22	Ser	Ser	Ser 22	Ser	Ser	Ser	Ser 22
Asp 57	Asp 56	Asp	Asp	Asp 56	Asp	Asp	Asp	Asp 56
Asp 98	Asp 96	Asp	Asp	Asp 97	Asp	Asp	Asp	Asp 97
Leu 99	Leu 97	Leu	Leu	Leu 98	Leu	Leu	Leu	Leu 98
Ser 129	Ser 135	Ser	Ser	Ser 137	Ser	Ser	Ser	Ser 137
Phe 130	Phe 136	Phe	Phe	Phe 138	Phe	Phe	Phe	Phe 138
(B)								
Gln 25	Gln 25	Gln	Gln	**Ala 25**	**Ala**	Gln	Gln	Gln 25
Tyr 147	Tyr 153	Tyr	Tyr	Tyr 155	Tyr	Tyr	**Phe**	Tyr 155
Met 150	Met 156	Met	Met	**His 158**	**Gln**	**His**	**His**	**His 158**
Trp 151	Trp 157	Trp	Trp	Trp 159	Trp	Trp	Trp	Trp 159
Leu 210	Leu 216	Leu	Leu	Leu 218	Leu	Leu	**Met**	**Met 218**
Ile 225	Ile 231	Ile	Ile	**Val 233**	**Val**	**Val**	Ile	**Leu 233**
Trp 226	Trp 232	Trp	Trp	**Leu 234**	**Leu**	Trp	**Leu**	**Leu 234**
Tyr 255	Tyr 261	Tyr	Tyr	Tyr 263	Tyr	Tyr	Tyr	Tyr 263
Met 308	Met 308	Met	Met	**Leu 336**	**Leu**	**Leu**	Met	Met 304
Val 311	Val 311	Val	Val	Val 339	**Leu**	Val	**Phe**	**Tyr 307**
Phe 347	Phe 347	Phe	Phe	Phe 376	Phe	Phe	**Ser**	**Cys 344**

The comparison of amino acids with importance for substrate binding is based on the active site of *C. breweri* SAMT (Zubieta et al. 2003). *A* Amino acids that are required for binding of SAM. *B* Amino acids with potential importance for binding of SA, BA and NA, respectively. Altered amino acids in comparison to the C.b.SAMT are shown in bold

[a] Pott et al. (2004)

relative to BA is inferred to have changed along the branch separating node A from B. At node B, the ancestor of all BSMT likely had nearly equal preference for SA and BA as indicated by the estimated ratio of MeSA to MeBA. This nearly fivefold reduction in the estimated MeSA:MeBA appears to have occurred along the same branch in which the important active site residue (Met 156) governing preference for methylation of SA by SAMT (Barkman et al. 2007) evolved to His (Fig. 6).

Discussion

Enzymatic divergence of orthologs

The approach taken in this study was to compare orthologous enzyme evolution as a result of divergence among closely related species as well as paralogous enzyme divergence as a result of gene duplication. Because the *Nicotiana* species diverged long after the duplication of SAMT and BSMT, any difference between their orthologous enzymes is most likely attributable to speciation. Within this comparative framework, we investigated evolutionary divergence at the level of gene expression, protein structure and enzyme activity. In terms of enzyme activity, SAMT did not vary substantially between species indicating that the preference for SA of the enzyme has not changed as species have diverged. The top four substrates for each SAMT were the same (SA, 2,3-dihydroxyBA, 2,5-dihydroxyBA and BA, respectively) suggesting that selection has largely maintained ancestral activity within *Nicotiana*. The active sites and substrate preferences of SAMTs from other species like *Clarkia breweri*, *Antirrhinum majus*, *Atropa belladonna*, *Datura wrightii* or BSMT from *Petunia hybrida* are highly similar (Fukami et al. 2002; Negre et al. 2002; Negre et al. 2003; Barkman et al. 2007). Only SAMT from *Stephanotis floribunda* is an exception since it differs in four amino acids within the active site (Pott et al. 2004; Effmert et al. 2005). These amino acid alterations of the *S. floribunda* SAMT seem to contribute to the lower substrate specificity of this enzyme. Together these are all indications that there is only a small range of variation within the active site of the SAMTs allowing for the effective binding of SA and simultaneous exclusion of other structurally similar substrates, particularly, BA.

On the other hand, there appears to have been divergence of BSMT enzyme activity among *Nicotiana* species. The top four substrates for N.sua.BSMT2 and N.ala.BSMT2 were 2-methoxyBA (*o*-anisic acid), BA, SA

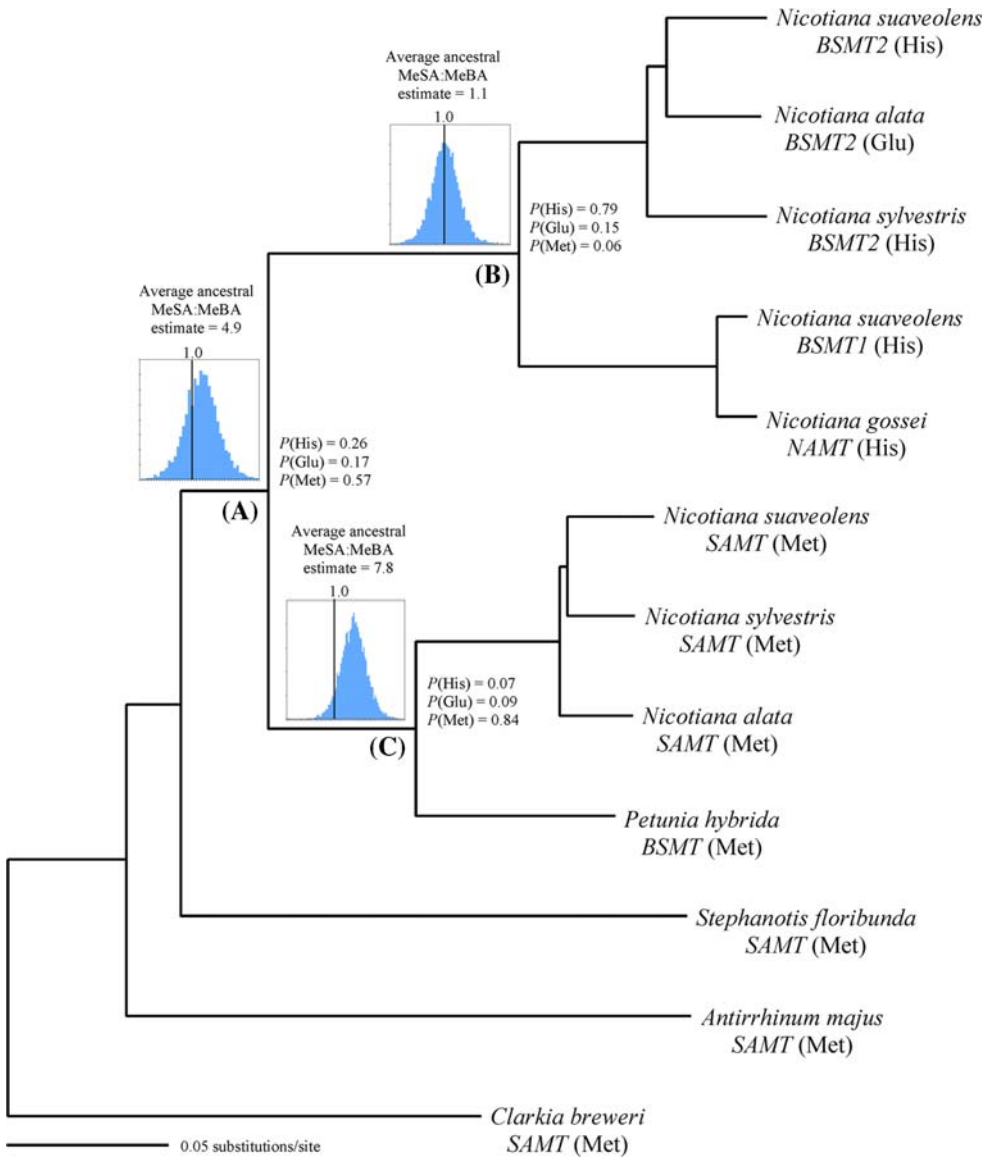

Fig. 6 Ancestral state estimates for the ratio of MeSA:MeBA produced by SAMT and BSMT and amino acid position 156/158 for Solanaceae SAMT and BSMT, respectively. Histograms at nodes **A–C** show the distribution of estimated ancestral states for the ratio of MeSA:MeBA based on the activities shown in Table 2 for modern-day enzymes. Probabilities of ancestral amino acids are shown at nodes A-C for His, Glu and Met. Node **A** shows that the ancestor of all BSMT and SAMT in Solanaceae likely exhibited a fivefold preference for methylation of SA as compared to BA (as indicated by the ratio of the products of these substrates). This ancestor also most likely possessed Met at one of the key residues previously shown to control enzyme preference for SA as compared to BA (Barkman et al. 2007). Node **B** shows that the ancestor of BSMT1 and BSMT2 likely exhibited little preference for SA over BA and that this activity is associated with the presence of His at the active site residue which controls preference for SA. Thus, the nearly fivefold reduction of ancestral BSMT preference for SA was likely concomitant with the active site residue change from Met to His along the branch separating node **A** from **B**. Node **C** shows that the ancestor of SAMT in *Nicotiana* likely retained the high preference for SA over BA and that the active site most likely remained Met along the branch separating node **A** from **C**

and anthranilic acid whereas N.syl.BSMT2 showed highest activity with BA, 3-hydroxyBA, 3,4-dihydroxyBA and 2-methoxyBA. The high activities of N.syl.BSMT2 with 3-hydroxyBA and of N.sua.BSMT1-1 with anthranilic acid indicate fundamental evolutionary changes to these enzymes; however, the importance of these enzymatic divergences for plant fitness remains unknown. *Nicotiana suaveolens* does emit low levels of methyl anthranilate (MeAA) from its flowers so perhaps the enzyme divergence enhances pollinator attraction.

Although only minimal orthologous enzyme divergence appears to have occurred among SAMT and BSMT of *N. alata*, *N. suaveolens*, and *N. sylvestris*, NAMT, a close ortholog of BSMT1-1, has evolved substantially in terms of

substrate preference. The phylogenetic analysis indicates that the *N.gos.NAMT* arose recently from an ancestral *BSMT1* gene (Fig. 2b) that only had minimal activity with NA. *Nicotiana gossei* and *N. suaveolens* are closely related Australian species whose flowers are very similar in morphology, differing mostly in floral tube length. They both express *BSMT1-1* orthologs at high levels in petal tissue as compared to leaves (Figs. 1b, 3). Yet, their enzyme properties differ substantially because N.sua.BSMT1-1 preferentially methylates BA and has only low activity with NA while N.gos.NAMT prefers NA above all others tested and catalyzes the formation of MeBA only at very low levels (Table 2). Determining the recent evolutionary changes allowing NAMT to diverge in enzyme activity will require site-directed mutagenesis studies aimed at determining the importance of the few amino acids (Leu 233, Tyr 307 and Cys 344; Fig. 5 and Table 3) that differ between it and N.sua.BSMT1-1. Although it is not possible to determine if speciation of *N. gossei* was promoted by this novel enzyme activity, it is clear that activity with NA evolved recently because of the recent divergence of *N. gossei* and *N. suaveolens.*

Expression divergence of orthologs

In the case of *SAMT*, there has been some degree of evolutionary change in gene expression patterns because *N. ala.SAMT* is expressed in all tissues whereas it is only expressed in roots of *N. suaveolens* and *N. sylvestris*. In the context of the phylogeny of *Nicotiana*, root-specific expression may be ancestral; however, the posterior probability of this ancestral state estimate is quite low ($P = 0.46$; data not shown). Given that gene expression changes likely evolve rapidly, determination of *SAMT* and *BSMT* expression patterns of more Solanaceae species is necessary in order to more confidently understand ancestral gene expression patterns. To our knowledge no one has ever reported volatile production from *Nicotiana* roots but our expression results indicate that these organs should be investigated for the presence of MeSA. It is not clear what the role of SAMT is in *N. alata* vegetative tissues because neither MeSA nor MeBA has been detected from its leaves or stems. Like other species, *SAMT* in *N. alata* and *N. sylvestris* appears to experience increased expression in response to SA treatment (Martins and Barkman 2005) making this a conserved, inducible leaf response indicative of a role in pathogen defense.

Patterns of *BSMT2* gene expression are largely conserved among *Nicotiana* species with expression highest in leaves and low or absent in roots and it is not induced by any of the treatments administered in this study. A role of *BSMT2* in leaf tissue is unclear except for potentially in *N. suaveolens* which does emit MeSA from untreated leaves (Raguso et al. 2003). Like *BSMT2*, *BSMT1* is expressed in multiple tissues including flowers. At least one role for these genes is for floral scent production and all three species studied do emit MeSA and/or MeBA from flowers. Inferring the directionality of gene expression evolution for *BSMT* is limited by our knowledge of the number of loci each species possesses. However, the patterns shown in Fig. 3 suggest that the ancestor of *Nicotiana* probably expressed *BSMT* throughout the plant, except perhaps, in roots.

Enzymatic divergence of paralogs

In contrast to enzyme evolution due to species divergences, the largest evolutionary changes noted in this study appear to be tied to gene duplications. Gene duplication in Solanaceae appears to have provided an opportunity for neofunctionalization, whereby SAMT appears to have maintained ancestral function (high level of activity with SA but not BA) and BSMT has evolved complementary enzyme function: a high level of activity with BA and other substrates, but lower activity with SA. This enzymatic divergence likely occurred early in the history of the Solanaceae long before the *Nicotiana* species evolved because ancestral state estimates indicate that the change in substrate preference occurred along the branch leading to the ancestor of all BSMTs (Fig. 6). In particular, estimates suggest a nearly fivefold reduction in the preference for SA evolved in the ancestor of BSMT from the preduplication enzyme. Divergence among the two BSMT-type enzymes (1 and 2) as a result of recent duplication within *Nicotiana* is not clear due to a lack of functionally characterized BSMT1s. The basis of the changes in enzyme activity appear to be the result of amino acid replacements affecting the active pocket. It is possible that the adaptive conflict model (Hughes 1994) explains our data instead of the neofunctionalization or subfunctionalization models. Future tests of historical patterns of selection will allow discrimination between these possibilities.

The phylogeny of SAMT and BSMT enzymes within Solanaceae, and the SABATH family in general, clearly indicates that most active site amino acid changes have occurred in the BSMT lineage while the SAMT lineage has apparently been under selection to maintain ancestral enzyme activity (5-sevenfold preference for SA over BA; Fig. 6; Table 3). In particular, Met 150 (according to C.b.SAMT) has undergone an evolutionary reversal in BSMT to the ancient ancestral residue, His, found in nearly all other characterized SABATH enzymes (Fig. 6). The evolutionary reversal to His (or Gln) at position 150 may have promoted specialization to other structurally related substrates to SA, like BA, 2-methoxyBA, and anthranilic acid. However, it should be noted that His also exists in other SAM-dependent carboxyl methyltransferases that use

jasmonic acid, gibberellic acid or indole-3-acetic acid as substrates and therefore is not a unique feature for BSMT enzymes (Seo et al. 2001; Qin et al. 2005; Varbanova et al. 2007). Rather, it appears that the Met-150 is a special feature of SAMTs that likely evolved in the ancestor of all angiosperms because nearly every sequence shown in Fig. 2b has Met at position 150 except for the Solanaceae BSMT. The importance of Met for SAMT results in a preference for SA as opposed to other substrates like BA as shown experimentally (Zubieta et al. 2003; Barkman et al. 2007) and by our activity results (Table 2). The evolution from His to Met or vice versa is a complicated set of mutations involving three changes of the single codon. The intermediate codon for Gln (CAG) may provide a functional intermediate because the Gln-containing N.ala.BSMT2 appears to be enzymatically comparable to the His-containing N.sua.BSMT2. However, a single inversion could result in the change between His and Met as well because the codons are reverse complements of each other.

The *Nicotiana* BSMTs methylated 2-methoxy BA (*o*-anisic acid) as well as, or better than, BA. Hitherto it is unknown whether *o*-anisic acid embodies a natural substrate for the BSMTs, since emission of methyl anisic acid has not been reported in *Nicotiana* although it is known from floral scents of other species (Knudsen et al. 2006). The *Nicotiana* BSMTs also methylated 3-hydroxyBA better than most other substrates tested. Furthermore, the BSMT-type enzymes also demonstrated moderate to high methylation activity with anthranilic acid. MeAA is emitted at low levels from *N. suaveolens* flowers (Raguso et al. 2003) making it possible that BSMT1 or 2 is responsible. *Arabidopsis thaliana* BSMT also exhibited high relative activity with 3-hydroxyBA and anthranilic acid in addition to BA (Chen et al. 2003). The fact that these two enzymes evolved independently from each other yet converged to have similar enzyme activities allows for future comparative approaches to dissect the amino acid substitutions resulting in the acquisitions of these properties. It should be noted that the synthesis of MeAA by methylating the carboxyl group of anthranilate in a SAM-dependent reaction has not been shown before. Previously, the formation of MeAA was demonstrated by the reaction of anthraniloyl-coenzyme A and methanol in *Vitis vinifera* (Wang and de Luca 2005). Experimental approaches will be required to determine the relative importance of either mechanism of MeAA production for plant biochemistry.

Expression divergence of paralogs

At the level of gene expression, it appears that there has been some degree of tissue specific complementation that has evolved between *SAMT* and *BSMT*. The gene duplication event leading to the divergence of these enzymes may have resulted in subfunctionalization. It is interesting to note that *SAMT* is largely expressed in roots whereas *BSMT2* is expressed mostly in other tissues besides roots. This expression divergence due to gene duplication or altered gene regulation appears to have promoted a role for *BSMT*, but not *SAMT*, in floral scent production in *Nicotiana suaveolens* and *N. sylvestris*. However, it is clear that other Solanaceae species, including *Petunia hybrida* and *Cestrum nocturnum*, express *SAMT* orthologs in petals as the primary enzyme producing MeBA and MeSA (Negre et al. 2003; Martins et al. 2007). Thus, it appears that duplicate gene expression patterns evolve rapidly making it difficult to ascribe general functions to one or the other enzyme in this family. The transcripts of *BSMT* and *SAMT* were found in plant organs other than flowers suggesting that they possess other functions than just pollinator attraction. While a root-specific function for *SAMT* remains obscure, an obvious potential role for both enzymes is in pathogen defense and the development of SAR (Koo et al. 2007). The presence of *BSMT* transcripts in uninfected leaves and the increase of *SAMT* expression in response to SA treatment of leaves suggests roles for both genes in the biosynthesis of MeSA in infected leaves as an endogenous signal transmitted to uninfected plant parts. Silencing studies in *N. tabacum*, suggested a role for a *BSMT* in SAR in response to tobacco mosaic virus infection (Park et al. 2007). Future expression and enzymatic studies of that enzyme and others should help further clarify the evolution of *SAMT*/*BSMT* function in Solanaceae.

Correlation of phenotype and enzyme characteristics

We relied on a correlative approach in this study to relate patterns of floral scent emission to gene/enzyme data. *Nicotiana suaveolens*, *N. alata*, *N. sylvestris*, and *N. gossei* are known to produce one or more of the volatile esters MeBA, MeSA, MeNA, and MeAA in flowers (Raguso et al. 2003). The enzyme activity and expression results all point to roles for BSMT, NAMT, and to a lesser extent, SAMT in the production of these volatiles in planta. Our results suggest that for *N. sylvestris*, only BSMT2 is likely involved in floral scent emission of MeBA. For *N. suaveolens* emission of MeBA, MeSA, MeNA, and MeAA at varying levels is difficult to correlate with the activity of any one enzyme because our studies showed the expression of at least three different *BSMT* genes within flowers of *N. suaveolens*. The participation of N.sua.BSMT1-1 in floral scent production was already shown by Pott et al. (2004) and the contribution of the newly isolated N.sua.BSMT1-2 and N.sua.BSMT2 may now be assumed. *Nicotiana alata* expressed both *SAMT* and *BSMT* in petals making it possible that both enzymes contribute to floral MeBA/MeSA emission. Finally, we have shown enzyme activity and

expression results consistent with a role of NAMT in MeNA production in *N. gossei*. Because of the overlapping expression patterns and enzyme activities, it is difficult to firmly establish the role of any one enzyme in volatile production in these *Nicotiana* species. However, silencing studies may be challenging due to the high level of sequence identity among the *BSMT/NAMT* sequences we have isolated. Furthermore, we also acknowledge that methyltransferase activity alone does not entirely account for the fragrance phenotypes. As was shown in *Petunia*, *Stephanotis* and *N. suaveolens*, available substrate pools may dictate the quality and quantity of floral volatile production to a larger degree than transcript abundance or enzyme substrate preference (Kolosova et al. 2001; Pott et al. 2004; Effmert et al. 2005).

Phylogenetic patterns of SABATH gene family evolution

The phylogeny of Fig. 2 implies that like IAMT (Zhao et al. 2008), SAMT is an ancient lineage of SABATH methyltransferases. At this point, it is not possible to determine which activity may be older within the gene family but functional characterization of SABATH enzymes from gymnosperms could provide valuable information in this regard. Recently, it was shown that an IAMT ortholog from *Picea* can catalyze methyl transfer to indole-3-acetic acid thereby extending the origin of this enzymatic function to the ancestor of seed plants (Zhao et al. 2009). However, the complex patterns of gene family member birth and death will ultimately make the inference of original protein family activity difficult. The phylogenetic patterns also indicate that BA and SA methylating enzymes do not form one monophyletic clade. Instead there are four lineages of enzymes that can form MeSA and/or MeBA. While multiple origins of SA or BA methylating ability has been suggested previously (D'Auria et al. 2003; Zhao et al. 2008), what has not been considered is that it is possible that these were the ancestral substrates for the entire family, or part of it. If this was the case, then the ability to methylate these substrates only evolved once during SABATH family evolution.

The phylogenetic approach used in this study allowed dissection of the potential roles of gene duplication and species divergence in enzyme evolution. The use of the same homologs from a minimum of three close relatives allowed for estimates of ancestral conditions and therefore inference of the directionality of evolutionary changes in enzyme activity and expression. Finally, while gene duplication may promote substantial enzyme divergence in terms of activity and expression patterns, it is clear from this study that species-specific evolutionary changes can be significant. In the case of NAMT from *N. gossei*, highly divergent enzyme activity evolved from a BSMT-like ancestral enzyme.

Acknowledgment The authors thank Anja Zeißler for helping with the gene expression experiments. We are grateful to Robert A. Raguso (Cornell University) for providing seeds. Talline Martins is thanked for helpful comments on earlier drafts of the manuscript. This work was supported by grants of the DFG (Pi 153/22 to B.P.) and NSF grant DEB 0344496 (to T.J.B).

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 24:1320–1329. doi:10.1093/molbev/msm053

Berenbaum MR (1995) The chemistry of defense: theory and practice. Proc Natl Acad Sci USA 92:2–8. doi:10.1073/pnas.92.1.2

Bradford MM (1976) A rapid and sensitive method for the quantification of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem 72:248–254

Chang S, Puryear J, Cairney J (1993) A simple and efficient method for isolating RNA from pine trees. Plant Mol Biol Report 11:113–116. doi:10.1007/BF02670468

Chase MW, Knapp S, Cox AV, Clarkson JJ, Butsko Y, Joseph J, Savolainen V, Parokonny AS (2003) Molecular systematics, GISH and the origin of hybrid taxa in *Nicotiana* (Solanaceae). Ann Bot 92:107–127. doi:10.1093/aob/mcg087

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 36:577–588. doi:10.1046/j.1365-313X.2003.01902.x

Clarkson JJ, Knapp S, Garcia VF, Olmstead RG, Leitch AR, Chase MW (2004) Phylogenetic relationships in *Nicotiana* (Solanaceae) inferred from multiple plastid DNA regions. Mol Phylogenet Evol 33:75–90. doi:10.1016/j.ympev.2004.05.002

D'Auria JC, Chen F, Pichersky E (2003) The SABATH family of MTs in *Arabidopsis thaliana* and other plant species. Recent Adv Phytochem 37:95–125

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 66:1211–1230. doi:10.1016/j.phytochem.2005.03.031

Firn RD, Jones CG (2000) The evolution of secondary metabolism—a unifying model. Mol Microbiol 37:989–994. doi:10.1046/j.1365-2958.2000.02098.x

Fitch WM (2000) Homology: a personal view on some of the problems. Trends Genet 16:227–231

Fraser AM, Mechaber WL, Hildebrand JG (2003) Electroantennographic and behavioral responses of the sphinx moth *Manduca sexta* to host plant headspace volatiles. J Chem Ecol 29:1813–1833. doi:10.1023/A:1024898127549

Fukami H, Asakura T, Hirano H, Abe K, Shimomura K, Yamakawa T (2002) Salicylic acid carboxyl methyltransferase induced in hairy root cultures of *Atropa belladonna* after treatment with exogeneously added salicylic acid. Plant Cell Physiol 43:1054–1058

Goodspeed TH (1954) The genus *Nicotiana*. Chronica Botanica 16:1–536

Guex N, Peitsch MC (1997) SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis 18:2714–2723. doi:10.1002/elps.1150181505

Hildebrand PW, Goede A, Bauer RA, Gruening B, Ismer J, Michalsky E, Preissner R (2009) SuperLooper—a prediction server for the modeling of loops in globular and membrane proteins. Nucleic Acids Res 37:W571–W574

Hoagland DR, Aronon DL (1938) The water-culture method of growing plants without soil. Calif Agric Exp Stn Circ 374:1–39

Hoballah ME, Stuurman J, Turlings TCJ, Guerin PM, Connétable S, Kuhlemeier C (2005) The composition and timing of flower odor emission by wild *Petunia axillaris* coincide with the antennal perception and nocturnal activity of the pollinator *Manduca sexta*. Planta 222:141–150. doi:10.1007/s00425-005-1506-8

Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogeny. Bioinformatics 17:754–755

Hughes AL (1994) The evolution of functionally novel proteins after gene duplication. Proc R Soc Lond B Biol Sci 256:119–124

Kapteyn J, Qualley AV, Xie Z, Fridman E, Dudareva N, Gang DR (2007) Evolution of cinnamate/p-coumarate carboxyl methyltransferases and their role in the biosynthesis of methylcinnamate. Plant Cell 19:3212–3229. doi:10.1105/tpc.107.054155

Kim YS, Uefuji H, Ogita S, Sano H (2006) Transgenic tobacco plants producing caffeine: a potential new strategy for insect pest control. Transgenic Res 15:667–672. doi:10.1007/s11248-006-9006-6

Knapp S, Chase MW, Clarkson JJ (2004) Nomenclatural changes and new sectional classification in *Nicotiana* (Solanaceae). Taxon 53: 73–82

Knudsen JT, Eriksson R, Gershenzon J, Ståhl B (2006) Diversity and distribution of floral scent. Bot Rev 72:1–120. doi:10.1663/0006-8101(2006)72[1:DADOFS]2.0.CO;2

Kolosova N, Gorenstein N, Kish CM, Dudareva N (2001) Regulation of circadian methyl benzoate emission in diurnally and nocturnally emitting plants. Plant Cell 13:2333–2347. doi:10.1105/tpc.13.10.2333

Koo YJ, Kim MA, Kim EH, Song JT, Jung C, Moon JK, Kim JH, Seo HS, Song SI, Kim JK, Lee JS, Cheong JJ, Choi YD (2007) Overexpression of salicylic acid carboxyl methyltransferase reduces salicylic acid-mediated pathogen resistance in *Arabidopsis thaliana*. Plant Mol Biol 64:1–15. doi:10.1007/s11103-006-9123-x

Loughrin JH, Hamilton-Kemp TR, Andersen RA, Hildebrand DF (1990) Headspace compounds from flowers of *Nicotiana tabacum* and related species. J Agric Food Chem 38:455–460

Martins TR, Barkman TJ (2005) Reconstruction of Solanaceae phylogeny using the nuclear gene SAMT. Syst Bot 30:435–447. doi:10.1600/0363644054223675

Martins TR, Stout JT, Todd SE, Kuipers K, Barkman TJ (2007) Molecular phylogenetic tests of floral scent evolution in the Solanaceae. Acta Hort 745:183–200

Michalsky E, Goede A, Preissner R (2003) Loops in proteins (LIP)—a comprehensive loop database for homology modelling. Protein Eng 16:979

Moore RC, Purugganan MD (2005) The evolutionary dynamics of plant duplicate genes. Curr Opinion Plant Biol 8:122–128

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 382:145–151. doi:10.1006/abbi.2000.2008

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 406:261–270. doi:10.1016/S0003-9861(02)00458-7

Negre F, Kish CM, Boatright J, Underwood B, Shibuya K, Wagner C, Clark DG, Dudareva N (2003) Regulation of methylbenzoate emission after pollination in snapdragon and petunia flowers. Plant Cell 15:2992–3006. doi:10.1105/tpc.016766

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

Nugroho LH, Verpoorte R (2002) Secondary metabolism in tobacco. Plant Cell Tiss Org Cult 68:105–125. doi:10.1023/A:1013853909494

Pagel M, Meade A, Barker D (2004) Bayesian estimation of ancestral character states on phylogenies. Syst Biol 53:673–684

Park SW, Kaimoyo E, Kumar D, Mosher S, Klessig DF (2007) Methyl salicylate is a critical mobile signal for plant systemic acquired resistance. Science 318:113–116. doi:10.1126/science.1147113

Posada D, Crandall KA (1998) Modeltest: testing the model of DNA substitution. Bioinformatics 14:817–818

Pott MB, Pichersky E, Piechulla B (2002) Evening specific oscillations of scent emission, SAMT enzyme activity, and SAMT mRNA in flowers of *Stephanotis floribunda*. J Plant Physiol 159: 925–934

Pott MB, Hippauf F, Saschenbrecker S, Chen F, Ross J, Kiefer I, Slusarenko A, Noel JP, Pichersky E, Effmert U, Piechulla B (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. doi:10.1104/pp.104.041806

Qin G, Gu H, Zhao Y, Ma Z, Shi G, Yang Y, Pichersky E, Chen H, Liu M, Chen Z, Qu LJ (2005) An indole-3-acetic acid carboxyl methyltransferase regulates *Arabidopsis* leaf development. Plant Cell 17:2693–2704. doi:10.1105/tpc.105.034959

Raguso RA, Light DM, Pichersky E (1996) Electroantennogram responses of *Hyles lineata* (Sphingidae: Lepidoptera) to volatile compounds from *Clarkia breweri* (Onagraceae) and other moth-pollinated flowers. J Chem Ecol 22:1735–1766

Raguso RA, Levin RA, Foose SE, Holmberg MW, McDade LA (2003) Fragrance chemistry, nocturnal rhythms and pollination "syndromes" in *Nicotiana*. Phytochemistry 63:265–284. doi:10.1016/S0031-9422(03)00113-4

Raguso RA, Schlumpberger BO, Kaczorowski RL, Holtsford TP (2006) Phylogenetic fragrance patterns in *Nicotiana* sections *Alatae* and *Suaveolentes*. Phytochemistry 67:1931–1942. doi:10.1016/j.phytochem.2006.05.038

Roeder S, Hartmann AM, Effmert U, Piechulla B (2007) Regulation of simultaneous synthesis of floral scent terpenoids by the 1, 8-cineole synthase of *Nicotiana suaveolens*. Plant Mol Biol 65:107–124. doi:10.1007/s11103-007-9202-7

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. doi:10.1006/abbi.1999.1255

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 USA 98:4788–4793. doi:10.1073/pnas.081557298

Seskar M, Shulaev V, Raskin I (1998) Endogenous methyl salicylate in pathogen-inoculated tobacco plants. Plant Physiol 116: 387–392

Shulaev V, Silverman P, Raskin I (1997) Airborne signalling by methyl salicylate in plant pathogen resistance. Nature 385:718–721. doi:10.1038/385718a0

Swofford DL (2003) PAUP*: phylogenetic analysis using parsimony (*and other methods). Version 4. Sinauer Associates, Sunderland

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 25:4876–4882. doi:10.1093/nar/25.24.4876

Van den Boom CEM, Van Beek TA, Posthumus MA, De Groot A, Dicke M (2004) Qualitative and quantitative variation among volatile profiles induced by *Tetranychus urticae* feeding on plants from various families. J Chem Ecol 30:69–89. doi:10.1023/B:JOEC.0000013183.72915.99

Van Poecke RMP, Posthumus MA, Dicke M (2001) Herbivore-induced volatile production by *Arabidopsis thaliana* leads to attraction of the parasitoid *Cotesia rubecula*: chemical, behavioral, and gene-expression analysis. J Chem Ecol 27:1911–1928. doi:10.1023/A:1012213116515

Varbanova M, Yamaguchi S, Yang Y, McKelvey K, Hanada A, Borochov R, Yu F, Jikumaru Y, Ross J, Cortes D, Ma C, Noel JP, Mander L, Shulaev V, Kamiya Y, Rodermel S, Weiss D, Pichersky E (2007) Methylation of gibberellins by *Arabidopsis* GAMT1 and GAMT2. Plant Cell 19:32–45. doi:10.1105/tpc.106.044602

Verdonk ML, Cole JC, Hartshorn MJ, Murray CW, Taylor RD (2003) Improved protein-ligand docking using GOLD. Proteins 52:609–623. doi:10.1002/prot.10465

Wang J, De Luca V (2005) The biosynthesis and regulation of biosynthesis of Concord grape fruit esters, including 'foxy' methylanthranilate. Plant J 44:606–619. doi:10.1111/j.1365-313X.2005.02552.x

Wang J, Dudareva N, Bhakta S, Raguso RA, Pichersky E (1997) Floral scent production in *Clarkia breweri* (Onagraceae). II. Localization and developmental modulation of the enzyme *S*-adenosyl-L-methionine:(iso)eugenol O-methyltransferase and phenylpropanoid emission. Plant Physiol 114:213–221

Wasternack C (2007) Jasmonates: an update on biosynthesis, signal transduction and action in plant stress response, growth and development. Ann Bot 100:681–697. doi:10.1093/aob/mcm079

Xu R, Song F, Zheng Z (2006) *OsBISAMT1*, a gene encoding *S*-adenosyl-L-methionine:salicylic acid carboxyl methyltransferase, is differentially expressed in rice defense responses. Mol Biol Rep 33:223–231. doi:10.1007/s11033-005-4823-x

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 448:123–132. doi:10.1016/j.abb.2005.08.006

Yang Y, Xu R, Ma C, Vlot AC, Klessig DF, Pichersky E (2008) Inactive methyl indole-3-acetic acid ester can be hydrolyzed and activated by several esterase belonging to *At*MES esterase family of Arabidopsis. Plant Physiol 147:1034–1045. doi:10.1104/pp.108.118224

Zhang J (2003) Evolution by gene duplication: an update. Trends Ecol Evol 18:292–298. doi:10.1016/S0169-5347(03)

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 146:455–467. doi:10.1104/pp.107.110049

Zhao N, Boyle B, Duval I, Ferrer JL, Lin H, Seguin A, Mackay J, Chen F (2009) SABATH methyltransferases from white spruce (*Picea glauca*): gene cloning, functional characterization and structural analysis. Tree Physiol 29:947–957

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 15:1704–1716. doi:10.1105/tpc.014548

Zwickl DJ (2006) Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. Ph.D. dissertation, The University of Texas at Austin