

**Functional Study of Plant SABATH Methyltransferases in
the Biosynthesis of Methyl Cinnamate, Juvenile Hormone
III and Methyl Gibberellins**

A Dissertation Presented for the
Doctor of Philosophy
Degree
The University of Tennessee, Knoxville

Chi Zhang
August 2019

Acknowledgement

I wished to express my deepest gratitude to my advisor Dr. Feng Chen for his mentoring, understanding and support through my Ph.D. study. Dr. Chen is a man with great wisdom, always willing to exchange opinions with me and give some constructive advice. He is a productive professor, always has many creative and original ideas. He encouraged me to develop independent thinking and research skills. He also offered many chances for me to improve my scientific writing, which benefited me a lot. I also thank Dr. Hong Guo, Dr. Robert Trigiano, Dr. William Klingeman, Dr. Max Cheng for their serving in my defense committee. They offered valuable suggestions for the dissertation, which brought me to understand my research from different perspectives. Besides my committee members, I also want to express my deep gratitude to Dr. Xinlu Chen, Dr. Minta Chaiprasongsuk, Dr. Guo Wei, Dr. Wangdan Xiong, Dr. Jianyu Fu for training me to improve my research skills and helping me overcome many difficulties in my researches and daily life. I want to thank the China Scholarship Council for funding support. I also want to thank Dr. Barbara Crandall-Stotler, Dr. Ping Qian, Dr. Guanglin Li, Dr. Yifan Jiang, Dr. Ying Chang for their kind cooperation during the research. I also appreciate all of the genuine help from people in the department. Finally, I want to dedicate this dissertation to my parents, Junting Zhang and Jun Chi and my wife, Xiuwen Li, for their unconditional love, support, encouragement, and care. My parents always put my education as a priority and support me even I am thus far away. Without their continued support and counsel, I could not complete this process. My wife accompanied me through all the hardship in my dissertation work and daily life. Without her persistent support and patience, I could not come this far.

Abstract

SABATH family of methyltransferases (MTs) is a group of plant MTs that are capable of methylating phytohormones and other small molecular compounds. This dissertation investigates biochemical function and evolution of *SABATH* genes with special interests in the biosynthesis of methyl cinnamate, juvenile hormone III, and methyl gibberellins. Methyl cinnamate is a fragrant volatile compound that occurs in a variety of land plants including great scent liverwort, *Conocephalum salebrosum* Szweykowski, Buczkowska and Odrzykoski. Using a comparative transcriptomic approach, we compare biosynthesis in liverworts and flowering plants and identified a cinnamic acid methyltransferase (CAMT) from *C. salebrosum* SABATH MTs. Structural and phylogenetic evidence indicate that methyl cinnamate biosynthesis in liverwort and flowering plants originated through convergent evolution. Juvenile hormones (JHs) are important in insect development and reproduction, however, JH III has been detected in some plants including *Cyperus iria* L. Unlike JH III biosynthetic pathway in insects that has been largely resolved, our understanding of JH III biosynthesis in plants remains limited. To reveal key enzymes involved in JH III biosynthesis in plants, a comparative transcriptomic approach was undertaken and identified *C. iria* homologs of those enzyme-coding genes in insects, but no homolog for the biosynthesis of methyl farnesoate, the immediate precursor of JH III in plants. However, *C. iria* SABATH MTs were used to identify a farnesoic acid methyltransferase (FAMT) that can produce methyl farnesoate. This discovery implied the independent evolution of JH III biosynthetic pathway in plants and insects. Gibberellins (GAs) are a class of plant hormones that have multiple roles in different physiological

processes. GA methylation is a recently discovered route of maintaining GA homeostasis which is catalyzed by gibberellic acid methyltransferase (GAMT). Because *GAMT* genes have previously identified only in the model plant *Arabidopsis*, the importance of this GA catabolic mechanism in other plants has remained unclear. To investigate the distribution, evolution, and biochemical functions of *GAMT* genes in land plants, we systemically analyzed 260 plant genomes and found that the *GAMT* gene arose early in the evolution of seed plants and was subsequently ~~got~~ lost in many flowering plants.

Table of Contents

Chapter I. Literature Review: The Discovery, Biochemical and Biological Functions and Evolution of SABATH Methyltransferases in Plants.....	1
1. Methylation in plant secondary metabolism.....	2
<i>1.1. General aspects of plant secondary metabolites and their biosynthesis.....</i>	<i>2</i>
<i>1.2. Methylation of small molecules and different types of methyltransferases in plants</i>	<i>3</i>
2. SABATH MTs and their biochemical and biological functions.....	4
3. SABATH MTs with unknown function.....	6
4. Functional evolution of the SABATH family.....	8
<i>4.1. Evolutionary relatedness and structural features of the SABATH family.....</i>	<i>8</i>
<i>4.2. Current understanding of functional evolution of the SABATH family.....</i>	<i>10</i>
5. SABATH gene discovery via comparative functional genomic approach	11
<i>5.1. Great scent liverwort (Conocephalum salebrosum) and rice flat sedge (Cyperus iria) as chosen plant models.....</i>	<i>13</i>
<i>5.2. Large scale phylogenetic study of SABATH family in the genome era</i>	<i>15</i>
References	17
Appendix A: Figures and Tables.....	30
Chapter II. Biosynthesis of Methyl (E)-cinnamate in the Liverwort <i>Conocephalum salebrosum</i> and Evolution of Cinnamic Acid Methyltransferase.....	31
Abstract	32
1. Introduction	34
2. Materials and methods.....	36
<i>2.1. Plant culture and axenic culture.....</i>	<i>36</i>
<i>2.2. Organic extraction and headspace collection.....</i>	<i>37</i>
<i>2.3. GC-MS analysis</i>	<i>37</i>
<i>2.4. Database search and phylogenetic analysis</i>	<i>38</i>
<i>2.5. Cloning full length cDNA of CsSABATHs</i>	<i>38</i>
<i>2.6. Semi-quantitative reverse-transcription PCR.....</i>	<i>39</i>
<i>2.7. Purification of CsSABATHs expressed in E. coli</i>	<i>39</i>
<i>2.8. Radiochemical methyltransferase activity assay</i>	<i>39</i>
<i>2.9. Protein structure modeling.....</i>	<i>40</i>
3. Results.....	41

3.1. Production of axenic culture of <i>C. salebrosum</i> and chemical profiling	41
3.2. Isolation of SABATH genes from <i>C. salebrosum</i> and gene expression analysis...	42
3.3. Biochemical characterization of CsSABATHs expressed in <i>E. coli</i>	43
3.4. Biochemical properties of CsCAMT	43
3.5. Comparison of CAMT genes in different chemotypes of the <i>C. conicum</i> complex	44
3.6. Structural modeling of CsCAMT, CsSAMT and ObCCMT1	44
3.7. Relatedness of CsCAMT with other SABATH methyltransferases	46
4. Discussion	47
5. Conclusions	50
Reference	52
Appendix B: Figures and Tables.....	63
Chapter III. Farnesoic acid methyltransferase involved in the biosynthesis of juvenile hormone III in rice flatsedge (<i>Cyperus iria</i>) is a member of the SABATH family	74
Abstract	75
1. Introduction	77
2. Material and methods	81
2.1. Plant culture	81
2.2. Metabolic profiling	81
2.3. GC-MS analysis	82
2.4. RNA preparation.....	82
2.5. Real-time qPCR	82
2.6. Cloning full length cDNA of CiSABATHs.....	83
2.7. Purification of CiSABATHs expressed in <i>E. coli</i>	83
2.8. Radiochemical CiSABATHs activity assay.....	84
2.9. Determination of kinetic parameters of CiSABATH3	84
2.10. Production of transcriptomes and analysis.	85
2.11. Phylogenic Analysis.....	86
3. Results.....	87
3.1. Chemical characterization of JH III and related compounds in <i>C. iria</i>	87
3.2. Transcriptomes of <i>Cyperus iria</i> root tissue.....	87
3.3. Identification of homologous genes for JH III biosynthesis in <i>C. iria</i>	88
3.4. Identification of SABATH genes among Unigenes	88

3.5. <i>Biochemical characterization of CiSABATHs expressed in E. coli</i>	89
3.6. <i>Detailed biochemical properties of CiFAMT</i>	89
3.7. <i>Comparison of FAMT gene expression in different tissues at different ages of C. iria</i>	90
3.8. <i>Phylogenetic Analysis of CiSABATHs with other SABATH members in Monocots</i>	90
4. Discussion	91
Reference	95
Appendix C: Figures and Tables.....	103
Chapter IV. Gibberellin Methyltransferase Gene is Ancient in Seed Plants and Got Lost in Many Flowering Plants.....	112
1. Introduction	115
2. Materials and methods.....	117
2.1. <i>Materials and reagents</i>	117
2.2. <i>Sequence retrieval and analysis</i>	118
2.3. <i>Phylogenetic reconstruction</i>	118
2.4. <i>Gene cloning and enzyme assays</i>	119
2.5. <i>Gene expression data retrieval</i>	120
2.6. <i>Accession numbers</i>	120
3. Results and discussion.....	120
3.1. <i>Comparative analysis of the SABATH family in 260 sequenced land plants</i>	120
3.2. <i>Catalytic activity of selected members in the GAMT clade and related non-GAMT enzymes</i>	122
3.3. <i>Apparent ortholog of GAMT gene is absent in about 2/3 of the 260 plants</i>	123
3.4. <i>GAMT-lacking Brachypodium distachyon lacks a GA-methylating gene</i>	124
3.5. <i>Retention after duplication of GAMT genes in Brassicaceae</i>	124
3.6. <i>Expressed patterns of GAMT genes in selected plant species</i>	125
4. Conclusions	126
Reference	128
Appendix D: Figures and Tables.....	134
Chapter V. Conclusions and Perspectives.....	160
1. Conclusions	161
2. Perspectives	163

2.1. <i>Biochemical functions of SABATH genes</i>	163
2.2. <i>Biological functions of SABATH genes</i>	164
2.3. <i>Evolution of SABATH genes</i>	165
References	167
Vita	169

List of Figures

- Figure A-1.** Sequences are from *A. thaliana* (blue), *O. sativa* (pink), *P. abies* (green) and other species. Branches were drawn to scale with the bar indicating 0.2 substitutions per site..... 30
- Figure B-1.** Chemical analysis of *C. salebrosum*. (A) GC chromatogram of the organic extraction of *C. salebrosum*. 1*, sabinene; 2, 1-octanol; 3, 1-octenyl acetate; 4*, methyl (E)-cinnamate; 5, unidentified sesquiterpene hydrocarbons; 6*, 10-epi-cubebol; 7*, 1-epi-cubanol; 8, 1-octadecyne; 9, E-2-tetradecen-1-ol; 10, phytol. Methyl (E)-cinnamate is the dominant component (50.52%) followed by sabinene (9.85%), 1-octadecyne (8.33%) and δ -cadinol (3.96%). (B) GC chromatogram of headspace collection of *C. salebrosum*. 1*, sabinene; 4*, methyl (E)-cinnamate; 5, unidentified sesquiterpene hydrocarbons. * indicates substrates that were verified by standard substances. 63
- Figure B-2.** Expression of six CsSABATHs genes in sterile thalli. Semi-quantitative RT-PCR of CsSABATH genes. Same amount of cDNA and respective specific primers were used for PCR. The cycle number is 25. 64
- Figure B-3.** Sequence alignment of CsSABATHs with ObCCMT1 and CbSAMT; CsSABATHs, *C. salebrosum* SABATHs; ObCCMT1, *Ocimum basilicum* cinnamate/p-coumarate carboxyl methyltransferase 1; CbSAMT, *Clarkia breweri* salicylic acid methyltransferase. Conserved residues are in shade with the more conserved the darker. Residues indicated with “&” are SAM/SAH-binding residues. Residues indicated with “*” are residues that interact with the carboxyl moiety of substrate. Residues indicated with “#” interact with the aromatic moiety of substrate and are important for substrate selectivity. 65
- Figure B-4.** Biochemical properties of CsCAMT. (A) pH optimum. Level of CsCAMT activity in the buffer of pH 6.5 was arbitrarily set at 1.0. (B) Thermostability. The activity of CsCAMT incubated at 4 oC for 30 min was arbitrarily set at 1.0. (C) Effect of ions. Individual ions were added to reactions in the form of chloride salts at final concentration of 5 mM. Level of activity without any ion added served as a control and was arbitrarily set at 1.0. (D) Steady-state kinetic measurements of CsCAMT. (E)-Cinnamic acid was used as substrate for enzyme assays in (A) to (D). Each value was the mean of three independent measurements. 67
- Figure B-5.** Authentic methyl (E)-cinnamate and the product of CsCAMT methylation. (A) GC chromatogram of authentic methyl (E)-cinnamate; (B) Mass spectrum of authentic methyl (E)-cinnamate; (C) GC chromatogram of the product of CsCAMT; (D) Mass spectrum of the product of CsCAMT..... 68
- Figure B-6.** Active sites of the structure models of CsCAMT, CsSAMT and ObCCMT1 with (E)-cinnamic acid or salicylic acid docked into the active sites. (A) CsCAMT (orange) complexed with (E)-cinnamic acid (green); the sulfur atom from SAH is shown as a yellow ball. Distances between the substrate and enzyme/SAH atoms are given in Å. (B) CsCAMT complexed with (E)-cinnamic acid. (C) CsCAMT complexed with salicylic acid. (D) CsSAMT complexed with salicylic acid. (E) ObCCMT1 complexed with (E)-cinnamic acid. (F) Superposition of ObCCMT1 (orange) and CsCAMT (cyan) with some active-site residues shown. (E)-Cinnamic acid and SAH are shown in line except for the sulfur atom. 69
- Figure B-7.** Phylogenetic tree of the SABATH family of methyltransferases in *C. salebrosum*, *Arabidopsis thaliana*, *Oryza sativa*, *Physcomitrella patens*, *Selaginella*

moellendorffii and ObCCMTs. The SABATH proteins from each species were color-coded. CCMTs: Cinnamate/p-coumarate methyltransferases; CAMT: cinnamic acid methyltransferase; IAMT: indole-3-acetic acid methyltransferase; SAMT: salicylic acid methyltransferase. AtIAMT, AtGAMTs, CsCAMT, CsSAMT, OsIAMT and ObCCMTs are indicated with arrows.	71
Figure C-1. Proposed pathway of JH III biosynthesis. Intermediate and final products are in bold. Lepidopteran specific pathway is indicated by the dashed line and arrow. Farnesoic acid methyltransferase is labeled in red.	103
Figure C-2. Identification of JH3 and related chemicals in <i>C. iria</i> . (A) GC chromatogram of the organic extraction of <i>C. iria</i> . 30DAG leaf in red color; 30DAG root in brown color; 60DAG leaf in blue color and 60DAG root in green color; (B) Mass spectrum of compound b; (C) Mass spectrum of compound c; (D) Mass spectrum of compound d. (E) Quantification of Juvenile hormone (JH) III, methyl farnesoate (MF) and farnesol (FO). Each identity was verified by an authentic standard. DAG, day after germination. 1-octanol was added as internal control.....	105
Figure C-3. The number of <i>C. iria</i> homologs of those enzyme-coding genes in insect and proposed JH III biosynthetic pathway in plants. Key enzymes are labeled in red; homolog numbers are in the parentheses.	107
Figure C-4. Sequence alignment of CiSABATHs with AtFAMT and CbSAMT. Conserved residues are in shade with the more conserved the darker. Residues indicated with “&” are SAM/SAH-binding residues. Residues indicated with “*” are residues that interact with the carboxyl moiety of substrate. Residues indicated with “#” interact with the aromatic moiety of substrate and are important for substrate selectivity. CiSABATHs, <i>Cyperus iria</i> SABATHs; AtFAMT, <i>Arabidopsis thaliana</i> farnesoic acid methyltransferase; CbSAMT, <i>Clarkia breweri</i> salicylic acid methyltransferase.	108
Figure C-5. Biochemical properties of CiSABATH3(CiFAMT) using FA as substrate. (A) Activity of CiSABATH3 with different substrates. The highest activity was arbitrarily set at 1.0. (B) pH effect on CiSABATH3 activity. The activity at pH 7.0 was arbitrarily set at 1.0. (C) Thermostability of CiSABATH3. The activity of CiSABATH3 with incubation at 0 oC for 30 min was arbitrarily set at 1.0. (D) Effects of metal ions on activity of CiSABATH3. The activity of CiSABATH3 without any metal ion added as control (Ctrl) was arbitrarily set at 1.00. (E) Km value of CiSABATH3.	109
Figure C-6. Phylogenetic tree of the SABATH family of methyltransferase in <i>C. iria</i> with SABATHs from selected monocots. The SABATH proteins from each species were color-coded. FAMTs: farnesoic acid methyltransferases; IAMTs: indole-3-acetic acid methyltransferases; BSMTs: benzoic acid/salicylic acid methyltransferases; JAMT: jasmonic acid methyltransferase. Proteins for different species are color coded.	110
Figure D-1. Known routes of catabolism of gibberellins (GAs). GA ₄ is one widely occurring bioactive GA. Methylation is the most recently discovered mechanism.	134
Figure D-2. Phylogenetic analysis of SABATH proteins. (A) Phylogenetic tree SABATH methyltransferases from 260 sequenced plants with five recognized groups; (B) phylogeny of group I is shown with bootstrap value.....	135
Figure D-3. Phylogenetic tree of the SABATH protein in the putative GAMT clade (A) and species tree (B).	136

Figure D-4. Presence/absence of SABATHs of the GAMT clade among families of sequenced plants. Three taxa with complete loss of GAMTs were shaded. Presence/absence of SABATH genes in the GAMT clade among families of sequenced plants. The phylogeny was redrawn from APG IV (2016). The lineages in gray include that no species from those lineages was analyzed in this study. The two numbers (red and black) includes in the number of species containing GAMT gene and the total number of species from that specific lineages analyzed in this study.	137
Figure D-5. Phylogenetic tree of SABATHs of the GAMT clade in Brassicaceae. The two clades in blue and in yellow depicts two clades that resulted from a whole gene duplication event occurred in the ancestor of Brassicaceae. The two smaller blocks indicate gene duplication due to Brassica-specific whole genome duplication.	139
Figure D-6. Expression patterns of GAMTs in different selected species. (A) <i>Ginkgo biloba</i> ; (B) <i>Picea abies</i> ; (C) <i>Phalaenopsis equestris</i> ; (D) <i>Musa acuminata</i> ; (E) <i>Vitis vinifera</i> ; (F) <i>Citrus sinensis</i> ; (G) <i>Arabidopsis thaliana</i> ; (H) <i>Camelina sativa</i> . O: Ovules; Mc: Male cones; Sl: Stem and leaves; S: Stem; L: Leaf; R: Root; Fs: Floral stalk; F: Flower; S: Sepal; P: Petal; La: Labellum; G: Gynostemium; 40F: 40-day-fruit; 60F: 60-day-fruit; 80F: 80-day-fruit; Yt: Young tendril; Yl: Young leaf; Sf: Senescent leaf; B: Bud; Yf: Young flower; Rp: Ripening pericarp; C: Callus; Fr: Fruit; Ro: Rosette; Cl: Cauline leaf; Se: Seed; The highest level of expression was arbitrarily set at 1.0. Enzyme expression profiles were obtained from open-source databases, as described above.	140

List of Tables

Table B-1. The relative activities of CsSABATH4 and CsSABATH6 from <i>C. salebrosum</i> with ten carboxyl acids.	73
Table D-1. Plant genomes used in phylogenetic analysis. The source of each genome used in this study was listed with species name and database or original literature.	143
Table D-2. Information of characterized SABATH methyltransferases. Enzyme names and their corresponding literatures were listed respectively. NCBI accession number was also provided if available.	151
Table D-3. SABATH gene family in <i>Brachypodium distachyon</i>	156
Table D-4. Gene-specific primers used in this study.	157
Table D-5. Activities of GAMTs from selected plant species against GA ₃ and GA ₄	158

Chapter I. Literature Review: The Discovery, Biochemical and Biological Functions and Evolution of SABATH Methyltransferases in Plants

1. Methylation in plant secondary metabolism

1.1. General aspects of plant secondary metabolites and their biosynthesis

Plant secondary metabolites are diverse specialized organic compounds that are required for the survival of plants but generally not essential for growth or development. Major classes of secondary metabolites include flavonoids and allied phenolic and polyphenolic compounds; terpenoids; nitrogen-containing alkaloids and Sulphur-containing compounds (Wink, 2010). These specialized compounds are involved in multiple biological and ecological processes, such as interactions with other organisms (attraction of beneficial organisms or deterrent against pathogens and herbivores) and protection under abiotic stresses (e.g., ultraviolet radiation) (Howlett, 2006; Kessler and Baldwin, 2007; Singh and Sharma, 2015; Tohge et al., 2016). Moreover, some secondary metabolites serve as precursors of important phytohormones and regulate the balance of primary metabolic pathways (Fujioka and Sakurai, 1997; Yamaguchi, 2008). Intricate biosynthetic pathways produce diverse secondary metabolites, involving a diverse array of modifications such as methylation, hydroxylation, glycosylation and halogenation (Noel et al., 2005). These modifications not only lead to a rich set of secondary metabolites but also control the homeostasis of some important intermediate and/or final products. Sometimes these modifications play vital roles in regulating availability, localization and biological activity of secondary metabolites (Koeduka et al., 2016) and sometimes they are just supplementary measures other than anabolism and catabolism (Varbanova et al., 2007).

1.2. Methylation of small molecules and different types of methyltransferases in plants

Among above-mentioned modifications, methylation is one of the most common and basic reaction. Methylation is catalyzed by a group of methyltransferases that transfer a methyl group from a methyl donor to an acceptor moiety. Usually, the donor in the cell is S-adenosyl-L-methionine (SAM). Based on the chemical nature of the substrates and the atoms to which a methyl group is transferred, methyltransferases (MTs) that catalyze the methylation of small molecules are divided into different types: carbon MT (CMT), nitrogen MT (NMT), sulfur MT (SMT) and oxygen MT (OMT). O-methyl esters make up the largest class of methylated specialized metabolites and MTs that catalyze O-methylation of small molecules have been identified and categorized into three groups based on their structures and functions (Noel et al., 2003).

Type I OMTs use phenylpropanoids, flavonoids, isoflavonoids, alkaloids, ketones and polyketides as substrates (Gang et al., 2002; Zubieta et al., 2001). This group of OMTs does not require divalent cations for activity compared with type II OMTs, which use CoA-esters of phenylpropanoids as substrates and they are involved in lignin biosynthesis (Ye et al., 1994). This group of OMTs requires a divalent cation to mediate the methyl group transfer (Boudet, 2000; Day et al., 2001; Maury et al., 1999). Type III OMTs share no sequence similarities with members of the previous two types but have similarity to each other in the group. The first enzyme that was characterized from this family is the salicylic acid carboxyl MT (SAMT) from *Clarkia breweri* (Ross et al., 1999). Interestingly, NMT involved in caffeine biosynthesis is highly homologous to SAMT, therefore it belongs to the same protein family. To distinguish this group of MTs from others, they were named SABATH based on

the first three enzymes to be characterized, SAMT, BAMT (benzoic acid MT), and Theobromine synthase with the last one being a NMT (Kato et al., 2000; Kato et al., 1999; Ogawa et al., 2001; Seo et al., 2001).

2. SABATH MTs and their biochemical and biological functions

Besides the first identification of SAMT from *Clarkia breweri* (Ross et al., 1999), SAMT was later identified in multiple species including *Antirrhinum majus*, *Stephanotis floribunda*, *Atropa belladonna*, *Datura wrightii*, *Solanum lycopersicum*, *Populus trichocarpa*, *Glycine max* and *Hoya carnosa* (Barkman et al., 2007; Effmert et al., 2005; Fukami et al., 2002; Lin et al., 2013; Negre et al., 2002; Pott et al., 2002; Tieman et al., 2010; Zhao et al., 2013). Many more SABATH members that utilize other carboxyl acids as substrates have been identified in recent years in different species of plants. Benzoic acid methyltransferase was identified from *Antirrhinum majus* (Murfitt et al., 2000). Benzoic acid/salicylic acid methyltransferases were identified in *Arabidopsis thaliana*, *Arabidopsis lyrata*, *Petunia hybrida*, *Nicotiana suaveolens* and *Oryza sativa* (Chen et al., 2003; Koo et al., 2007; Negre et al., 2003; Pott et al., 2004). The volatile ester methyl salicylate has several biological functions in plants. It is a constituent of flower scents of various angiosperms. In California wildflower *Clarkia breweri*, it makes up about 5% of the total scent and is considered as an attractant to pollinators (Ross et al., 1999). Flowers of the Japanese loquat *Eriobotrya japonica* emit approximately 20 different floral scent compounds, including methyl benzoate, methyl cinnamate and methyl *p*-methoxybenzoate (Koeduka et al., 2016). Tobacco plants emit methylsalicylate from their leaves when they are infected by the tobacco mosaic virus

(Seskar et al., 1998). Indole-3-acetic acid methyltransferases were identified from *Picea abies*, *Arabidopsis thaliana*, *Populus trichocarpa*, *Oryza sativa*, *Solanum lycopersicum*, and *Picea glauca* (Chaiprasongsuk, 2016; Qin et al., 2005; Zhao et al., 2009; Zhao et al., 2008; Zhao et al., 2007). Jasmonic acid methyltransferases were identified from *Arabidopsis thaliana*, *Solanum lycopersicum*, *Populus trichocarpa*, *Fragaria vesca*, *Brassica juncea* and *Oryza sativa* (Meur et al., 2015; Preuß et al., 2014; Seo et al., 2001; Tieman et al., 2010; Zhao et al., 2013). In the case of methyl jasmonate, it plays as a cellular regulator by inducing the expression of defense-related genes when plants are damaged by herbivores or challenged by pathogen infection in addition to being a floral scent compound (Cheong et al., 2003; Robet et al., 1992; Howe et al., 2004; Koo et al., 2009). Gibberellic acid methyltransferases were identified only in *Arabidopsis thaliana* (Varbanova et al., 2007). Although there is evidence that phytohormone methylated products like MeJA and MeSA can be converted back to JA and SA respectively, ancymidol inhibition experiment indicated that the function of GAMT methylating GAs in developing seeds of *Arabidopsis thaliana* is an irreversible deactivation of GAs (Varbanova et al., 2007). Farnesoic acid methyltransferase was only identified in *Arabidopsis thaliana* (Yang et al., 2006). Anthranilic acid methyltransferases were only identified in *Zea mays* (Köllner et al., 2010). Loganic acid methyltransferases were identified in *Camellia irrawadiensis* and *Catharanthus roseus* (Murata et al., 2008; Petronikolou et al., 2018). Norbixin carboxyl methyltransferase was only indentified in *Bixa orellana* (Bouvier et al., 2003). Cinnamate/p-coumarate carboxyl methyltransferases were identified in *Ocimum basilicum* (Kapteyn et al., 2007). p-Methoxybenzoic acid methyltransferase was identified in *Eriobotrya japonica* (Koeduka et al., 2016). More N-methyltransferases that were related to

caffeine biosynthesis were identified from *Paullinia cupana*, *Citrus sinensis*, *Camellia ptilophylla*, *Camellia irrawadiensis*, *Coffea arabica* and *Theobroma cacao* (Huang et al., 2016; Ogawa et al., 2001; Uefuji et al., 2003; Yoneyama et al., 2006). In tea plant, heterologous expression of TCS1 in *E. coli* showed enzyme activity of catalyzing 3-N-methylation and 1-N-methylation but not 7-N-methylation which suggested TCS1 encodes caffeine synthase (Kato et al., 2000). In coffee plant, tissue expression study showed CaMXMT accumulates in young leaves and stems. The 7-methylxanthine methyltransferase activity of CaMXMT suggested that it functions in the production of theobromine as the second step of the three independent methylation steps in caffeine biosynthesis (Ogawa et al., 2001). These NMTs were shown to be structurally related to SAMT and BAMT (Uefuji et al., 2003; Kato et al., 2000). A thiobenzoic acid methyltransferase was identified from *Physcomitrella patens* (Zhao et al., 2012).

3. SABATH MTs with unknown function

As various plant genomes were sequenced, a large number of SABATH genes were discovered from plants (Chaiprasongsuk, 2016; Zhao et al., 2008). However more than half of SABATH family are still mysterious to us with their unknown function. For example, in *Arabidopsis thaliana* 24 putative SABATH genes were identified and six of them have been biochemically characterized including IAMT, BSMT, JAMT, GAMT and FAMT (Chen et al., 2003; Qin et al., 2005; Seo et al., 2001; Varbanova et al., 2007; Yang et al., 2006). However, the remaining majority are still functionally unknown. Gene expression data showed that SABATH genes in *Arabidopsis thaliana* express in their specific patterns which implied that

they may have different biochemical properties for diverse biological functions. Another example, in *Oryza sativa*, only three members of SABATH family have been characterized from 41 of the whole family and these three characterized SABATH were IAMT, BSMT and JAMT which were not strange to us since they were identified in other species before (Koo et al., 2007; Qi et al., 2016; Zhao et al., 2008). Expression data of all 41 SABATH genes from *Oryza sativa* also showed specific patterns through different tissues which indicated the functional diversity of this family in rice (Zhao et al., 2008). Interestingly in *Physcomitrella patens* only four SABATH genes have been identified from its genome which is far less than that in seed plants. Only one of them was functionally characterized with S-methyltransferase activity which was not detected from any other plants before (Zhao et al., 2012). This triggered curiosity that how its activity evolved and what the function of the remaining three was. Since several familiar substrates were used to screen the activity of these SABATH enzymes but none was detected, we can infer these three must hold some unknown catalytic activity.

Overall, although inspiring advances have been achieved in the research on biochemical and biological functions of the SABATH family and genomic information has helped to identify the whole family of SABATHs in several species, a serial of interesting questions still needs to be answered. First, what are the function of uncharacterized SABATH members? Second, are some of those known function conserved among species? Three, are there some species-specific functions? Fourth, how did some SABATH genes achieve diverged functions? Fifth, what's the original form of SABATH? To address these questions,

we need to continue to isolate and characterize SABATH members especially those with unknown function from plants of proper evolutionary context.

4. Functional evolution of the SABATH family

4.1. Evolutionary relatedness and structural features of the SABATH family

All SABATH proteins characterized so far contain amino acids between the number of 357 and 389 with molecular weights ranging from 40-49 kDa. Size exclusion chromatography shows that the molecular mass of native enzymes is roughly 80kDa, indicating a dimeric structure (Kapteyn et al., 2007). None of the SABATH proteins characterized so far contains any subcellular targeting signals except newfound PbBSMT (*Plasmodiophora brassicae* benzoic acid and salicylic acid methyltransferase), which contains a signal peptide targeting the protein to the secretory pathway (Ludwig-Muller et al., 2015).

Pairwise alignment and maximum likelihood phylogenetic construction (Figure A-1) indicate that all known IAMTs cluster into a single clade, implying that IAMT is an evolutionarily ancient member of the SABATH family (Zhao et al., 2008). The clustering of other MTs is not constrained by their substrate specificity. For example, 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 significant structural difference between the two enzyme types is that the Met (150th in *C. breweri*) residue in the SAMT-type is replaced by a His residue in the BAMT-type enzymes. SAMT and BAMTs appear to be polyphyletic (Hippauf et al., 2010).

To further understand the structural basis of substrate recognition in SAMT, the 3.0 Å crystal structure of *Clarkia breweri* SAMT in complex with the substrate salicylic acid and the demethylated product S-adenosyl-L-homocysteine (SAH) was obtained (Chole et al., 2003). A protein structure that possesses a helical active site capping domain and a unique dimerization interface have been revealed. The substrate specificity of SAMT by site-directed mutagenesis and activity assays against small carboxyl molecules gave us more information about SABTH proteins. This helped the identification of IAMT in *Arabidopsis* (Chole et al., 2003).

The crystal structure of *Arabidopsis thaliana* IAMT (AtIAMT) was determined and refined to 2.75 Å resolution (Zhao et al., 2008). The overall tertiary and quaternary structures closely resemble the two-domain bilobed monomer and the dimeric arrangement respectively. The residues of AtIAMT1 that likely interact with the carboxyl moiety of the IAA substrate are strictly conserved with respect to CbSAMT including Lys-10, Gln-25, and Trp-162.

Recently a fatty acid O-methyltransferase (FAMT) from a bacterium *Mycobacterium marinum* (MmFAMT) was biochemically characterized and its crystal structures with different chemicals as substrates were compared with methyltransferases that are involved in plant natural product metabolism (Petronikolou and Nair, 2015). It was found that an architectural fold employed in plant natural product biosynthesis is used in bacterial fatty acid O-methylation. From the crystal structure of MmFAMT, it shows that MmFAMT can methylate medium-chain fatty acids. This interesting discovery not only broadens our view of the presence of SABATH analog in different taxa but also intrigued the thought of the origin of SABATH family.

In the protein domain database Pfam, 20 unique domain organizations of architecture in which the Methyltransferase_7 (SABATH) are classified. From bacteria, fungi to embryophyte, SABATH coding sequences are present in 174 known species.

4.2. Current understanding of functional evolution of the SABATH family

The SABATH family has been studied in a number of plant species. In all these plants of diverse taxonomic groups, SABATH genes exist as a group of more than two members which implies divergence in evolution (Zhao et al., 2008). It is interesting to look into the evolutionary process of SABATH family and understand how they achieve high specificity towards individual substrates.

Using the *Oryza sativa* SABATH family (OsSABATH) as an example, forty-one OsSABATHs were identified (Zhao et al., 2008). OsSABATH4 is most similar to AtIAMT1 which was previously identified as an IAMT. OsSABATH4 was characterized as having the highest level of catalytic activity toward IAA. Its kinetic properties are similar to those of AtIAMT1 and poplar IAMT (PtIAMT1). Since the crystal structure of AtIAMT1 was determined and refined to 2.75 Å resolution, it was then used as a template to model the structure of OsIAMT1(OsSABATH4) and PtIAMT1. Conservative structural features of these IAMTs were found within the active-site cavity (Lys-10, Gln-25, and Trp-162) and they are divergent from other SABATH members which have different catalytic preference, like CbSAMT. A monophyletic group was formed with IAMTs from these three species (*Arabidopsis*, rice, and poplar). All this information supports the hypothesis that IAMT is an

evolutionarily ancient member of the SABATH family, at least in flowering plants (Zhao et al., 2008).

It is hard to predict the catalytic activity of SABATHs merely based on sequence and phylogenetic analysis. A recent study showed that OsSABATH3 has BSMT activity in vitro, but it doesn't fall into the same clade of others known SAMTs (Koo et al., 2007). It implies that SAMTs may evolve several times during the course of SABATH gene evolution. The sequence identity between theobromine synthase and caffeine synthase within the genus of *Camellia* is very high, but the sequence identity between the N-methyltransferases from *Camellia* and *Theobroma* is lower. This suggests that caffeine biosynthetic pathways may have evolved independently several times in plants as well (Yoneyama et al., 2006).

As one example study of the functional evolution of SABATH genes, relative enzyme activities of 13 SAMT, BSMT, and NAMT with 18 substrates were tested and compared in species from Apocynaceae and Solanaceae (Huang et al., 2012). Most SAMTs show a high preference for SA relative to BA, most BSMTs show a high preference for BA than SA and NAMT show high preference for nicotinic acid (NA). Two gene duplication events were found, one occurred early in Solanaceae evolutionary leading to the functional division of SAMT and BSMT and a later one only in *Nicotiana* (Huang et al., 2012).

5. SABATH gene discovery via comparative functional genomic approach

Although several SABATH genes have been cloned and their corresponding proteins characterized, our knowledge of this family of methyltransferases is still limited. A number of

outstanding questions from chemical perspectives, biological perspectives, and evolutionary perspectives remain to be answered.

First, there are some specialized methylated products of small molecular weights such as juvenile hormone in plants (Atherton et al., 2010; Toong et al., 1988). How are these chemicals biosynthesized? They could be synthesized by uncharacterized SABATH members. Thus, it is interesting to determine whether they are indeed synthesized by the action of SABATH proteins or other methyltransferases.

Second, most of the biochemical and biological studies of SABATH methyltransferases are in seed plants especially flowering plants. In contrast, the functional study of SABATH methyltransferases in non-seed plants is very limited. And most of the characterized SABATH methyltransferases are related to phytohormone methylation. Thus, it is necessary to study SABATH family in non-seed plants, for example, bryophytes, and identify their biochemical function and roles in biological processes.

Third, structural, biochemical and phylogenetic analyses have suggested that IAMT is an evolutionarily ancient member of the SABATH family (Zhao et al., 2008). However, other phytohormone methyltransferases have also been isolated and characterized from several different species, so how did these MTs evolve? So far, GAMTs have been identified and characterized only from *Arabidopsis*, thus we need to figure out whether other species also have GAMTs and whether GAMT is an evolutionarily ancient member like IAMT. A large-scale phylogenetic study of SABATH family across different species seems like a good way to start.

5.1. Great scent liverwort (*Conocephalum salebrosum*) and rice flat sedge (*Cyperus iria*) as chosen plant models

In this dissertation, *C. salebrosum* and *C. iria* are used as plant models. *C. salebrosum* was chosen as a plant model for the isolation and functional characterization of cinnamic acid methyltransferase (CAMT) and for studying SABATH family in liverworts. This liverwort is widely distributed in North America, Europe, and East Asia. It is well known for its great scent and snakeskin (Atherton et al., 2010). The major compound made by *C. salebrosum* is identified as methyl (*E*)-cinnamate (Harinantenaina et al., 2007; Toyota, 2000; Wood et al., 1996). However, later studies put *C. salebrosum* into *C. conicum* complex as genotype S (Ludwiczuk et al., 2013; Szweykowski et al., 2005; Wood et al., 1996). Interestingly three chemotypes of *C. conicum* complex were then identified: bornyl acetate-dominant, sabinene-dominant and methyl (*E*)-cinnamate-dominant (Toyota, 2000; Toyota et al., 1997). That brought a question: what's the molecular basis of these different chemotypes? Here, *C. salebrosum* was chosen as a plant model to study its methyl (*E*)-cinnamate biosynthesis and compare it to that in seed plants for gaining knowledge of underlying SABATH family. The plant material is collectible from multiple sites across the Knoxville area and our cooperators. The transcriptome data of *C. conicum* complex is available through the OneKP database (Matasci et al., 2014).

C. iria was chosen as a plant model for studying the biosynthesis of JH III in plants and monocot SABATH family. *C. iria* is known as an invasive rice field weed and it is also called “grasshopper’s *Cyperus*” (Toong et al., 1988). JH III was unexpectedly identified from *C. iria* extract in high content (Toong et al., 1988). Since JH III is an essential insect

hormone, it has been recognized as a potential insecticide for herbivores. Aside from chemical biosynthesis, a lot of effort has been put to decode its biological synthesis firstly in insects. The insect biosynthetic pathway to JH III is now clear: farnesyl diphosphate to farnesol, farnesol to farnesal, farnesal to farnesoic acid, farnesoic acid to methyl farnesoate and finally methyl farnesoate into JH III (Baker et al., 1983; Feyereisen et al., 1981; Goldstein and Brown, 1990; Shinoda and Itoyama, 2003). The last two steps might differ depending on different species (farnesoic acid to JH acid, then JH acid to JH III) (Bhaskaran et al., 1986). Several key enzyme-coding genes have been characterized including a cytochrome P450 that catalyzes the epoxidation of methyl farnesoate to JH III and a methyltransferase that methylates JH acid into JH III (Bhaskaran et al., 1986; Shinoda and Itoyama, 2003). On the contrary, the progress on the JH III biosynthesis in plants is relatively limited. Most of the studies relied on *C. iria* cell suspension cultures. By enzyme inhibition, labeling studies and precursor feeding assays, the biochemical pathway in *C. iria* appears to be similar to the one in insects (Bede et al., 1999a, 2000; Bede et al., 1999b; Bede et al., 2001; Chappell et al., 1995). However, none of the key enzyme-coding genes have been isolated and biochemically characterized. For the last but one step in the biosynthesis, an assumption is that a homologous gene to that in insect encodes a methyltransferase or an unrelated gene coding other methyltransferase has a similar function. In the latter situation, SABATH methyltransferase becomes a major candidate since this family of methyltransferases can catalyze a variety of carboxyl acids. However, no genome or expression information about *C. iria* is available so far. We decided to collect plant tissues and performed an RNA-seq to get transcriptome sequence data.

5.2. Large scale phylogenetic study of SABATH family in the genome era

Early studies of SABATH methyltransferases focused on biochemical characterization. As more and more SABATH members being characterized, the diversity of its substrates is beyond prediction. Some clues were gained about the evolution of various catalytic actives when phylogenetic the method was performed. For example, IAMT was identified from multiple species and proved to be an ancient member of SABATH family (Zhao et al., 2008). However, most of the phylogenetic studies on SABATH family so far are based on one species or several closely related species. Questions like whether IAMT or its orthologs exist in all species are hard to answer. Not to mention that GAMTs have only been studied in *Arabidopsis* (Varbanova et al., 2007). As enzyme functional evolution studies can give hint about the diversity of catalytic activities (Hippauf et al., 2010; Huang et al., 2012), it is necessary to bring the phylogenetic studies to a new level. To date, about 300 plant species have been sequenced, the quality and quantity of new genomes are steadily on the increase. A large-scale phylogenetic analysis that combines all available genome information will certainly deepen our knowledge about the evolution of SABATH family and eliminate possible bias in previous studies due to a limited number of sequences.

Taking advantage of fully sequenced genomes and transcriptomes of various vascular and non-vascular plants, I propose to use a comparative functional genomic approach to identify the SABATH families in liverwort, rice flat sedge, and many other plants and then to investigate the biochemical and evolution of identified members of SABATH family. The purpose of this Ph.D. research is to significantly enhance our understanding of the function and evolution of the SABATH family.

This dissertation contains five specific objectives. In objective 1, identification, expression analysis and phylogenetic analysis will identify the SABATH members from *Conocephalum salebrosum* and *Cyperus iria*. In objective 2, *C. salebrosum* CAMT will be cloned and its biochemical and biological functions will be studied. In objective 3, FAMT from *C. iria* will be studied. Phylogenetic and structural analysis will be performed to understand the evolution of the SABATH family. Objective 4 is GAMTs from *Brassica napus* and *Ginkgo biloba* will be identified, and its biochemical will be studied in order to understand GAMTs in brassicas and gymnosperm. In objective 5, GAMTs from a large number of plants with full genome sequenced will be identified, synthesized and biochemically studied to understand the evolution of GAMT in land plants.

References

Atherton, I., Bosanquet, S., Lawley, M., 2010. Mosses and liverworts of Britain and Ireland: a field guide. British Bryological Society Plymouth.

Baker, F. C., Mauchamp, B., Tsai, L. W., Schooley, D. A., 1983. Farnesol and farnesal dehydrogenase (s) in corpora allata of the tobacco hornworm moth, *Manduca sexta*. Journal of lipid research 24, 1586-1594.

Barkman, T. J., Martins, T. R., Sutton, E., Stout, J. T., 2007. Positive selection for single amino acid change promotes substrate discrimination of a plant volatile-producing enzyme. Molecular biology and evolution 24, 1320-1329.

Bede, J., Goodman, W., Tobe, S., 1999a. Developmental distribution of insect juvenile hormone III in the sedge, *Cyperus iria* L. Phytochemistry 52, 1269-1274.

Bede, J., Goodman, W., Tobe, S., 2000. Quantification of juvenile hormone III in the sedge *Cyperus iria* L.: comparison of HPLC and radioimmunoassay. Phytochemical Analysis: An International Journal of Plant Chemical and Biochemical Techniques 11, 21-28.

Bede, J., Teal, P., Tobe, S., 1999b. Production of insect juvenile hormone III and its precursors in cell suspension cultures of the sedge, *Cyperus iria* L. Plant cell reports 19, 20-25.

Bede, J. C., Teal, P. E., Goodman, W. G., Tobe, S. S., 2001. Biosynthetic pathway of insect juvenile hormone III in cell suspension cultures of the sedge *Cyperus iria*. Plant physiology 127, 584-593.

Bhaskaran, G., Sparagana, S. P., Barrera, P., Dahm, K. H., 1986. Change in corpus allatum function during metamorphosis of the tobacco hornworm *Manduca sexta*: regulation at the terminal step in juvenile hormone biosynthesis. Archives of insect biochemistry and physiology 3, 321-338.

Boudet, A.-M., 2000. Lignins and lignification: selected issues. Plant Physiology and Biochemistry 38, 81-96.

Bouvier, F., Dogbo, O., Camara, B., 2003. Biosynthesis of the food and cosmetic plant pigment bixin (annatto). Science 300, 2089-2091.

Chaiprasongsuk, M., 2016. Biochemistry and Evolution of the Phytohormone-methylating SABATH Methyltransferase in Plants.

Chappell, J., Wolf, F., Proulx, J., Cuellar, R., Saunders, C., 1995. Is the reaction catalyzed by 3-hydroxy-3-methylglutaryl coenzyme A reductase a rate-limiting step for isoprenoid biosynthesis in plants? Plant Physiology 109, 1337-1343.

Chen, F., D'auria, J. C., Tholl, D., Ross, J. R., Gershenzon, J., Noel, J. P., Pichersky, E., 2003. An *Arabidopsis thaliana* gene for methylsalicylate biosynthesis, identified by a biochemical genomics approach, has a role in defense. The Plant Journal 36, 577-588.

Day, A., Dehorter, B., Neutelings, G., Czeszak, X., Chabbert, B., Belingheri, L., David, H., 2001. Caffeoyl-coenzyme A 3-O-methyltransferase enzyme activity, protein and transcript accumulation in flax (*Linum usitatissimum*) stem during development. *Physiologia plantarum* 113, 275-284.

Effmert, U., Saschenbrecker, S., Ross, J., Negre, F., Fraser, C. M., Noel, J. P., Dudareva, N., Piechulla, B., 2005. Floral benzenoid carboxyl methyltransferases: From in vitro to in planta function. *Phytochemistry* 66, 1211-1230.

Feyereisen, R., Pratt, G. E., Hamnett, A. F., 1981. Enzymic synthesis of juvenile hormone in locust corpora allata: Evidence for a microsomal cytochrome P-450 linked methyl farnesoate epoxidase. *European Journal of Biochemistry* 118, 231-238.

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 and cell physiology* 43, 1054-1058.

Gang, D. R., Lavid, N., Zubieta, C., Chen, F., Beuerle, T., Lewinsohn, E., Noel, J. P., Pichersky, E., 2002. Characterization of phenylpropene O-methyltransferases from sweet basil: facile change of substrate specificity and convergent evolution within a plant O-methyltransferase family. *The Plant Cell* 14, 505-519.

Goldstein, J. L., Brown, M. S., 1990. Regulation of the mevalonate pathway. *Nature* 343, 425.

Harinantenaina, L., Kida, S., Asakawa, Y., 2007. Phytochemistry of three selected liverworts: *Conocephalum conicum*, *Plagiochila barteri* and *P. terebrans*. *Arkivoc* 7, 22-29.

Hippauf, F., Michalsky, E., Huang, R., Preissner, R., Barkman, T. J., Piechulla, B., 2010. Enzymatic, expression and structural divergences among carboxyl O-methyltransferases after gene duplication and speciation in *Nicotiana*. *Plant molecular biology* 72, 311-330.

Howlett, B. J., 2006. Secondary metabolite toxins and nutrition of plant pathogenic fungi. *Current opinion in plant biology* 9, 371-375.

Huang, R., Hippauf, F., Rohrbeck, D., Haustein, M., Wenke, K., Feike, J., Sorrelle, N., Piechulla, B., Barkman, T. J., 2012. Enzyme functional evolution through improved catalysis of ancestrally nonpreferred substrates. *Proceedings of the National Academy of Sciences* 109, 2966-2971.

Huang, R., O'Donnell, A. J., Barboline, J. J., Barkman, T. J., 2016. Convergent evolution of caffeine in plants by co-option of exapted ancestral enzymes. *Proceedings of the National Academy of Sciences* 113, 10613-10618.

Kapteyn, J., Qualley, A. V., Xie, Z., Fridman, E., Dudareva, N., Gang, D. R., 2007. Evolution of cinnamate/p-coumarate carboxyl methyltransferases and their role in the biosynthesis of methylcinnamate. *The Plant Cell* 19, 3212-3229.

Kato, M., Mizuno, K., Crozier, A., Fujimura, T., Ashihara, H., 2000. Plant biotechnology: Caffeine synthase gene from tea leaves. *Nature* 406, 956.

Kato, M., Mizuno, K., Fujimura, T., Iwama, M., Irie, M., Crozier, A., Ashihara, H., 1999. Purification and characterization of caffeine synthase from tea leaves. *Plant physiology* 120, 579-586.

Kessler, D., Baldwin, I. T., 2007. Making sense of nectar scents: the effects of nectar secondary metabolites on floral visitors of *Nicotiana attenuata*. *The Plant Journal* 49, 840-854.

Koeduka, T., Kajiyama, M., Suzuki, H., Furuta, T., Tsuge, T., Matsui, K., 2016. Benzenoid biosynthesis in the flowers of *Eriobotrya japonica*: molecular cloning and functional characterization of p-methoxybenzoic acid carboxyl methyltransferase. *Planta* 244, 725-736.

Köllner, T. G., Lenk, C., Zhao, N., Seidl-Adams, I., Gershenzon, J., Chen, F., Degenhardt, J., 2010. Herbivore-induced SABATH methyltransferases of maize that methylate anthranilic acid using S-adenosyl-L-methionine. *Plant physiology* 153, 1795-1807.

Koo, Y. J., Kim, M. A., Kim, E. H., Song, J. T., Jung, C., Moon, J.-K., Kim, J.-H., Seo, H. S., Song, S. I., Kim, J.-K., 2007. Overexpression of salicylic acid carboxyl methyltransferase reduces salicylic acid-mediated pathogen resistance in *Arabidopsis thaliana*. *Plant molecular biology* 64, 1-15.

Lin, J., Mazarei, M., Zhao, N., Zhu, J. J., Zhuang, X., Liu, W., Pantalone, V. R., Arelli, P. R., Stewart Jr, C. N., Chen, F., 2013. Overexpression of a soybean salicylic acid methyltransferase gene confers resistance to soybean cyst nematode. *Plant biotechnology journal* 11, 1135-1145.

Ludwiczuk, A., Odrzykoski, I. J., Asakawa, Y., 2013. Identification of cryptic species within liverwort *Conocephalum conicum* based on the volatile components. *Phytochemistry* 95, 234-241.

Matasci, N., Hung, L.-H., Yan, Z., Carpenter, E. J., Wickett, N. J., Mirarab, S., Nguyen, N., Warnow, T., Ayyampalayam, S., Barker, M., 2014. Data access for the 1,000 Plants (1KP) project. *GigaScience* 3, 17.

Maury, S., Geoffroy, P., Legrand, M., 1999. Tobacco O-methyltransferases involved in phenylpropanoid metabolism. The different caffeoyl-coenzyme A/5-hydroxyferuloyl-coenzyme A 3/5-O-methyltransferase and caffeic acid/5-hydroxyferulic acid 3/5-O-methyltransferase classes have distinct substrate specificities and expression patterns. *Plant Physiology* 121, 215-224.

Meur, G., Shukla, P., Dutta-Gupta, A., Kirti, P., 2015. Characterization of *Brassica juncea*–*Alternaria brassicicola* interaction and jasmonic acid carboxyl methyl transferase expression. *Plant gene* 3, 1-10.

Murata, J., Roepke, J., Gordon, H., De Luca, V., 2008. The leaf epidermome of *Catharanthus roseus* reveals its biochemical specialization. *The Plant Cell* 20, 524-542.

Murfitt, L. M., Kolosova, N., Mann, C. J., 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*. Archives of biochemistry and biophysics 382, 145-151.

Negre, F., Kish, C. M., Boatright, J., Underwood, B., Shibuya, K., Wagner, C., Clark, D. G., Dudareva, N., 2003. Regulation of methylbenzoate emission after pollination in snapdragon and petunia flowers. The Plant Cell 15, 2992-3006.

Negre, F., Kolosova, N., Knoll, J., Kish, C. M., 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. Archives of biochemistry and biophysics 406, 261-270.

Noel, J. P., Austin, M. B., Bomati, E. K., 2005. Structure–function relationships in plant phenylpropanoid biosynthesis. Current opinion in plant biology 8, 249-253.

Noel, J. P., Dixon, R. A., Pichersky, E., Zubieta, C., Ferrer, J.-L., 2003. Chapter two Structural, functional, and evolutionary basis for methylation of plant small molecules. Recent advances in phytochemistry, vol. 37. Elsevier, pp. 37-58.

Qi, J., Li, J., Han, X., Li, R., Wu, J., Yu, H., Hu, L., Xiao, Y., Lu, J., Lou, Y., 2016. Jasmonic acid carboxyl methyltransferase regulates development and herbivory-induced defense response in rice. Journal of integrative plant biology 58, 564-576.

Qin, G., Gu, H., Zhao, Y., Ma, Z., Shi, G., Yang, Y., Pichersky, E., Chen, H., Liu, M., Chen, Z., 2005. An indole-3-acetic acid carboxyl methyltransferase regulates *Arabidopsis* leaf development. *The Plant Cell* 17, 2693-2704.

Ogawa, M., Herai, Y., Koizumi, N., Kusano, T., Sano, H., 2001. 7-Methylxanthine Methyltransferase of Coffee Plants GENE ISOLATION AND ENZYMATIC PROPERTIES. *Journal of Biological Chemistry* 276, 8213-8218.

Petronikolou, N., Hollatz, A. J., Schuler, M. A., Nair, S. K., 2018. Loganic Acid Methyltransferase: Insights into the Specificity of Methylation on an Iridoid Glycoside. *ChemBioChem* 19, 784-788.

Pott, M. B., Hippauf, F., Saschenbrecker, S., Chen, F., Ross, J., Kiefer, I., Slusarenko, A., Noel, J. P., Pichersky, E., Effmert, U., 2004. Biochemical and structural characterization of benzenoid carboxyl methyltransferases involved in floral scent production in *Stephanotis floribunda* and *Nicotiana suaveolens*. *Plant physiology* 135, 1946-1955.

Pott, M. B., Pichersky, E., Piechulla, B., 2002. Evening specific oscillations of scent emission, SAMT enzyme activity, and SAMT mRNA in flowers of *Stephanotis floribunda*. *Journal of Plant Physiology* 159, 925-934.

Preuß, A., Augustin, C., Figueroa, C. R., Hoffmann, T., Valpuesta, V., Sevilla, J. F., Schwab, W., 2014. Expression of a functional jasmonic acid carboxyl methyltransferase is negatively correlated with strawberry fruit development. *Journal of plant physiology* 171, 1315-1324.

Qin, G., Gu, H., Zhao, Y., Ma, Z., Shi, G., Yang, Y., Pichersky, E., Chen, H., Liu, M., Chen, Z., 2005. An indole-3-acetic acid carboxyl methyltransferase regulates *Arabidopsis* leaf development. *The Plant Cell* 17, 2693-2704.

Ross, J. R., Nam, K. H., D'Auria, J. C., 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. *Archives of biochemistry and biophysics* 367, 9-16.

Seo, H. S., Song, J. T., Cheong, J.-J., Lee, Y.-H., Lee, Y.-W., Hwang, I., Lee, J. S., Do Choi, Y., 2001. Jasmonic acid carboxyl methyltransferase: a key enzyme for jasmonate-regulated plant responses. *Proceedings of the National Academy of Sciences* 98, 4788-4793.

Shinoda, T., Itoyama, K., 2003. Juvenile hormone acid methyltransferase: a key regulatory enzyme for insect metamorphosis. *Proceedings of the National Academy of Sciences* 100, 11986-11991.

Singh, B., Sharma, R. A., 2015. Plant terpenes: defense responses, phylogenetic analysis, regulation and clinical applications. *3 Biotech* 5, 129-151.

Szweykowski, J., Buczkowska, K., Odrzykoski, I., 2005. *Conocephalum salebrosum* (Marchantiopsida, Conocephalaceae)—a new Holarctic liverwort species. *Plant systematics and evolution* 253, 133-158.

Tieman, D., Zeigler, M., Schmelz, E., Taylor, M. G., Rushing, S., Jones, J. B., Klee, H. J., 2010. Functional analysis of a tomato salicylic acid methyl transferase and its role in synthesis of the flavor volatile methyl salicylate. *The Plant Journal* 62, 113-123.

Tohge, T., Wendenburg, R., Ishihara, H., Nakabayashi, R., Watanabe, M., Sulpice, R., Hoefgen, R., Takayama, H., Saito, K., Stitt, M., 2016. Characterization of a recently evolved flavonol-phenylacyltransferase gene provides signatures of natural light selection in Brassicaceae. *Nature communications* 7, 12399.

Toong, Y. C., Schooley, D. A., Baker, F. C., 1988. Isolation of insect juvenile hormone III from a plant. *Nature* 333, 170.

Toyota, M., 2000. Phytochemical study of liverworts *Conocephalum conicum* and *Chiloscyphus polyanthos*. *Yakugaku Zasshi: Journal of the Pharmaceutical Society of Japan* 120, 1359-1372.

Toyota, M., Saito, T., Matsunami, J., Asakawa, Y., 1997. A comparative study on three chemo-types of the liverwort *Conocephalum conicum* using volatile constituents. *Phytochemistry* 44, 1265-1270.

Uefuji, H., Ogita, S., Yamaguchi, Y., Koizumi, N., Sano, H., 2003. Molecular cloning and functional characterization of three distinct N-methyltransferases involved in the caffeine biosynthetic pathway in coffee plants. *Plant physiology* 132, 372-380.

Varbanova, M., Yamaguchi, S., Yang, Y., McKelvey, K., Hanada, A., Borochoy, R., Yu, F., Jikumaru, Y., Ross, J., Cortes, D., 2007. Methylation of gibberellins by *Arabidopsis* GAMT1 and GAMT2. *The Plant Cell* 19, 32-45.

Wink, M., 2010. Introduction: biochemistry, physiology and ecological functions of secondary metabolites. *Annual plant reviews volume 40: Biochemistry of plant secondary metabolism*, 1-19.

Wood, W. F., Lancaster, W. C., Fisher, C. O., Stotler, R. E., 1996. Trans-methyl cinnamate: The major volatile from some populations of the liverwort, *Conocephalum conicum*. *Phytochemistry* 42, 241-242.

Yang, Y., Yuan, J.S., Ross, J., Noel, J.P., Pichersky, E., Chen, F., 2006. An *Arabidopsis thaliana* methyltransferase capable of methylating farnesoic acid. *Arch. Biochem. Biophys.* 448, 123-132.

Ye, Z.-H., Kneusel, R. E., Matern, U., Varner, J. E., 1994. An alternative methylation pathway in lignin biosynthesis in *Zinnia*. *The Plant Cell* 6, 1427-1439.

Yamaguchi, S., 2008. Gibberellin metabolism and its regulation. *Annu. Rev. Plant Biol.* 59, 225-251.

Yoneyama, N., Morimoto, H., Ye, C.-X., Ashihara, H., Mizuno, K., Kato, M., 2006. Substrate specificity of N-methyltransferase involved in purine alkaloids synthesis is

dependent upon one amino acid residue of the enzyme. *Molecular genetics and genomics* 275, 125-135.

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

Zhao, N., Ferrer, J.-L., Moon, H. S., Kapteyn, J., Zhuang, X., Hasebe, M., Stewart Jr, C. N., Gang, D. R., Chen, F., 2012. A SABATH Methyltransferase from the moss *Physcomitrella patens* catalyzes S-methylation of thiols and has a role in detoxification. *Phytochemistry* 81, 31-41.

Zhao, N., Ferrer, J.-L., Ross, J., Guan, J., Yang, Y., Pichersky, E., Noel, J. P., 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 physiology* 146, 455-467.

Zhao, N., Guan, J., Lin, H., Chen, F., 2007. Molecular cloning and biochemical characterization of indole-3-acetic acid methyltransferase from poplar. *Phytochemistry* 68, 1537-1544.

Zhao, N., Yao, J., Chaiprasongsuk, M., Li, G., Guan, J., Tschaplinski, T. J., Guo, H., Chen, F., 2013. Molecular and biochemical characterization of the jasmonic acid methyltransferase gene from black cottonwood (*Populus trichocarpa*). *Phytochemistry* 94, 74-81.

Zubieta, C., He, X.-Z., Dixon, R. A., Noel, J. P., 2001. Structures of two natural product methyltransferases reveal the basis for substrate specificity in plant O-methyltransferases. *Nature Structural & Molecular Biology* 8, 271.

Appendix A: Figures and Tables

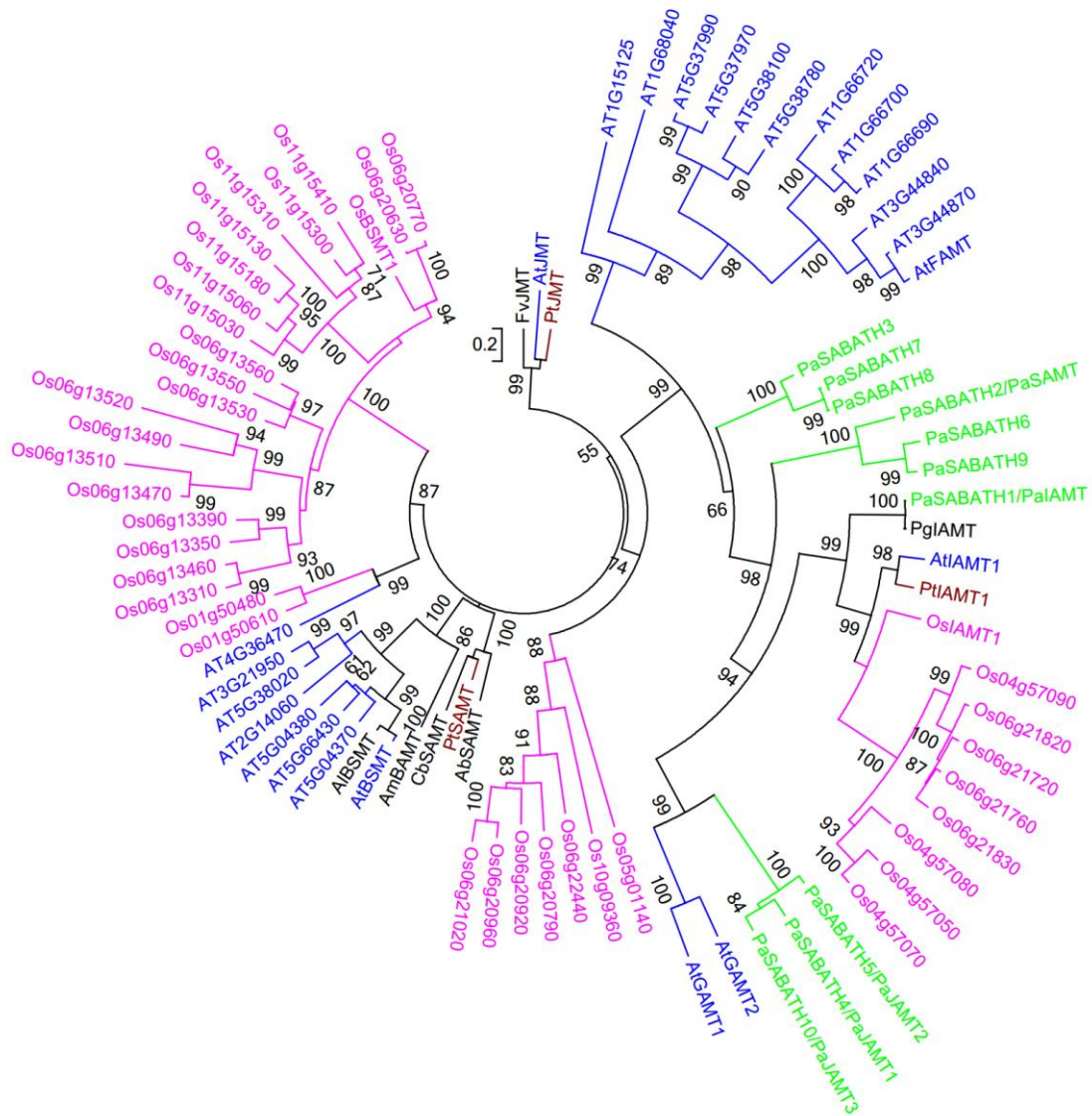


Figure A-1. Sequences are from *A. thaliana* (blue), *O. sativa* (pink), *P. abies* (green) and other species. Branches were drawn to scale with the bar indicating 0.2 substitutions per site.

Chapter II. Biosynthesis of Methyl (*E*)-cinnamate in the Liverwort

***Conocephalum salebrosum* and Evolution of Cinnamic Acid**

Methyltransferase

Adapted from:

Chi Zhang, Xinlu Chen, Barbara Crandall-Stotler, Ping Qian, Tobias G. Köllner, Hong Guo and Feng Chen (2019) Biosynthesis of methyl (*E*)-cinnamate in the liverwort *Conocephalum salebrosum* and evolution of cinnamic acid methyltransferase. *Phytochemistry* 164: 50-59

Abstract

Methyl (*E*)-cinnamate is a specialized metabolite that occurs in a variety of land plants. In flowering plants, it is synthesized by cinnamic acid methyltransferase (CAMT) that belongs to the SABATH family. While rarely reported in bryophytes, methyl (*E*)-cinnamate is produced by some liverworts of the *Conocephalum conicum* complex, including *C. salebrosum*, which produces methyl (*E*)-cinnamate as the dominant compound. To characterize methyl (*E*)-cinnamate biosynthesis in *C. salebrosum*, six full-length SABATH genes, which were designated *CsSABATH1-6*, were cloned. These six genes showed different levels of expression in the thallus of *C. salebrosum*. Next, *CsSABATH1-6* were expressed in *Escherichia coli* to produce recombinant proteins, which were tested for methyltransferase activity with cinnamic acid and a few related compounds as substrates. Among the six SABATH proteins, *CsSABATH6* exhibited the highest level of activity with cinnamic acid. It was renamed *CsCAMT*. The apparent *K_m* value of *CsCAMT* using cinnamic acid as substrate was determined to be 50.5 μ M. In contrast, *CsSABATH4* was demonstrated to function as salicylic acid methyltransferase and was renamed *CsSAMT*. Interestingly, the *CsCAMT* gene from a sabinene-dominant chemotype of *C. salebrosum* is identical to that of the methyl (*E*)-cinnamate-dominant chemotype. Structure models for *CsCAMT*, *CsSAMT* and one flowering

plant CAMT (ObCCMT1) in complex with (*E*)-cinnamic acid and salicylic acid were built, which provided structural explanations to substrate specificity of these three enzymes. In phylogenetic analysis, CsCAMT and ObCCMT1 were in different clades, implying that methyl (*E*)-cinnamate biosynthesis in bryophytes and flowering plants originated through convergent evolution.

Keywords: *Conocephalum salebrosum*, Liverworts, Conocephalaceae, SABATH methyltransferase, specialised metabolism, convergent evolution

1. Introduction

Plants produce diverse volatile specialized metabolites that include carboxyl methyl esters. These methyl esters play important roles in various biological processes (Chen et al., 2003a; Knudsen et al. 1993; Seo et al., 2001). Methyl (*E*)-cinnamate, methyl ester of (*E*)-cinnamic acid, is one such compound. Its odor is defined as “balsamic, strawberry, fruity, cherry”. Methyl (*E*)-cinnamate occurs as a scent compound in many flowers such as orchids (Kaiser, 1993) and *Narcissus* species (Arai, 1994). It is also made by some fruits such as strawberry (Gomes da Silva and Chaves das Neves, 1999) and plum (Ismail et al., 1980). Its occurrence in leaves of some plants is also known, such as in basil (Viña and Murillo, 2003) and *Eucalyptus* species (Curtis et al., 1990). In flowers and fruits, methyl (*E*)-cinnamate is probably involved in attracting pollinators and seed dispersers, respectively (Dodson et al., 1969; Eltz and Lunau, 2005). In leaves, this compound most likely functions as a chemical defense (Hattori et al., 1992). It has been shown to inhibit the growth of pathogenic fungi and bacteria (Ali et al., 2010). The molecular basis of methyl (*E*)-cinnamate biosynthesis has so far only been studied in basil (Kapteyn et al., 2007), where cinnamic acid/*p*-coumaric acid methyltransferases (ObCCMTs) catalyze the formation of methyl (*E*)-cinnamate using *S*-adenosyl-L-methionine (SAM) as methyl donor (Kapteyn et al., 2007).

ObCCMTs belong to the family of methyltransferases known as SABATH (D'Auria et al., 2003). Most substrates of SABATH methyltransferases are important phytohormones including indole-3-acetic acid, gibberellins, salicylic acid and jasmonic acid (Chen et al., 2003a; Qin et al., 2005; Seo et al., 2001; Varbanova et al., 2007). Other substrates of SABATH methyltransferases include benzoic acid (Murfitt et al., 2000), *p*-methoxybenzoic

acid (Koeduka et al., 2016), farnesoic acid (Yang et al., 2006), anthranilic acid (Köllner et al., 2010), nicotinic acid (Hippauf et al., 2010) and theobromine (Kato et al., 2000; McCarthy and McCarthy, 2007). Such diverse substrates within a same protein family make SABATH proteins useful models for studying substrate specificity evolution (Huang et al., 2012). Within the *SABATH* family, indole-3-acetic acid methyltransferase gene (*IAMT*) has been proposed to be an ancient member in seed plants (Zhao et al., 2008). *ObCCMTs* are closely related to *IAMT* (Kapteyn et al., 2007), making it tempting to speculate that *ObCCMTs* may have evolved from *IAMT* through gene duplication and functional divergence.

Liverworts are considered extant plants that are most closely related to ancestral land plants (Bowman et al., 2017). Like flowering plants, liverworts make a diverse array of specialised metabolites, especially terpenoids (Chen et al., 2018). While rarely reported, methyl (*E*)-cinnamate is known to be made by liverworts within the *Conocephalum conicum* complex (Ghani et al., 2016; Harinantenaina et al., 2007; Toyota, 2000; Wood et al., 1996). Species of the *Conocephalum conicum* complex, known as great scented liverworts, are widely distributed in North America, Europe and East Asia (Szweykowski et al., 2005). They can be easily recognized by their snake skin-like surface and aromatic odor (Atherton et al., 2010). The *C. conicum* complex is comprised of two morphologically distinct species, namely *C. conicum* and *C. salebrosum*, and multiple cryptic species (Szweykowski et al., 2005). Although when Szweykowski et al. (2005) named *C. salebrosum*, they indicated that it mostly corresponded to *C. conicum* genotype S, North American specimens with genotype A from North Carolina, were also included. This species level relationship between genotypes A and S (= *C. salebrosum*) was further confirmed in Ludwiczuk et al. (2013). In an early study

of presumed *C. conicum* populations in the in United States, methyl (*E*)-cinnamate was identified as a major constituent from a sample of unknown natural origin as well as a population from Southern Illinois, but was not detected in the extracts of wild *C. conicum* from coastal northern California (Wood et al., 1996). In subsequent studies by a Japanese group, three chemotypes of *C. conicum* were identified: methyl (*E*)-cinnamate-dominant, sabinene-dominant, and bornyl acetate-dominant (Toyota, 2000; Toyota et al., 1997). In this study, we aim at elucidating the molecular basis of methyl (*E*)-cinnamate biosynthesis in the North American member of the *C. conicum* complex (i.e., *C. salebrosum*), comparing it to that in flowering plants, and gaining understanding about the molecular basis for methyl (*E*)-cinnamate variations among *C. conicum* populations.

2. Materials and methods

2.1. Plant culture and axenic culture

Two populations of *C. salebrosum* were used this study. One population was collected from Illinois: Williamson Co., Rocky Bluff Nature Preserve, nr. Devils Kitchen Lake, growing over moist sandstone rocks; 37°38'32.34" N, 89°05'51.25" W; elevation 157 m, 16 Oct. 2017, *J. Henry s.n.* [F]). This population is methyl (*E*)-cinnamate-dominant. The other population was collected from Tennessee: Knox Co., Campbell station park, nr. North Fork Turkey Creek, growing over moist rocks; 35°53'13.5"N 84°10'02.4"W and is sabinene-dominant. The Illinois plants collected from the field were rinsed with running water until thalli were clean. Explants were sterilized using 70% ethanol solution for 2 mins, then rinsed

3 times with sterilized distilled water. The explants were further surface-sterilized using 10% bleach(v/v) for 5 mins, then rinsed 3 times using sterilized distilled. The sterilized explants were transferred onto Hatcher medium which was prepared following the protocol of Dr. Crandall-Stotler's Lab (<http://bryophytes.plant.siu.edu>).

2.2. Organic extraction and headspace collection

Tissues were used for organic extraction as described (Chen et al., 2003b) using ethyl acetate. The 5 μ L extracts were used for GC-MS analysis or stored in -20°C. A SPME fiber coated with 100- μ m polydimethylsiloxane was inserted into the headspace of the axenic culture to start volatile collection. After 1.5 h, the SPME fiber was retracted and inserted into the injector port for GC-MS analysis.

2.3. GC-MS analysis

Samples from either organic extraction or headspace collection were analyzed with Shimadzu QP5050A GC-MS with a Restek R xi-5Sil MS column. The compound separation using GC was carried out with the following conditions: carrier gas Helium with the flow rate of 5mL/min, a splitless injection of 5 μ L, and a temperature gradient of 5°C/min from 40°C to 240°C after an initial 3-min hold. Compounds were identified based on comparisons to standard substances or internal MS spectrum database.

2.4. Database search and phylogenetic analysis

To identify putative *SABATH* genes from the *C. conicum* complex, the protein sequences of sample ID ILBQ from OneKP database (Matasci et al., 2014) were searched using Pfam profile PF03492 (Finn et al., 2015) via the hmmsearch (Finn et al., 2011). For phylogeny reconstruction, MAFFT (version 7.369b, under L-INS-I strategy) was used to perform multiple protein sequence alignments (Katoh and Standley, 2013). FastTree (version 2.1.10, under the JTT + CAT model with 1000 resamples) was used to construct Approximately-Maximum-Likelihood trees (Price et al., 2010). The trees were further edited using MEGA (version 7.0.21) (Kumar et al., 2016).

2.5. Cloning full length cDNA of CsSABATHs

Total RNA was extracted from vegetative thalli using a RNeasy Plant Mini Kit (Qiagen, Valencia, CA) and reverse-transcribed into first strand cDNA in a 15 µL reaction volume using the First-strand cDNA Synthesis Kit (Amersham Biosciences, Piscataway, NJ) as previously described (Chen et al., 2003a). *CsSABATH* full-length cDNAs were amplified using forward primer and reverse primers corresponding with the 5' and 3' ends of the *CsSABATH* coding region (Table S2), respectively. The PCR was carried out using the following program: 94 °C for 3 min followed by 32 cycles at 94 °C for 30 s, 58 °C for 30 s, 72 °C for 1 min 30 s, and followed by a final extension at 72 °C for 10 min. Products of PCR were isolated on 1.0% agarose gel and purified using QIAquick Gel Extraction kit (Qiagen, Valencia, CA). The cDNAs were then cloned into a pEXP-5-CT/TOPO vector following the vendor's protocol (Invitrogen, Carlsband, CA). The cloned cDNAs were fully sequenced.

2.6. Semi-quantitative reverse-transcription PCR

Semi-quantitative reverse-transcription PCR (RT-PCR) was performed as previously described (Chen et al., 2003a). The cDNA from a single reverse-transcription reaction was used as template and 6 pairs of specific primers (Table S2) were used in the PCR reaction. PCR was performed using same program as described in Section 5.5 except that the cycle number was 25. 30 μ L PCR products and 3 μ L 1kb ladder marker were loaded in the agarose gel.

2.7. Purification of CsSABATHs expressed in *E. coli*

CsSABATHs were subcloned into the vector of pET32a (Invitrogen, Carlsband, CA). To express the CsSABATH protein, the protein expression construct was transformed into the *E. coli* strain BL21 (DE3) CodonPlus (Stratagene, La Jolla, CA) and cultured under 25 °C. Isopropyl β -D-1-thiogalactopyranoside (IPTG) was added into bacterial culture at a concentration of 500 μ M to induce protein expression. The bacterial cells were lysed by sonication after 18h. His-tagged CsSABATH protein was enriched from the *E. coli* cell lysate using Ni-NTA (Invitrogen, Carlsband, CA). An empty pET32a vector without any insert was set to be a negative control.

2.8. Radiochemical methyltransferase activity assay

CsSABATH enzyme assays were carried out following a radiochemical protocol as previously described (Zhao et al., 2007). Optimum pH, thermostability, the effect of cations and kinetic parameters were measured using the method of Zhao et al. (2013) based on a

radiochemical assay. Final concentration of each tested substrate in assay was 1 μ M. Three independent assays were performed to calculate mean and standard deviation for each measurement. A non-radiochemical assay followed by extraction and GC-MS analysis was performed to determine the chemical identity of the product of CsSABATH methylation.

2.9. Protein structure modeling

Based on the crystallographic structure (PDB code: 1M6E) of CbSAMT, the homology models of CsCAMT, CsSAMT and ObCCMT1 were built by using the tools in the MOE program [Molecular Operating Environment (MOE), version 2013.08 (2015) Chemical Computing Group Inc., Montreal]. SAH in each of the models was built based on the superposition of the model with the X-ray structure of the CbSAMT complex containing both SAH and SA. (*E*)-Cinnamic acid molecule was manually docked into the active site in each case based on the superposition of the (*E*)-cinnamic acid carboxylate group with that of SA in the CbSAMT complex that is close to the sulfur atom of SAH. Such correct binding mode of the carboxyl moiety in the enzyme-substrate complex is expected to be necessary for the enzyme's function as methyltransferases (Yao et al., 2012, Yao et al., 2010). The similar idea has been used in our earlier study of other plants (Qian et al., 2015).

3. Results

3.1. Production of axenic culture of *C. salebrosum* and chemical profiling

The sample collected from Illinois (IL) morphologically matched *C. salebrosum*, a segregate species of *C. conicum* (Szweykowski et al., 2005). This population is the chemotype with methyl (*E*)-cinnamate as the dominant compound described by Wood et al. (1996), and was therefore chosen as the model chemotype for most of the experiments in this study. Live plants were collected in the field and grown in axenic culture. Chemical profiling analysis was performed using gas chromatography–mass spectrometry (GC-MS) to determine the volatile components in the axenically grown gametophytes of this species of the *C. conicum* complex. From organic extraction, a total of 10 volatiles were detected (Figure B-1A), of which methyl (*E*)-cinnamate was the dominant component, accounting for 50.5% of the total volatiles. A monoterpene sabinene is the second most abundant with a concentration about 20% of that of methyl (*E*)-cinnamate. The thallus of *C. salebrosum*, when growing in the wild, possesses a characteristic aromatic odor. To determine whether *C. salebrosum* grown axenically on culture medium emits volatiles, we performed headspace collection of explants of *C. salebrosum*, which was analyzed by GC-MS. Three dominant compounds were detected (Figure B-1B). All of them were also detected by organic extraction. Similar to organic extraction, methyl (*E*)-cinnamate was the most abundant compound in the volatile bouquet of *C. salebrosum* (Figure B-1B).

3.2. Isolation of SABATH genes from *C. salebrosum* and gene expression analysis

We hypothesized that in the *C. salebrosum* methyl (*E*)-cinnamate is synthesized by the action of SABATH methyltransferase. To identify SABATH genes, we analyzed the transcriptome of *C. conicum* prepared from vegetative thalli

(<https://sites.google.com/a/ualberta.ca/onekp>), which was produced by the 1KP consortium (Matasci et al., 2014). Via a HMMR search (Finn et al., 2011) using Methyltransf_7 Pfam profile (Finn et al., 2015) as query, a total of nine unigenes encoding SABATH proteins were identified. Six of them appear to be in full-length when compared to other SABATH proteins.

The chemotype of the specimen used for transcriptome analysis was unknown. To determine whether all six full-length SABATH genes are expressed in the methyl (*E*)-cinnamate-dominant chemotype from IL, we renamed them to *CsSABATH1-6* (Table S1) and performed gene expression analysis. By semi-quantitative reverse transcription polymerase chain reaction (RT-PCR) analysis, transcripts for five of the six genes can be detected from thallus tissue, the exception being *CsSABATH1* (Figure B-2). *CsSABATH4* showed the highest level of expression and *CsSABATH2* and *CsSABATH6* showed relatively high levels of expression. *CsSABATH3* and *CsSABATH5* showed relatively low levels of expression. The lengths of six proteins range from 383 to 443 amino acids (Figure B-3), Sequence similarities among *CsSABATHs* range from 37% to 52%. Their similarities to ObCCMT1, one known CAMT from basil, are between 36% and 48%. The amino acid residues for binding the methyl donor SAM among *CsSABATHs* and ObCCMT1 are identical. The residues that interact with the carboxyl moiety of the substrate are also conserved. However, the residues that interact with the aromatic moiety are significantly different (Figure D-3).

3.3. Biochemical characterization of CsSABATHs expressed in *E. coli*

Gene expression analysis (Figure B-2) suggested that *CsSABATH2*, *CsSABATH4* and *CsSABATH6* are the most probable candidates as *CAMT*. To validate this prediction, we characterized the proteins encoded by all six *CsSABATH* genes using a biochemical approach. Full-length cDNA for each of the six *CsSABATH* genes was cloned by RT-PCR into a protein expression vector without any tag and expressed in *E. coli* to produce recombinant proteins. Each of the recombinant *CsSABATH* enzymes was assayed with (*E*)-cinnamic acid and an additional nine carboxyl acids: indole-3-acetic acid, gibberellin A3, salicylic acid, jasmonic acid, anthranilic acid, nicotinic acid, benzoic acid, *p*-coumaric acid and abscisic acid (Table B-1). Among the six *CsSABATH*s, *CsSABATH6* was highly active using (*E*)-cinnamic acid as substrate. It also showed activity with benzoic acid and *p*-coumaric acid, which was about 32% and 18% of the activity with (*E*)-cinnamic acid. *CsSABATH4* also showed some activity with (*E*)-cinnamic acid, but the activity was very low. Instead, *CsSABATH4* was most active with salicylic acid as substrate. Its activity with (*E*)-cinnamic acid was only about 6% of that with salicylic acid. *CsSABATH6* and *CsSABATH4* were renamed *CsCAMT* (GenBank accession no. MK673137) and *CsSAMT* (GeneBank accession no. MK673138), respectively. *CsSABATH1*, *CsSABATH2*, *CsSABATH3* and *CsSABATH5* did not show activity with any of the substrates tested.

3.4. Biochemical properties of *CsCAMT*

With the identification of *CsCAMT* in *C. salebrosum*, next, we performed biochemical assays to determine its detailed biochemical properties using a purified *CsCAMT*

containing a N-terminal His-tag. The optimum pH for CsCAMT was determined to be 6.5 (Figure B-4A). For temperature stability, CsCAMT was relatively stable when the temperatures were below 42 °C (Figure B-4B). CsCAMT activity can be affected by various ions (Figure B-4C). Na⁺ and Mg²⁺ ions slightly stimulated the activity of CsCAMT. K⁺, NH₄⁺, Ca²⁺, and Mn²⁺ had a mild inhibitory effect. Fe²⁺, Zn²⁺, Cu²⁺, or Fe³⁺ completely inhibited the activity of CsCAMT (Figure B-4C). The kinetic property of CsCAMT was also measured. The apparent *K_m* value of CsCAMT using cinnamic acid as substrate was determined to be 50.5 μM (Figure B-4D).

3.5. Comparison of CAMT genes in different chemotypes of the *C. conicum* complex

As previously reported, some populations of *C. conicum* showed a monoterpene-dominant chemical profile (Craft et al., 2016; Ludwiczuk et al., 2013; Toyota, 2000). To gain some understanding into the molecular basis of such variations, we collected a population of *C. salebrosum* from Tennessee (TN). This population exhibited sabinene-dominant chemotype (Figure 8). Next, a putative ortholog of *CsCAMT* gene was cloned from the sabinene-dominant plants and fully sequenced. It is identical to the *CsCAMT* gene isolated from methyl (*E*)-cinnamate-dominant plants.

3.6. Structural modeling of *CsCAMT*, *CsSAMT* and *ObCCMT1*

The structure models for *CsCAMT*, *CsSAMT* and one basil CAMT (*ObCCMT1*) (Kapteyn et al., 2007) were built using the X-ray structure of *CbSAMT* (Zubieta et al., 2003) as the template. The active sites of the models for *CsCAMT* and *CsSAMT* in complex with

(*E*)-cinnamic acid are plotted in Figure B-6A and 6B, respectively. In each of these cases, the model was first superposed with the protein structure of CbSAMT that is complexed with salicylic acid and *S*-adenosyl-L-homocysteine (SAH) at the active site. (*E*)-Cinnamic acid was then docked into the active site such that its cinnamate carboxylate group was superposed with that of salicylic acid to form the reactive configuration for the methyl transfer. Figure 6A shows that (*E*)-cinnamic acid can fit well into the active site of CsCAMT with its carboxyl moiety located at the suitable position for accepting the methyl group from SAM. The result is consistent with the experimental observation that (*E*)-cinnamic acid is the preferred substrate for CsCAMT. In contrast, when (*E*)-cinnamic acid was docked into the active site of CsSAMT (Figure B-6B) in the similar fashion, the ring of (*E*)-cinnamic acid becomes too close to some of the residues of CsSAMT. For instance, the distances between two of the carbon atoms on the ring are only about 2.4 Å and 2.1 Å to the C_α and C_β carbons of Ile237, respectively. Such close contacts would create significant repulsions between the enzyme and (*E*)-cinnamic acid and is likely to destroy the reactive configuration for the methyl transfer.

The active sites of the models for CsCAMT and CsSAMT in complex with salicylate are shown in Figure B-6C and 6D. Figure B-6D shows that, unlike (*E*)-cinnamic acid, salicylic acid can fit into the active site of CsSAMT without the close contacts observed for (*E*)-cinnamic acid. This is expected, as salicylic acid is significantly smaller than (*E*)-cinnamic acid. Trp159 of CsCAMT (Figure B-6C) and Trp157 of CsSAMT (Figure B-6D) are equivalent to Trp151 of CbSAMT, one of the two key residues for ensuring proper orientation and proximity of salicylic acid to SAM for the methylation reaction in CbSAMT (Zubieta et al., 2003). It is of interest to note from Figure B-6C that, unlike in CbSAMT (Trp151) and

CsSAMT (Trp157), Trp159 in CsCAMT seems to have the potential to interact not only with the carboxyl group of salicylic acid, but also with the 2-OH group. Such additional interaction with the 2-OH group might distort the optimal arrangement between the methyl acceptor and donor and therefore interferes with the methyl transfer. Consistent with this suggestion, Table B-1 shows that CsCAMT does not have activity on salicylic acid, even though it is active on benzoic acid.

In Figure B-6E, the structure model for ObCCMT1 complexed with (*E*)-cinnamic acid is plotted on the left, and the superposition of the active sites for CsCAMT and ObCCMT1 is shown in Figure B-6F. Similar to the case of CsCAMT, (*E*)-cinnamic acid can fit into the active site of ObCCMT1 without close contacts with the active site residues (Figure B-6E), in agreement with an earlier study on this enzyme (Kapteyn et al., 2007). It is of interest to note from Figure 6F that a number of the active site residues for ObCCMT1/CsCAMT are conserved in these two enzymes (e.g., Tyr29/Tyr20, Gln36/Gln27, Asp68/Asp59, Ser72/Ser63, Asn76/Asn67, Asp108/Asp98, Leu109/Leu99, Tyr141/Tyr139, Phe157/Phe155, His160/His158, Trp161/Trp159 and Tyr272/Tyr268).

3.7. Relatedness of CsCAMT with other SABATH methyltransferases

An approximately-maximum-likelihood phylogenetic tree was constructed using the six CsSABATHs, three ObCCMTs and SABATH proteins from four sequenced plants that include *Arabidopsis thaliana*, *Oryza sativa*, *Physcomitrella patens* and *Selaginella moellendorffii* to understand their evolutionary relatedness (Figure B-7). All SABATHs in seed plant except gibberellic acid methyltransferases from *A. thaliana* (AtGAMTs) resolved

in a single clade. ObCCMTs from basil resolved in a clade with IAMTs from *A. thaliana* (AtIAMT) and *O. sativa* (OsIAMT). CsCAMT and CsSAMT resolved with other uncharacterized CsSABATHs from *C. salebrosum* and together they are resolved in a clade with AtGAMTs. Surprisingly, liverwort SABATHs do not resolve with the other major SABATHs in seed plants, including IAMTs and CCMTs.

4. Discussion

With the isolation and characterization of six *SABATH* genes from *C. salebrosum*, this is the first report on functional study of *SABATH* genes in liverworts. Particularly, we successfully identified *CsCAMT* (*CsSABATH6*) as the gene with the capability for synthesis of methyl (*E*)-cinanmate in *C. salebrosum* (Table B-1). In addition, *CsSABATH4* was determined to encode SAMT (Table B-1). CsCAMT has a higher affinity with the substrate (*E*)-cinnamic acid than ObCCMT1 from basil (Kapteyn et al., 2007). Its K_m value (50.5 μM) is about the half of that for ObCCMT1 (124 μM). CsCAMT and ObCCMT1 also showed different selectivity towards *p*-coumaric acid and benzoic acid. ObCCMT1 had about 30% and 10% of activities of (*E*)-cinnamic acid with *p*-coumaric acid and benzoic acid, respectively (Kapteyn et al., 2007). CsCAMT exhibited the opposite trend: its activities with *p*-coumaric acid and benzoic acid were 18% and 32% respectively of that with (*E*)-cinnamic acid (Table B-1). The identification of CsCAMT and CsSAMT (*CsSABATH4*) together with the availability of basil ObCCMT1 makes it possible to understand the structural basis underlying their substrate specificities.

We demonstrated from computer modeling that while the active site of CsCAMT can accommodate (*E*)-cinnamic acid and generate the reactive configuration for the methyl transfer (Figure B-6A), CsSAMT may not be able to do so. This is because some of the active site residues are too close to the ring of (*E*)-cinnamic acid (Figure B-6B), and such close contacts would create significant repulsions between CsSAMT and (*E*)-cinnamic acid, leading to the non-reactive configuration for the methyl transfer. This suggestion is supported by the lack of the activity of CsSAMT on (*E*)-cinnamic acid (see Table B-1). Similar discussion has been made in our earlier quantum mechanical/molecular mechanical (QM/MM) study for understanding the substrate specificity of CbSAMT (Yao et al., 2010). Interestingly, as observed in other characterized SABATH methyltransferases, our results on ObCCMT1/CsCAMT (Figure B-6F) show that both SAM/SAH-binding residues and carboxyl moiety interacting residues are very conservative, but residues that interact with the aromatic moiety of the substrate are different (Figure B-2). This indicates that similar interaction between active sites and substrate is achieved by the combination of different residues.

Based on our phylogenetic study (Figure B-7), ObCCMTs share a common ancestor with IAMTs from *Arabidopsis* and rice. Although IAMT in basil has not been identified, these results suggest that ObCCMTs and IAMTs, which belong to an ancient SABATH group in seed plants evolved from a common ancestral type. However, CsCAMT together with other CsSABATH members share a common ancestor with *Arabidopsis* GAMTs, which is beyond our expectation. This demonstrates that CsCAMT may have evolved from other uncharacterized SABATH methyltransferases and emerged independently from ObCCMT in

basil. The phylogeny in combination with structural modeling implies that the evolution of methyl (*E*)-cinnamate synthesis in the liverwort and flowering plants is the result of convergent evolution.

C. conicum and *C. salebrosum* usually inhabit shady moist environments, which can be hotbeds for microbe proliferation. Several studies showed that methyl (*E*)-cinnamate has antimicrobial activity (Gilles et al., 2010; Padalia et al., 2017). This could be a good reason for some populations of *C. salebrosum* steadily producing volatile methyl (*E*)-cinnamate in relatively large amounts (Figure B-1B). Aside from the capability of inhibiting the growth of fungi and bacteria, methyl (*E*)-cinnamate has larvicidal activity (Fujiwara et al., 2017). Larvae of the moth *Epimartyri pardella* have been reported to feed on *C. conicum* (Davis and Landry, 2012). This can be another rational cause for liverwort to defend itself from herbivores. Ghani et al. (2016) have shown that production of methyl (*E*)-cinnamate increases in populations of *C. conicum* from Japan when they are exposed to stressful growth conditions, suggesting that gene expression may be influenced by environment, while Szweykowsky et al. (2005) note that *C. salebrosum* often grows in habitats that vary from flooded to seasonally dry conditions.

The *in vitro* product of CsSAMT, methyl salicylate, has been shown to be a critical signaling molecule in plant defenses, particularly as the mobile signal in systematic acquired resistance against pathogens (Park et al., 2007; Shulaev et al., 1997). It may also affect plant defense against insects in both direct and indirect ways (Ament et al., 2010; Zhu and Park, 2005). The identification of CsSAMT indicates that methyl salicylate might play similar roles in liverworts. As in the case of CsCAMT in the phylogenetic analysis, CsSAMT is more closely

related to other CsSABATH members rather than to other characterized SAMTs or BSMTs (benzoic acid/salicylic acid MT) in seed plants. This observation implies that SAMTs might have evolved independently in liverworts and in flowering plants as well.

5. Conclusions

In this study, we identified and characterized the enzyme CsCAMT that is capable for the biosynthesis of methyl (*E*)-cinnamate in the liverwort *C. salebrosum*, a species in the *C. conicum* complex. CsCAMT belongs to the SABATH family of methyltransferases. The SABATH family in *C. salebrosum* has at least nine members based on transcriptome analysis, six of which were judged to be full-length and selected for characterization in this study. It is evident that members of the SABATH family in *C. salebrosum* have different substrate specificities, which is similar to the SABATH family in flowering plants such as *Arabidopsis* (Varbanova et al., 2007; Yang et al., 2006), rice (Zhao et al., 2008) and gymnosperms (Chaiprasongsuk et al., 2018). This implies repeated gene duplication followed by functional divergence of the SABATH family. Selected members of the *C. salebrosum* SABATH family are not conserved with their counterparts in flowering plants, evidenced by CAMT and SAMT in this study (Figure B-7). Particularly for CAMT, we can infer that this enzyme in *C. salebrsum* and in flowering plants evolved independently. There are strong similarities between the active site of CsCAMT and that of ObCCMT1, despite their relatively distant relatedness among the SABATH family, indicating similarities in catalytic mechanisms. There are a number of interesting future directions. With CsCAMT and ObCCMT1 resulting from convergent evolution, it will be interesting to ask whether they evolved from similar or

different ancestral activities. Because the two chemotypes of *C. salebrosum* contain the same *CAMT* gene, it will be interesting to ask what causes the difference in the contents of methyl (*E*)-cinnamate: different expression levels of CsCAMT or the different concentrations of its substrate cinnamic acid or both? With methyl (*E*)-cinnamate accumulated in the thallus and also emitted as a volatile compound, it will also be interesting to ask what are the biological functions CsCAMT and its product methyl (*E*)-cinnamate. In addition, for the four CsSABATHs that did not show activity with any of the ten carboxylic acids tested, their *in vivo* substrates and biological functions can be an important future study.

Reference

- Ab, N.G., Ludwiczuk, A., Ismail, N.H., Asakawa, Y., 2016. Volatile components of the stressed liverwort *Conocephalum conicum*. Nat. Prod. Commun. 11, 103-104.
- Ali, N., Rahmani, M., Shaari, K., Ali, A., Cheng Lian, G., 2010. Antimicrobial activity of *Cinnamomum impressicostatum* and *C. pubescens* and bioassay-guided isolation of bioactive (*E*)-methyl cinnamate. J. Biol. Sci. 10, 101-6.
- Ament, K., Krasikov, V., Allmann, S., Rep, M., Takken, F.L., Schuurink, R.C., 2010. Methyl salicylate production in tomato affects biotic interactions. Plant J. 62, 124-134.
- Arai, T., 1994. Volatile compounds of *Narcissus tazetta* var. *chinensis* flowers. Nippon Koryo Kyokai. 184, 105-111.
- Atherton, I., Bosanquet, S., Lawley, M., 2010. Mosses and liverworts of Britain and Ireland: a field guide. British Bryological Society, Plymouth UK.
- Bowman, J. L., Kohchi, T., Yamato, K.T., Jenkins, J., Shu, S., Ishizaki, K., Yamaoka, S., Nishihama, R., Nakamura, Y., Berger, F., Adam, C., Aki, S.S., Althoff, F., Araki, T., Arteaga-Vazquez, M.A., Balasubramanian, S., Barry, K., Bauer, D., Boehm, C.R., Briginshaw, L., Caballero-Perez, J., Catarino, B., Chen, F., Chiyoda, S., Chovatia, M., Davies, K.M., Delmans, M., Demura, T., Dierschke, T., Dolan, L., Dorantes-Acosta, A. E., Eklund, D.M., Florent, S.N., Flores-Sandoval, E., Fujiyama, A., Fukuzawa, H., Galik, B., Grimanelli, D., Grimwood, J., Grossniklaus, U., Hamada, T., Haseloff, J., Hetherington, A.J., Higo, A., Hirakawa, Y., Hundley, H.N., Ikeda, Y., Inoue, K., Inoue, S.-I., Ishida, S., Jia, Q., Kakita, M.,

Kanazawa, T., Kawai, Y., Kawashima, T., Kennedy, M., Kinose, K., Kinoshita, T., Kohara, Y., Koide, E., Komatsu, K., Kopischke, S., Kubo, M., Kyozuka, J., Lagercrantz, U., Lin, S.-S., Lindquist, E., Lipzen, A.M., Lu, C.-W., De Luna, E., Martienssen, R.A., Minamino, N., Mizutani, M., Mizutani, M., Mochizuki, N., Monte, I., Mosher, R., Nagasaki, H., Nakagami, H., Naramoto, S., Nishitani, K., Ohtani, M., Okamoto, T., Okumura, M., Phillips, J., Pollak, B., Reinders, A., Rövekamp, M., Sano, R., Sawa, S., Schmid, M.W., Shirakawa, M., Solano, R., Spunde, A., Suetsugu, N., Sugano, S., Sugiyama, A., Sun, R., Suzuki, Y., Takenaka, M., Takezawa, D., Tomogane, H., Tsuzuki, M., Ueda, T., Umeda, M., Ward, J.M., Watanabe, Y., Yazaki, K., Yokoyama, R., Yoshitake, Y., Yotsui, I., Zachgo, S. and Schmutz, J., 2017. Insights into land plant evolution garnered from the *Marchantia polymorpha* genome. *Cell*. 171, 287-304.

Bradford, M.M., 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* 72, 248-254.

Chaiprasongsuk, M., Zhang, C., Qian, P., Chen, X., Li, G., Trigiano, R.N., Guo, H., Chen, F., 2018. Biochemical characterization in Norway spruce (*Picea abies*) of SABATH methyltransferases that methylate phytohormones. *Phytochemistry*. 149, 146-154.

Chen, F., D'auria, J.C., Tholl, D., Ross, J.R., Gershenzon, J., Noel, J.P., Pichersky, E., 2003a. An *Arabidopsis thaliana* gene for methylsalicylate biosynthesis, identified by a biochemical genomics approach, has a role in defense. *Plant J.* 36, 577-588.

Chen, F., Tholl, D., D'auria, J.C., Faroog, A., Pichersky, E., Gershenzon, J., 2003b. Biosynthesis and emission of terpenoid volatiles from *Arabidopsis* flowers. *Plant Cell*. 15, 481-494.

Chen, F., Ludwiczuk, A., Wei, G., Chen, X., Crandall-Stotler B., Bowman, J.L., 2018. Terpenoid secondary metabolites in bryophytes: chemical diversity, biosynthesis and biological functions. *Crit. Rev. Plant Sci*. 37, 210-231.

Craft, J.D., Harrelson, D., Setzer, W.N., 2016. Chemotypic variation of *Conocephalum salebrosum* in the southeastern Appalachian range: a search for cryptic plant biodiversity around the Tennessee River Valley. *Nat. Prod. Commun*. 11, 1009-1014.

Curtis, A., Southwell, I.A., Stiff, I.A., 1990. *Eucalyptus*, a new source of E-methyl cinnamate. *J. Essent. Oil Res*. 2, 105-110.

D'auria, J.C., Chen, F., Pichersky, E., 2003. The SABATH family of MTS in *Arabidopsis thaliana* and other plant species. In: Romeo, J.T. (Ed.), *Recent Advances in Phytochemistry*, vol. 37. Elsevier, Amsterdam, pp. 253-283.

Davis, D.R., Landry, J.-F., 2012. A review of the North American genus *Epimartyria* (Lepidoptera, Micropterigidae) with a discussion of the larval plastron. *ZooKeys*. 37-83.

Dodson, C.H., Dressler, R.L., Hills, H.G., Adams, R.M., Williams, N.H., 1969. Biologically active compounds in orchid fragrances. *Science*. 164, 1243-1249.

Eltz, T., Lunau, K., 2005. Antennal response to fragrance compounds in male orchid bees. *Chemoecology*. 15, 135-138.

Finn, R.D., Clements, J., Eddy, S.R., 2011. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res.* 39, W29-W37.

Finn, R.D., Coghill, P., Eberhardt, R.Y., Eddy, S.R., Mistry, J., Mitchell, A.L., Potter, S.C., Punta, M., Qureshi, M., Sangrador-Vegas, A., Salazar, G.A., Tate, J., Bateman A., 2015. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* 44, D279-D285.

Forrest, L.L., Davis, E.C., Long, D.G., Carandall-Stotler, B.J., Clark, A., Hollingsworth, M.L., 2006. Unraveling the evolutionary history of the liverworts (Marchantiophyta): multiple taxa, genomes and analyses. *Bryologist*. 109, 303-335.

Fujiwara, G.M., Annies, V., de Oliveira, C.F., Lara, R.A., Gabriel, M.M., Betim, F.C., Nadal, J.M., Farago, P.V., Dias, J.F., Miguel, O.G., Miguel, M.D., Marques, F.A., Zanin, S.M., 2017. Evaluation of larvicidal activity and ecotoxicity of linalool, methyl cinnamate and methyl cinnamate/linalool in combination against *Aedes aegypti*. *Ecotoxicol. Environ. Saf.* 139, 238-244.

Ghani, N. A., Ludwiczuk, A., Ismail, N.H., Asakawa, Y., 2016. Volatile components of the stressed liverwort *Conocephalum conicum*. *Nat. Prod. Commun.* 11, 103-104.

Gilles, M., Zhao, J., An, M., Agboola, S., 2010. Chemical composition and antimicrobial properties of essential oils of three Australian *Eucalyptus* species. Food Chem. 119, 731-737.

Gomes da Silva, M.D., Chaves das Neves, H.J., 1999. Complementary use of hyphenated purge-and-trap gas chromatography techniques and sensory analysis in the aroma profiling of strawberries (*Fragaria ananassa*). J. Agr. Food Chem. 47, 4568-4573.

Harinantenaina, L., Kida, S., Asakawa, Y., 2007. Phytochemistry of three selected liverworts: *Conocephalum conicum*, *Plagiochila barteri* and *P. terebrans*. Arkivoc. 7, 22-29.

Hattori, M., Sakagami, Y., Marumo, S., 1992. Oviposition deterrents for the limabean pod borer, *Etiella zinckenella* (Treitschke)(Lepidoptera: Pyralidae) from *Populus nigra* L.c.v. Italica leaves. Appl. Entomol. Zool. 27, 195-204.

Hippauf, F., Michalsky, E., Huang, R., Preissner, R., Barkman, T.J., Piechulla, B., 2010. Enzymatic, expression and structural divergences among carboxyl O-methyltransferases after gene duplication and speciation in *Nicotiana*. Plant Mol. Biol. 72, 311-330.

Huang, R., Hippauf, F., Rohrbeck, D., Haustein, M., Wenke, K., Feike, J., Sorrelle, N., Piechulla, B., Barkman, T.J., 2012. Enzyme functional evolution through improved catalysis of ancestrally nonpreferred substrates. Proc. Natl. Acad. Sci. 109, 2966-2971.

Ismail, H.M., Williams, A.A., Tucknott, O.G., 1980. The flavour components of plum. Eur. Food Res. Technol. 171, 24-27.

Kaiser, R. 1993. The scent of orchids: olfactory and chemical investigations. Elsevier Science Publishers B.V..

Kapteyn, J., Qualley, A.V., Xie, Z., Fridman, E., Dudareva, N., Gang, D.R., 2007. Evolution of cinnamate/*p*-coumarate carboxyl methyltransferases and their role in the biosynthesis of methylcinnamate. Plant Cell. 19, 3212-3229.

Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30, 772-780.

Kato, M., Mizuno, K., Crozier, A., Fujimura, T., Ashihara, H., 2000. Caffeine synthase gene from tea leaves. Nature. 406, 956-957.

Knudsen, J.T., Tollsten, L., Bergström, L.G., 1993. Floral scents—a checklist of volatile compounds isolated by head-space techniques. Phytochemistry. 33, 253-280.

Koeduka, T., Kajiyama, M., Suzuki, H., Furuta, T., Tsuge, T., Matsui, K., 2016. Benzenoid biosynthesis in the flowers of *Eriobotrya japonica*: molecular cloning and functional characterization of *p*-methoxybenzoic acid carboxyl methyltransferase. Planta. 244, 725-736.

Köllner, T.G., Lenk, C., Zhao, N., Seidl-Adams, I., Gershenzon, J., Chen, F., Degenhardt, J., 2010. Herbivore-induced SABATH methyltransferases of maize that methylate anthranilic acid using S-adenosyl-L-methionine. *Plant Physiol.* 153. 1795-1807.

Kumar, S., Stecher, G., Tamura, K., 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Bio. Evol.* 33, 1870-1874.

Ludwiczuk, A., Odrzykoski, I.J., Asakawa, Y., 2013. Identification of cryptic species within liverwort *Conocephalum conicum* based on the volatile components. *Phytochemistry.* 95, 234-241.

Matasci, N., Hung, L., Yan, Z., Carpenter, E.J., Wickett, N. J., Mirarab, S., Nguyen, N., Warnow, T., Ayyampalayam, S., Barker, M., Burleigh, J. G., Gitzendanner, M. A., Wafula, E., Der, J. P., dePamphilis, C. W., Roure, B., Philippe, H., Ruhfel, B. R., Miles, N. W., Graham, S. W., Mathews, S., Surek, B., Melkonian, M., Soltis, D. E., Soltis, P. S., Rothfels, C., Pokorny, L., Shaw, J. A., DeGironimo, L., Stevenson, D. W., Villarreal, J. C., Chen, T., Kutchan, T. M., Rolf, M., Baucom, R. S., Deyholos, M. K., Samudrala, R., Tian, Z., Wu, X., Sun, X., Zhang, Y., Wang, J., Leebens-Mack, J., Wong, G. K., 2014. Data access for the 1,000 Plants (1KP) project. *GigaScience.* 3, 17.

McCarthy, A.A., McCarthy, J.G., 2007. The structure of two N-methyltransferases from the caffeine biosynthetic pathway. *Plant Physiol.* 144, 879-889.

Murfitt, L.M., Kolosova, N., Mann, C.J., 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.

Padalia, R.C., Verma, R.S., Chauhan, A., Goswami, P., Singh, V.R., Verma, S.K., Darokar, M.P., Singh, N., Saikia, D., Chanotiya, C.S., 2017. Essential oil composition and antimicrobial activity of methyl cinnamate-Linalool chemovariant of *Ocimum basilicum* L. from India. Rec. Nat. Prod. 11, 193.

Park, S.-W., Kaimoyo, E., Kumar, D., Mosher, S., Klessig, D.F., 2007. Methyl salicylate is a critical mobile signal for plant systemic acquired resistance. Science. 318, 113-116.

Price, M. N., Dehal, P. S., Arkin, A. P., 2010. FastTree 2—approximately maximum-likelihood trees for large alignments. PloS one. 5, e9490.

Qian, P., Zhao, N., Chen, F., Guo, H., 2015. Understanding Substrate Specificity of Related Plant Methyltransferases (MESs) from Computational Investigations. Chem. J. Chinese U. 36, 2283-2291.

Qin, G., Gu, H., Zhao, Y., Ma, Z., Shi, G., Yang, Y., Pichersky, E., Chen, H., Liu, M., Chen, Z., 2005. An indole-3-acetic acid carboxyl methyltransferase regulates *Arabidopsis* leaf development. Plant Cell. 17, 2693-2704.

Seo, H.S., Song, J.T., Cheong, J.-J., Lee, Y.-H., Lee, Y.-W., Hwang, I., Lee, J.S., Choi, Y.D., 2001. Jasmonic acid carboxyl methyltransferase: a key enzyme for jasmonate-regulated plant responses. *Proc. Natl. Acad. Sci. U.S.A.* 98, 4788-4793.

Shulaev, V., Silverman, P., Raskin, I., 1997. Airborne signalling by methyl salicylate in plant pathogen resistance. *Nature*. 385, 718.

Szweykowski, J., Buczkowska, K., Odrzykoski, I., 2005. *Conocephalum salebrosum* (Marchantiopsida, Conocephalaceae)—a new Holarctic liverwort species. *Plant Syst. Evol.* 253, 133-158.

Toyota, M., 2000. Phytochemical study of liverworts *Conocephalum conicum* and *Chiloscyphus polyanthos*. *Yakugaku Zasshi*. 120, 1359-1372.

Toyota, M., Saito, T., Matsunami, J., Asakawa, Y., 1997. A comparative study on three chemo-types of the liverwort *Conocephalum conicum* using volatile constituents. *Phytochemistry*. 44, 1265-1270.

Varbanova, M., Yamaguchi, S., Yang, Y., McKelvey, K., Hanada, A., Borochoy, R., Yu, F., Jikumaru, Y., Ross, J., Cortes, D., 2007. Methylation of gibberellins by *Arabidopsis* GAMT1 and GAMT2. *Plant Cell*. 19, 32-45.

Viña, A., Murillo, E., 2003. Essential oil composition from twelve varieties of basil (*Ocimum spp*) grown in Colombia. *J. Brazil. Chem. Soc.* 14, 744-749.

Wood, W.F., Lancaster, W.C., Fisher, C.O., Stotler, R.E., 1996. (*E*)-methyl cinnamate: The major volatile from some populations of the liverwort, *Conocephalum conicum*. *Phytochemistry*. 42, 241-242.

Yang, Y., Yuan, J.S., Ross, J., Noel, J.P., Pichersky, E., Chen, F., 2006. An *Arabidopsis thaliana* methyltransferase capable of methylating farnesoic acid. *Arch. Biochem. Biophys.* 448, 123-132.

Yao, J., Chu, Y., An, R., Guo, H., 2012. Understanding product specificity of protein lysine methyltransferases from QM/MM molecular dynamics and free energy simulations: the effects of mutation on SET7/9 beyond the Tyr/Phe switch. *J. Chem. Inf. Model.* 52, 449-456.

Yao, J., Xu, Q., Chen, F., Guo, H., 2010. QM/MM free energy simulations of salicylic acid methyltransferase: Effects of stabilization of ts-like structures on substrate specificity. *J. Phys. Chem. B.* 115, 389-396.

Zhao, N., Yao, J., Chaiprasongsuk, Li, G., Guan, J., Tschaplinski, T.J., Chen, F., 2013. Molecular and biochemical characterization of the jasmonic acid methyltransferase gene from black cottonwood (*Populus trichocarpa*). *Phytochemistry*. 94, 74-81.

Zhao, N., Ferrer, J.-L., Moon, H.S., Kapteyn, J., Zhuang, X., Hasebe, M., Stewart Jr, C.N., Gang, D.R., Chen, F., 2012. A SABATH Methyltransferase from the moss *Physcomitrella patens* catalyzes S-methylation of thiols and has a role in detoxification. *Phytochemistry*. 81, 31-41.

Zhao, N., Ferrer, J.-L., Ross, J., Guan, J., Yang, Y., Pichersky, E., Noel, J.P., 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.

Zhao, N., Guan, J., Lin, H., Chen, F., 2007. Molecular cloning and biochemical characterization of indole-3-acetic acid methyltransferase from poplar. *Phytochemistry*. 68, 1537-1544.

Zhu, J., Park, K.-C., 2005. Methyl salicylate, a soybean aphid-induced plant volatile attractive to the predator *Coccinella septempunctata*. *J. Chem. Ecol.* 31, 1733-1746.

Zubieta, C., Ross, J. R., Koscheski, P., Yang, Y., Pichersky, E., Noel, J.P., 2003. Structural basis for substrate recognition in the salicylic acid carboxyl methyltransferase family. *Plant Cell*. 15, 1704-1716.

Appendix B: Figures and Tables

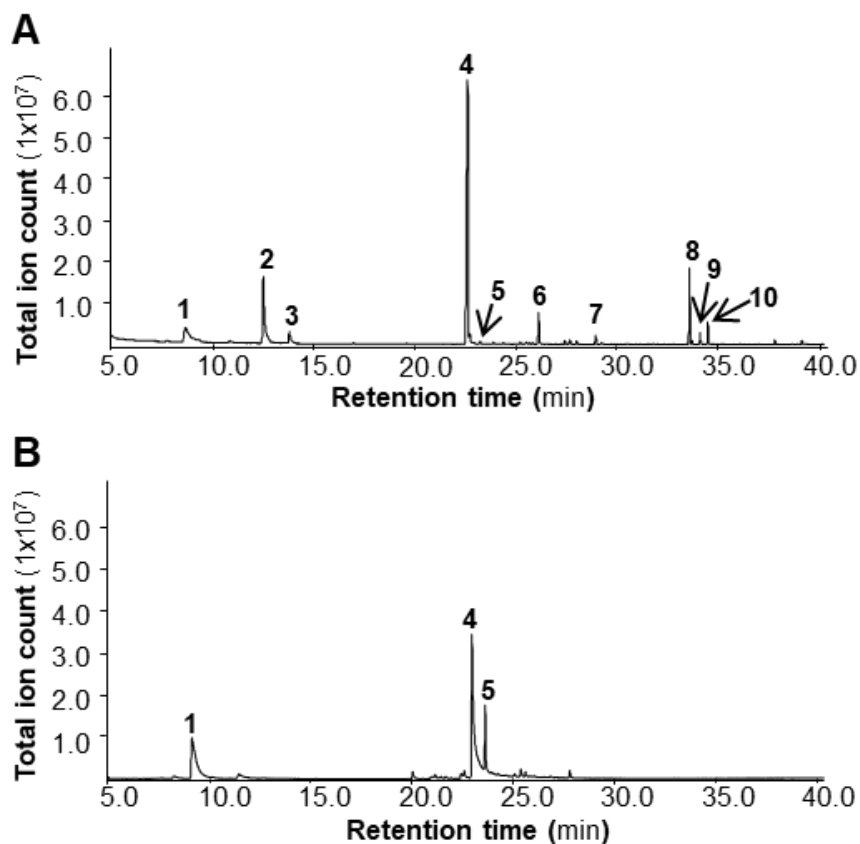


Figure B-1. Chemical analysis of *C. salebrosum*. (A) GC chromatogram of the organic extraction of *C. salebrosum*. 1*, sabinene; 2, 1-octanol; 3, 1-octenyl acetate; 4*, methyl (*E*)-cinnamate; 5, unidentified sesquiterpene hydrocarbons; 6*, 10-epi-cubebol; 7*, 1-epi-cubenol; 8, 1-octadecyne; 9, E-2-tetradecen-1-ol; 10, phytol. Methyl (*E*)-cinnamate is the dominant component (50.52%) followed by sabinene (9.85%), 1-octadecyne (8.33%) and δ -cadinol (3.96%). (B) GC chromatogram of headspace collection of *C. salebrosum*. 1*, sabinene; 4*, methyl (*E*)-cinnamate; 5, unidentified sesquiterpene hydrocarbons. * indicates substrates that were verified by standard substances.

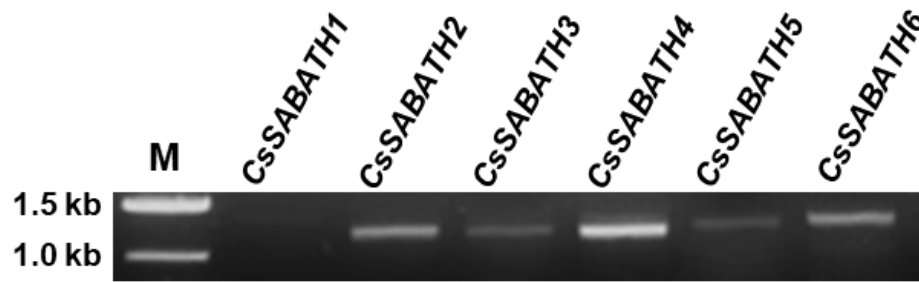


Figure B-2. Expression of six CsSABATHs genes in sterile thalli. Semi-quantitative RT-PCR of CsSABATH genes. Same amount of cDNA and respective specific primers were used for PCR. The cycle number is 25.

Figure B-3. Sequence alignment of CsSABATHs with ObCCMT1 and CbSAMT;

CsSABATHs, *C. salebrosum* SABATHs; ObCCMT1, *Ocimum basilicum* cinnamate/*p*-coumarate carboxyl methyltransferase 1; CbSAMT, *Clarkia breweri* salicylic acid

methyltransferase. Conserved residues are in shade with the more conserved the darker.

Residues indicated with “&” are SAM/SAH-binding residues. Residues indicated with “*” are

residues that interact with the carboxyl moiety of substrate. Residues indicated with “#”

interact with the aromatic moiety of substrate and are important for substrate selectivity.

CsSABATH1 : MTAGVDRAGMGFGALDVPSSNAKLRYSSNAAEPNQQLSPGDMVANN-----ALSIAEVSELNALAMKGG : 66
CsSABATH2 : MAIG-----EPGLSS-----GAAVRRRSATHDHF-MGKG : 28
CsSABATH3 : -----MTID-----NPKTARHTAMEKVFAMNAG : 23
CsSABATH4 : MGP-----DGTSSNSSPAFCGQGTQG : 22
CsSABATH5 : MKT-----DMSTNYKREVLASQVETLIVGVCDLTMITTT-----MESTVKSPMPKLQQVRVMTG : 54
CsSABATH6 : M-----DMTGSSPCKFLNEHTQSPMPNVFKDLSSSITTCMVPKQVQDVEHEEPGSANSPLPT-----MGKG : 60
ObCCMT1 : -----MARK-----ENYVVSNNMVESVLCMKGG : 23
CbSANT : -----MDVROVLHMKGG : 12

CsSABATH1 : EGDASYAKNS-HTCAQGF-RFVQFVLQAATVRMDIPT--QGVVRIADLGCSSGPNTHNMNHIVHTLRDRMAAAS--QQL : 140
CsSABATH2 : DRKENHVRNSYNSQMLNSNLFPAFLEAIERIRPNPQLGPTVAEFRSBSGTSVDNVGTVLQRFKERRLNEI--GQ- : 105
CsSABATH3 : SCENSYAKNS-NPQQFLVKNLHABELLEALDQALASN-DGQPVVVDLGCSSGPNTHNMNLILNHLKQRAAAA--DS- : 98
CsSABATH4 : ECENSYAKNS-SVQATLM-RKILKFFDVMKSMILPDP-DCAFRIADLGCSSGPNTHANVEAITEQVFAKKEEGLGDS- : 98
CsSABATH5 : DGVDSYARNS-HPGDFM-SRLPFTVFDVINRSSLPQ--RGVTAVDLGCSSGPNTHANVDATINRMFAKGTGEG-- : 125
CsSABATH6 : LGKNSYVRNS-NGQADFV-ERVFETVLKPVDEMKILEKDTDVVRVVEFGCCSSGPNTHIRNVETILRRMKRNLS--GE- : 133
ObCCMT1 : KCEDSYDNNS-KMCEQHA-RSLHLLMEALDGVGSSVAAAGAFVVDLGCSSGPNTHANVTMEFMNHLTEHTVTA--EE- : 98
CbSANT : AGENSYAMNS-FICRQV-SITKITEAATITALYSGDVTVTRLAADLGCSSGPNTHAVTELTKTEELRKKMG--RE- : 87

CsSABATH1 : QLSSSSRKAPGGAPFEFFVYFADLPSNDFNSLRRHSRDHSRSQRDLRGASSGYSSCVAGSFYCRLFPRASVNVVVSFS : 220
CsSABATH2 : -----QPEYQVYFCDLPTTDFNMILKLRNAA--FLSTENSAEADYHAAAVPGTIYGRLFPHYSLHFAFTFA : 171
CsSABATH3 : ---NSSIK----IPDEHVFFNDLPSNDFNYLQQLSEQ-----RDSLDYHAAAGVPGSFYGRLFPRASVHIFSSLC : 162
CsSABATH4 : -----VPEIQVYFCDLPTTDFNLTFFKHLFAQ-PNSE--VETVRNYSAAVPGSFYCDLRFPKSINIAMSSFA : 162
CsSABATH5 : -----PEYQAYFCDLPTTDFNLTFRLLNLK-QKSEPGGKVVPMPYFAAGVPGSFYDRLFPHVASLHFMSSFA : 191
CsSABATH6 : -----CPEFQAFCDLHETDFNLTFRFFKTV-SANSNGSDDDLKYHAGVAGSGCYGRIFPKSVHFAAMSTFA : 200
ObCCMT1 : -----PPEFSAFCDLPSNDFNLTQLPPS-----DGSSGSYTAGVAGSFYRRLFPAKSVDEFFYSAS : 158
CbSANT : -----NSPEYQIFLNDLPSNDFNLTFRSLPIE-----NDVDGVCHINGVPGSFYGRLFPRNLTLEHHSYS : 148

CsSABATH1 : LHWLSKIPDVWQKDSPAYNNGFVWHGGKPAVAKAYAEQSRSDVSLKRADEMVPCCGLMFLILKGRKDS--ETIQY : 298
CsSABATH2 : LNCLSRIPESVDRSSPAYNGGQTGLHQSSVATPAYSAQAREDCFKLGRADEMADDGVLCNFRORDSDNSTPYST- : 250
CsSABATH3 : LHMMSKVEDAVLDKNSPAFNKCDMMVHGRSEVGAARKEQADLDURFLSARABELAPGGLFVMLFTGRGNSD--PARQW : 240
CsSABATH4 : LHWLAQVFDVARDRESPAYNGGHTDIYRSSLATQAYAEQADKDLDNFLAARAEVAPCGGLIFTFAIROSEF--PYIY- : 239
CsSABATH5 : LHWLSQIPEAVLDKTSPAYNGGHTELHLSSISTVDAYAEQAKRDISSFLAARAEVMPKCGSMMLVFGIRENYF--PYRI- : 268
CsSABATH6 : LHWLSRIPFVYDQNDYFAGAHNELVFASKATHAAYAEQANIDERNFLDARATEMVPCCGMVTLVFGVRKITY--PYSI- : 277
ObCCMT1 : LHWLSQIPEKEMEGSAAYNBERVTINGAKESTVNAVKKQFQSDLGVLFRSRKELKPGGSMFLMLLGRTPSPD--EADQ- : 235
CbSANT : LHWLSQVFIGTESNKGNIYMA----NTCPQSVLNAYYKQFQEHALFLRCRAQEVVPCGRMVLITILGRSED--RAST- : 220

CsSABATH1 : DPNGST--FGTHMEAVENELISEGLDAEVRDTFNIPMYVPTVNEVREAVDESGA-FLVDRLEFLNDVP-ANPSTKQEL : 374
CsSABATH2 : -P-----HEVIGDSINDDVEKGLISASMRDSNPSPHYWRSVQEVKEVSEFSSEFQVHKQSMHQVHMLAGAHGATTG : 322
CsSABATH3 : -PDNSYAYEVWKCMDESNDSMVTQGHITEDORDANVPLYNRSLEEVREATESCGSDRVERKLVTRDTGG--TKVSALFP : 317
CsSABATH4 : ---EPTI----HLQEEVWNAVLVDVSAEVRDSYNIPIYFRTIEDVKNQVEKYGSVEEVQREVKIKL---NDDLSLMG : 308
CsSABATH5 : ---GSTI----ELQEKINDLVVEGLVEAKLRDSHNNYVYFRTISEVDGVSSFSSTLTVKREVCPSPL--SSFSMNP : 339
CsSABATH6 : ---GPTD----MTGEKVNDMISEGIVDAKLKESPKSYMYFRYKDEEVDLSFSSMAKTEQRHV-----SPFFTYPP : 343
ObCCMT1 : ---GAWILTFSTRYQDANDLVQEGLSSEKRDFTFNIPYTPSEEFKEVVERDGA-HINLKLQFHG---GSALIIDD : 307
CbSANT : ---ECCL---IWQLLAMALNOMVSEGLIEEEKMKENIROYTPSPTEEAELKESG-FLDHIHASEIYW-SSCTKDGDG : 293

CsSABATH1 : EASPTDCGRKLAKTGRSLGLVLDSEHMCVAVADQFFLLEERATARS-----QVEALRHRAAYSHVNLAV-LIRK- : 443
CsSABATH2 : PSSEAVVPKGVPTLCRGPSSGMLRAHEGREVTNMLERYAQVIREGI-----GNQTSYYQPTTWVVLV--LERKP : 391
CsSABATH3 : NTPTSFPRKRITSFMSMHSPVYEAHICKELGRVYWNSEVRMLENKIEDLIL---TLERDPAHKEHFDLRMV-LIRS- : 392
CsSABATH4 : TENLREAAARVNNRGSTSTGYFEAFFCKTATDLYFERWEDAIYEQL----LLWQSGQEQVGIKPKSVDMIANGLRRK- : 383
CsSABATH5 : ASNAEIAERITSITGAGGNLEEHECKLSTRLEMERYRKALVEGF-----NNKTVEDLANARKHLVIVSRK- : 409
CsSABATH6 : GLTAREQGARFVSLQGINSTVFEAHECKEATNIEFWERYEELVRGF-----SDGTHDDEENYVILVVVATRR- : 413
ObCCMT1 : PNDAVEISRAYVSLGRSLTGGLVDAAHICDQLGHELFSLLSQAQDQA-----KELMDQFQL--VHIVASLTLA- : 373
CbSANT : GGSVEEGYNVARCMAVAEPLLDLHFCGATIEDVFEHYKLLITERM-----SKEKTK-----FNNIVSLTIKKS : 358

CsSABATH1 : ---- : -
CsSABATH2 : RVFD : 395
CsSABATH3 : ---- : -
CsSABATH4 : ---- : -
CsSABATH5 : ---- : -
CsSABATH6 : ---- : -
ObCCMT1 : ---- : -
CbSANT : ---D : 359

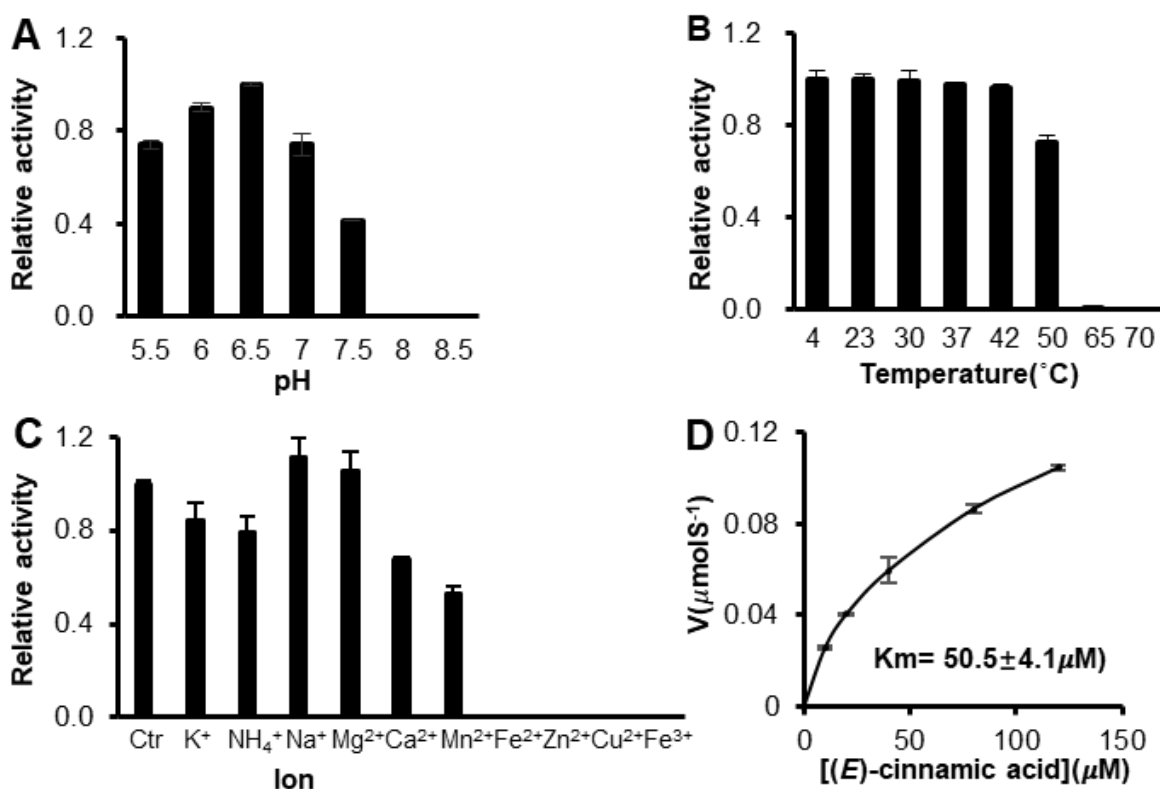


Figure B-4. Biochemical properties of CsCAMT. (A) pH optimum. Level of CsCAMT activity in the buffer of pH 6.5 was arbitrarily set at 1.0. (B) Thermostability. The activity of CsCAMT incubated at 4 oC for 30 min was arbitrarily set at 1.0. (C) Effect of ions. Individual ions were added to reactions in the form of chloride salts at final concentration of 5 mM. Level of activity without any ion added served as a control and was arbitrarily set at 1.0. (D) Steady-state kinetic measurements of CsCAMT. (E)-Cinnamic acid was used as substrate for enzyme assays in (A) to (D). Each value was the mean of three independent measurements.

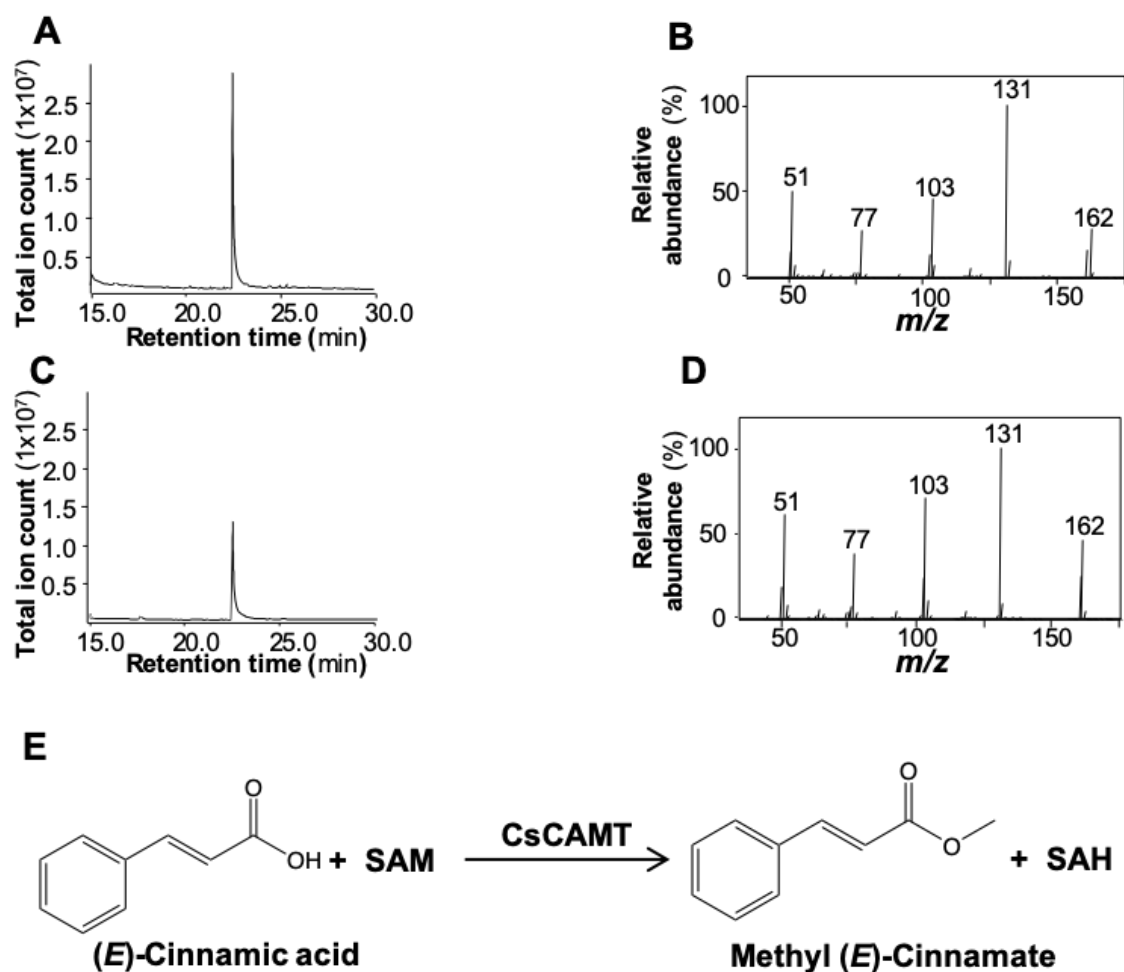
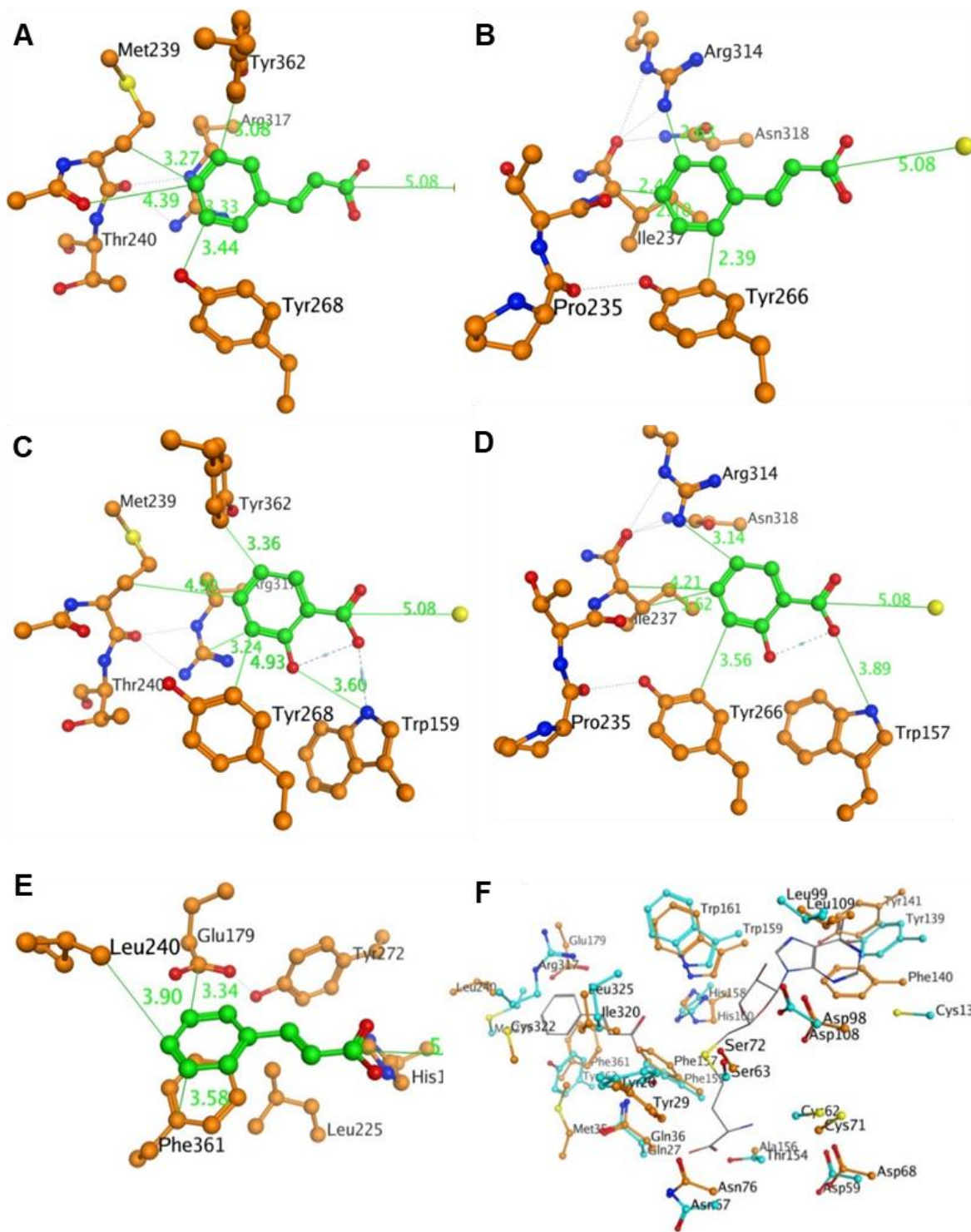


Figure B-5. Authentic methyl (*E*)-cinnamate and the product of CsCAMT methylation. (A) GC chromatogram of authentic methyl (*E*)-cinnamate; (B) Mass spectrum of authentic methyl (*E*)-cinnamate; (C) GC chromatogram of the product of CsCAMT; (D) Mass spectrum of the product of CsCAMT.

Figure B-6. Active sites of the structure models of CsCAMT, CsSAMT and ObCCMT1 with (*E*)-cinnamic acid or salicylic acid docked into the active sites. (A) CsCAMT (orange) complexed with (*E*)-cinnamic acid (green); the sulfur atom from SAH is shown as a yellow ball. Distances between the substrate and enzyme/SAH atoms are given in Å. (B) CsCAMT complexed with (*E*)-cinnamic acid. (C) CsCAMT complexed with salicylic acid. (D) CsSAMT complexed with salicylic acid. (E) ObCCMT1 complexed with (*E*)-cinnamic acid. (F) Superposition of ObCCMT1 (orange) and CsCAMT (cyan) with some active-site residues shown. (*E*)-Cinnamic acid and SAH are shown in line except for the sulfur atom.



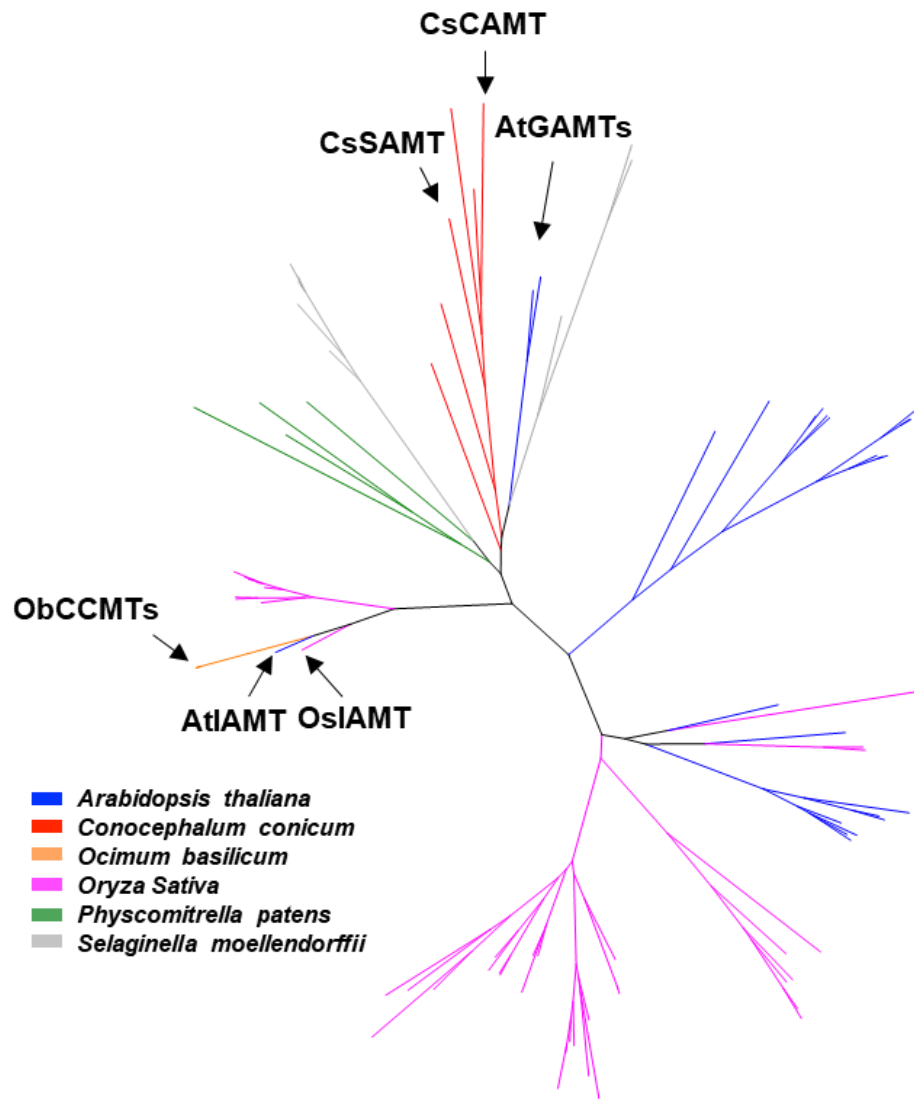


Figure B-7. Phylogenetic tree of the SABATH family of methyltransferases in *C. salebrosum*, *Arabidopsis thaliana*, *Oryza sativa*, *Physcomitrella patens*, *Selaginella moellendorffii* and ObCCMTs. The SABATH proteins from each species were color-coded. CCMTs: Cinnamate/*p*-coumarate methyltransferases; CAMT: cinnamic acid methyltransferase; IAMT: indole-3-acetic acid methyltransferase; SAMT: salicylic acid methyltransferase. AtIAMT, AtGAMTs, CsCAMT, CsSAMT, OsiIAMT and ObCCMTs are indicated with arrows.

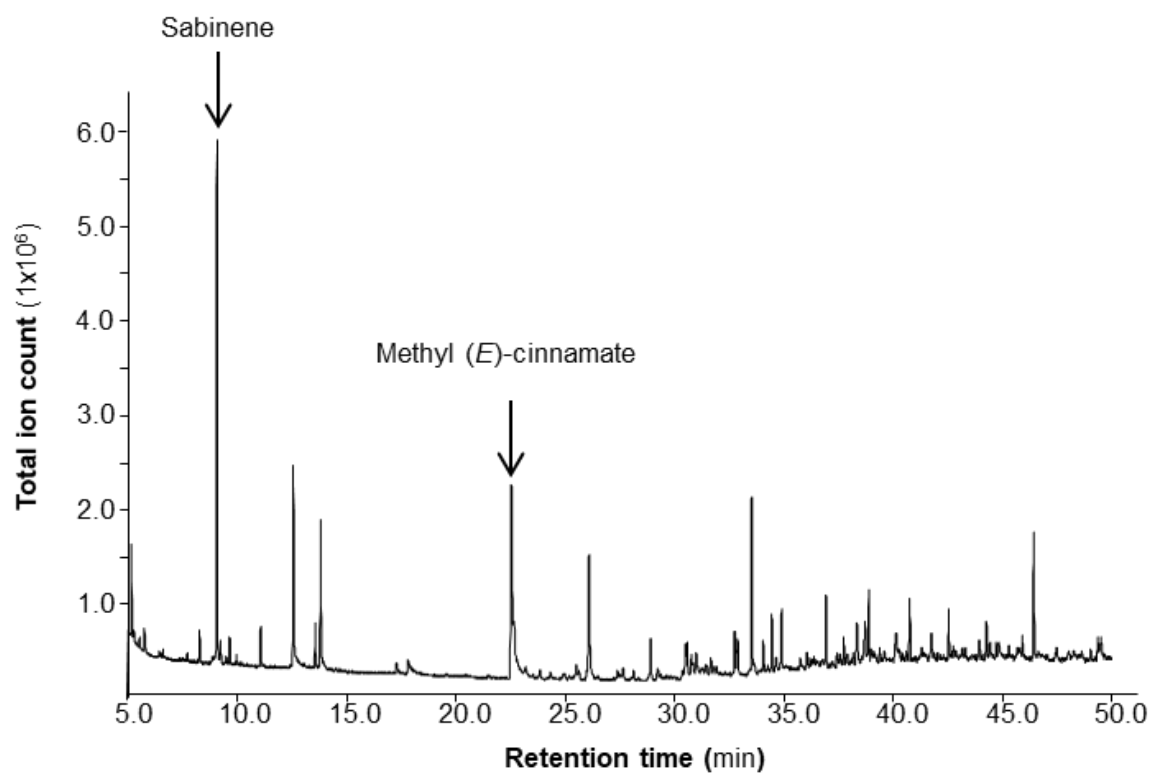


Figure B-8. Chemical analysis of Tennessee (TN) population of *Conocephalum salebrosum*. TN population exhibited sabinene-dominant chemotype in organic extraction.

Table B-1. The relative activities of CsSABATH4 and CsSABATH6 from *C. salebrosum* with ten carboxyl acids.

	CsSABATH4 (CsSAMT)	CsSABATH6 (CsCAMT)
(<i>E</i>)-Cinnamic acid	6%	100%*
<i>p</i> -Coumaric acid	4%	18%
Nicotinic acid	0%	3%
Anthranilic acid	12%	2%
Benzoic acid	51%	32%
Salicylic acid	100%*	0%
Absciscic acid	0%	0%
Jasmonic acid	1%	0%
Gibberellic acid	0	0%
Indole-3-acetic acid	5%	3%

*The highest level of activity was arbitrarily set at 100%.

Chapter III. Farnesoic acid methyltransferase involved in the biosynthesis of juvenile hormone III in rice flatsedge (*Cyperus iria*) is a member of the SABATH family

Adapted from:

Zhang Chi, Chaiprasongsuk Minta, Li Guanglin, Chen Xinlu, Jiang Yifan, Chen Feng.

Farnesoic acid methyltransferase involved in the biosynthesis of juvenile hormone III in rice flatsedge (*Cyperus iria*) is a member of the SABATH family. Drafted

Abstract

Juvenile hormones (JHs) are a group of sesquiterpenoids that regulate diverse aspects of insect physiology. Interestingly, some plants are known to synthesize JHs or their analogs. While the biosynthetic pathway to JHs in insects have been largely solved, our understanding of the pathway in plants is very limited. Here, we investigated the biosynthesis of JH III in *Cyperus iria*, commonly known as rice flatsedge, with a focus on farnesoic acid methyltransferase (FAMT), which is a key enzyme of the pathway. We first produced a transcriptome from the roots of *C. iria*, which accumulate higher concentrations of JH III. Comparing to the biosynthesis of JH III in insects, phosphatase, farnesol dehydrogenase and farnesal dehydrogenase homologs were identified but no homologs of insect FAMT were identified from the transcriptome. As such, we hypothesized that the enzyme is encoded by the member of the SABATH family. By comparative transcriptome analysis in rice flatsedge, *Cyperus iria*, we have isolated three full-length SABATH methyltransferase genes, which were designated CiSABATH1-3. Via in vitro biochemical assays against FA and a few related compounds, CiSABATH3 exhibited the highest level of activity with FA. It was renamed CiFAMT. The CiFAMT cDNA encodes an ORF of 372 amino acids with a calculated molecular mass of 41.57 kDa and the apparent K_m value of CiFAMT using FA as substrate

was determined to be 60.42 μ M. In contrast, CiSABATH1 and CiSABATH2 were demonstrated to function as benzoic acid/salicylic acid methyltransferase and jasmonic acid methyltransferase respectively. They were renamed CiBSMT and CiJAMT accordingly. Structure models for CiFAMT, CiBSMT and insect FAMT in complex with farnesoic acid and salicylic acid were built, which provided structural explanation to substrate specificity of these three enzymes. In summary, we have identified in this study FAMT involved in the biosynthesis of JH III in *C. iria*. Unlike FAMT from insects, the CiFAMT is a member of the SABATH family, indicating independent evolution of JH III biosynthetic pathway in plants and insects.

1. Introduction

Juvenile hormones (JHs) are a group of sesquiterpenoids that regulate metamorphosis and reproduction of insects (Gilbert et al., 2000; Nijhout, 1998; Raikhel et al., 2005; Riddiford, 1994). JHs prevents mainly metamorphosis during larval molting and regulates female reproduction. Since the discovery of JHs and recognition of their importance to insects, there has been enormous interest in the synthesis of these compounds and their analogs for use as insect control agents. At the same time, many natural products of plant origin have been screened for JH activities on miscellaneous insects. The ongoing search for plant natural products with JH activities has surprisingly resulted in the identification of JH III from a few related sedge species that belong to the Cyperaceae family.

The plant-derived JH III was originally identified from *Cyperus iria*, which is one of the most serious weeds causing economic losses in rice fields (Catling 1993; Holm et al. 1977). The selection of *C. iria* for extraction was predicated on its native common name, “grasshopper’s Cyperus” (Toong et al., 1988). Analyses by NMR and mass spectrometry (MS) indicated a compound consistent with the existence of JH III and the chirality of the compound is 10R (Toong et al., 1988), which is the same as that possessed by the natural insect JH III. *Cyperus aromaticus*, a sedge species closely related to *C. iria*, also contains high levels of JH III (Toong et al., 1988). In addition to JH III, other structurally related sesquiterpenoids have been isolated from *Cyperus microiria* (Iwamura, 1979) and *Cyperus polystachyos* (Iwamura, 1979), suggesting that JH III and other similar compounds may exist in diverse plant species.

The developmental distribution of JH III in *C. iria* had been determined from the seedling to the senescent plant (Bede et al., 1999a). JH III levels increased in immature plants and were maintained at a high level until about four months when flowering occurred. At this time, a transient decrease in JH III content was observed. In mature plants of about six months of age, JH levels in all plant tissues increased again. About 85% of the total JH III in *C. iria* was present in the roots, suggesting that the roots are a site of synthesis and/or storage of JH III.

Because of the central role of JH III to insect development and reproduction, its biosynthesis in insects has been relatively well studied. The insect biosynthetic pathway to JH III is conventionally divided into the early step and the late step. In the early step, farnesyl diphosphate is generated via the mevalonate pathway that is conserved in plants and animals (Goldstein and Brown, 1990). Farnesyl diphosphate is a common precursor for a number of branch pathways. For example, in plants, all sesquiterpenoids and phytol are derived from farnesyl diphosphate. In the late step, JH III is synthesized from farnesyl diphosphate, which involves five enzymatic reactions (Shinoda and Itoyama, 2003). The first reaction, from farnesyl diphosphate to farnesol, is likely catalyzed by a phosphatase or a terpene synthase. The second reaction, from farnesol to farnesal, is catalyzed by an alcohol dehydrogenase. From farnesal to farnesoic acid, the third step of the pathway, is catalyzed by an aldehyde dehydrogenase. Using cell free preparations of the insect corpora allata, in which JH III is synthesized, Baker and his coworkers (1983) identified NADP-dependent farnesol and/or farnesal dehydrogenase activity in the tobacco hornworm. From farnesoic acid to JH III, two enzymes are involved: one carboxyl methyl transferase for methylation and one cytochrome

P450-dependent monooxygenase for epoxidation (Feyereisen et al., 1981). Depending on the particular system, either methylation or epoxidation occurs first (Bhaskaran et al., 1986). The outline of the proposed JH III-specific pathway in insects is shown in Figure. C-1. In recent years, significant progress has been made in the identification of genes encoding for the enzymes involved in pathway. One characterized gene encodes a cytochrome P450 that catalyzes the epoxidation of methyl farnesoate to JH III in cockroach (Bhaskaran et al., 1986). Another gene identified is the juvenile hormone acid methyl transferase gene from silkworm (Shinoda and Itoyama, 2003). Interestingly, the protein encoded by BmJHAMT does not share significant sequence similarities to other known methyl transferases (Shinoda and Itoyama, 2003).

The biochemical pathway leading to JH III in plants appears similar to the pathway occurring in insects. Using *C. iria* cell suspension cultures, which accumulate JH III, as a model system, Bede and her coworkers (1999b) identified methyl farnesoate, the immediate precursor of JH III. Moreover, enzyme inhibition, labeling studies and precursor feeding assays provided more convincing evidence that the biosynthesis of JH III in plants proceeds via a similar pathway to the one in insects (Bede et al., 2000). HMG-CoA reductase is a key regulatory enzyme in the mevalonate pathway for synthesizing farnesyl diphosphate (Chappell et al., 1995). Treating *C. iria* cell suspension cultures with mevinolin, a potent competitive inhibitor of HMG-CoA which had no effect on cell growth, caused >85% inhibition of JH III production (Chappell et al., 1995). When mevalonate, the product of the enzyme, was added to mevinolin-treated cell suspension cultures, the production of JH III was restored (Bede et al., 2001). Taken together, these data suggest that the backbone of JH III is

coming from the mevalonate pathway. To understand the pathway from farnesyl diphosphate to JH III, Bede et al. (2001) performed precursor feeding assays. By supplying the mevinolin-inhibited cell suspension cultures of *C. iria* with farnesol, farnesal, farnesoic acid, or methyl farnesoate, the production of JH III in the cell suspension cultures increased, suggesting that all these compounds are involved in the pathway for JH III biosynthesis. Finally, Bede et al. (2001) tested the involvement of cytochrome P450 in the biosynthesis of JH III in the *C. iria* cell suspension cultures. From the insect model, an NADP-dependent cytochrome P450 monooxygenase is involved in the late stage of JH III biosynthesis (Feyereisen et al., 1981). To test whether the same type of enzyme catalyzes the late step of JH III biosynthesis in plants, enzyme inhibitor assays were performed. When the inhibitors of P450 enzyme were added into the *C. iria* cell suspension cultures, JH III production was completely inhibited. Adding methyl farnesoate into the cell suspension cultures did not restore the production of JH III. These results suggest that the last step of JH III biosynthesis in *C. iria* is catalyzed by a P450 enzyme (Bede et al., 2001). In plants, the molecular aspects of the mevalonate pathway leading to the formation of FPP has been very well characterized (Lichtenthaler, 1999). In a sharp contrast, nothing is known at the molecular level about the specific pathway leading to the formation of JH III from farnesyl diphosphate in plants.

Although these advances gave us some clue about JH III biosynthetic pathway in *C. iria*, what are the key enzymes and whether they evolved independently from those in insects are still unknown. For example, a plant specific methyltransferase family is responsible methylating various small carboxyl acid (D'Auria et al., 2003) and it might also play similar function in the last but one step of JH III biosynthesis in *C. iria*. Since all the previous study

of JH III biosynthesis in plant are based on *C. iria* cell suspension culture and enzymatic inhibition, none of enzyme-coding genes in the pathway have been isolated which deter further research and application. The study of biosynthetic pathway was merely in vitro because the lack of sequence information from the model plant *C. iria*. Here we proposed to take advantage of transcriptome that acquired via RNA-seq to decode the key enzyme-coding genes in JH III biosynthetic pathway. We first quantified the content of JH III and its precursors in plant and send the tissue for RNA-seq. Then we analyzed the transcriptome and found homologous genes or analogous genes corresponding to those of insects. Finally, we selected one key enzyme for biochemical characterization.

2. Material and methods

2.1. Plant culture

Cyperus iria seeds were originally from Dr. Jacquie Bede. 50 *Cyperus iria* seeds were germinated and grown in pots with soil from local grassland under greenhouse condition (North Greenhouse, UTIA, 18 hours of light and 6 hours of darkness, irrigated every two days). Leaves and roots of the plants were collected separately on the 30th, 60th day after for germination.

2.2. Metabolic profiling

Tissues were used for organic extraction as described using ethyl acetate (Sasidharan et al., 2011). Extracts were used for GC-MS analysis or stored in -20°C.

2.3. GC-MS analysis

Samples from organic extraction were eluted into methylchloride and analyzed with Shimadzu QP5050A GC-MS instrument. The experiments were carried out at the following conditions, flow rate of 5mL/min of carrier gas Helium, a splitless injection of 3 μ L, and a temperature gradient of 5°C/min from 40°C (3-min hold) to 240°C.

2.4. RNA preparation

Both leaves and roots tissues were harvested for the total RNA extraction. Total RNA was extracted from the tissue using RNeasy Plant Mini Kit (Qiagen, Valencia, CA) and DNA contamination was removed with an on-column DNase (Qiagen, Valencia, CA) treatment. 1.5 μ g of total RNA was reverse transcribed into first strand cDNA in a 15 μ L reaction volume using the First-strand cDNA Synthesis Kit (Amersham Biosciences, Piscataway, NJ) as previously described (Chen et al., 2003b).

2.5. Real-time qPCR

Real-time quantitative PCR (RT-PCR) was performed on an ABI7000 Sequence Detection System (Applied Biosystems, Foster City, CA) using SYBR green fluorescence dye (Bio-Rad Laboratories) as previously described (Yang et al., 2006). The gene-specific primers were designed and two primers used for the PCR amplification of ubiquitin was designed as the internal control.

2.6. Cloning full length cDNA of CiSABATHs

Total RNA was extracted using a RNeasy Plant Mini Kit (Qiagen, Valencia, CA) with DNA contamination removed using an off-column DNase treatment (Qiagen, Valencia, CA). After purification, total RNA (1.5 µg) was reverse-transcribed into first strand cDNA in a 15 µL reaction volume using the First-strand cDNA Synthesis Kit (Amersham Biosciences, Piscataway, NJ) as previously described (Chen et al., 2003b). CiSABATHs full-length cDNA was amplified using forward primer and reverse primer corresponding with the beginning and end of the CiSABATH coding region (Table 1), respectively. The PCR was conducted using the following program: 94 °C for 2 min followed by 30 cycles at 94 °C for 30 s, 57 °C for 30 s, 72 °C for 1 min 30 s, and a final extension at 72 °C for 10 min. Products of PCR were separated on 1.0% agarose gel. The target band was sliced from the gel, purified using QIAquick Gel Extraction kit (Qiagen, Valencia, CA), and cloned into a pCRT7/CT-TOPO vector using the protocol recommended by the vendor (Invitrogen, Carlsbad, CA). Cloned cDNA in pCRT7/CT-TOPO vector was sequenced using T7 and V5 primers.

2.7. Purification of CiSABATHs expressed in *E. coli*

CiSABATHs were subcloned from cDNA from pCRT7/CT-TOPO into the vector of pET32a (Invitrogen, Carlsbad, CA). To express the CiSABATHs protein, the protein expression construct was transformed into the *E. coli* strain BL21 (DE3) CodonPlus (Stratagene, La Jolla, CA). Protein expression was induced by isopropyl β-D-1-thiogalactopyranoside (IPTG) at a concentration of 500 µM for 18 h at 25 °C, with cells lysed by sonication. His-tagged CiSABATH protein was purified from the *E. coli* cell lysate using

Ni-NTA agarose following manufacturer instructions (Invitrogen, Carlsbad, CA). Protein purity was verified by SDS-PAGE and protein concentrations was determined by the Bradford assay (Bradford, 1976).

2.8. Radiochemical CiSABATHs activity assay

Radiochemical CiSABATH assays was performed with a 50 μ L volume containing 50 mM Tris-HCl, pH 7.5, 1 mM substrate, and 3 μ M ^{14} C-SAM with a specific activity of 51.4 mCi/mmol (Perkin Elmer, Boston, MA). The assay was initiated by addition of SAM and kept at room temperature for 30 min. The reaction was stopped by the addition of 150 μ L ethyl acetate, vortexed, and phase-separated by a 1 min centrifugation at 14,000g. The upper organic phase was counted using a liquid scintillation counter (Beckman Coulter, Fullerton, CA). The amount of radioactivity in the organic phase indicated the amount of synthesized of the methylated substrates. Three independent assays were performed for each compound.

2.9. Determination of kinetic parameters of CiSABATH3

In all kinetic analyses, the appropriate enzyme concentrations and incubation time was chosen so that the reaction velocity was linear during the reaction time period. For the CiSABATH3, to determine a K_m value for SAM, concentrations of SAM were independently varied from 3 to 120 μ M, while substrates were held constant at 1 mM. To determine the K_m for substrates, concentrations substrates was independently varied from 10 to 200 μ M, while SAM was held constant at 200 μ M. Assays was conducted at 25 $^{\circ}$ C for 30 min. The kinetic parameters K_m and V_{max} was calculated with GraphPad Prism 5 software for Windows

(GraphPad Software Inc.), using standard settings for non-linear regression curve fitting in Michaelis–Menten equation.

2.10. Production of transcriptomes and analysis.

RNA-Seq and Expression abundance. RNA was isolated from the leaf and root of rice flatsedge samples (three biological replicates for each tissue). Samples were sequenced on Illumina Hiseq 2000 following the manufacturer's instructions. Raw reads were filtered, from which adapter, contamination and low-quality reads, clean reads were removed. Each replicate has 47~49 million reads (90bp pair-end reads). All reads of leaf and root transcriptomes were de novo assembled for one transcript set using Trinity (v2.0.6). To calculate transcript abundance, reads were mapped to transcript set by bowtie and numbers of supporting read were used to estimate transcript expression abundance by RPKM (Reads Per Kilobase per Million mapped reads).

Functional annotation. The peptides of transcripts were translated by EMBOSS transeq and filtered with length not less than 50 bp. ORF (open reading frame) in transcripts were identified based on peptides location. Interproscan5 was used to annotate the protein sequence translated from our transcript set, and the NCBI nr database was also used for the transcript annotation. Pfam Domain information can be accessed in introproscan annotation results. We used Pfam domain or PIRSF to identify multiple gene families, like p450 (PF00067) and TPS genes (PF01397, PF03936).

Gene Cluster. As rice flatsedge genome has not been sequenced, the standard gene set is unknown. To analyze the detailed information of the gene expressed in different tissue,

multiple transcripts needed to be clustered into the same gene or gene family with artificial arguments. As genes in different tissues may express with different expression levels or transcript variants, root and leaf transcriptomes were also assembled, respectively. So, three transcript sets (root, leaf, and integral transcript set) were clustered by the next steps. We first blasted all sequence files to themselves and filter the results, which satisfy 95% identity and 60% coverage for either sequence of matches. By further eliminating self matches and integrating multiple matches that share the same sequences, transcripts in different transcript set were clustered into the same gene or gene families.

2.11. Phylogenic Analysis

To identify putative SABATH proteins, an HMM search as previously described (Chaiprasongsuk et al., 2018) was performed against proteins that were either transcribed from *C. iria* unigenes or retrieved from Phytozome (Goodstein et al., 2011) and data source (for other species). All identified SABATH protein sequence with a minimum length of 200 amino acids were used to render a multiple sequence alignment via MAFFT (version 7.369b, under L-INS-I strategy) (Katoh and Standley, 2013). RAxML (version 8.2.10) was used to construct maximum likelihood analyses with 1000 bootstrap replicates through CIPRES Science Gateway (Miller et al., 2010; Stamatakis, 2014).

3. Results

3.1. Chemical characterization of JH III and related compounds in *C. iria*.

To determine JH III and its precursors in *C. iria*, we first divided plant into two parts, aboveground tissue (leaf and stem) and underground tissue (root). Then two time points were selected, 30 day after germination (30DAG) and 60 day after germination (60DAG). GC-MS based measurement was applied to each sample. Three compounds in JH III biosynthetic pathway were detected, farnesol, methyl farnesoate and JH III (Figure C-2). When comparing spatial distribution of these three compounds, they were found exclusively in underground tissue. When comparing underground tissues of 30DAG and 60DAG, farnesol concentration increased from 0.008 mg/g to 0.033 mg/g, MF increased from 0.010 mg/g to 0.0154 mg/g and JH III increased from 0.145 mg/g to 0.343 mg/g (Figure C-2E).

3.2. Transcriptomes of *Cyperus iria* root tissue

Since JHIII and its precursors were only found exclusively in root tissue, we speculated that JHIII biosynthetic pathway was primarily working in root tissue. For subsequent pathway assembling and molecular cloning, a transcriptomic study of *C. iria* was performed. The raw reads were qualified, and adapters were removed, yielding 29.58 Gb of sequence data. All of the reads obtained from transcriptomes were assembled, resulting in 97564 transcripts (N50:1562) with a mean length of 973 bp.

3.3. Identification of homologous genes for JH III biosynthesis in *C. iria*

In this study, we first assumed that *C. iria* possessed homologous genes for JH III biosynthesis similar to that in insects. Thus, we performed multiple Blast search using insect genes as queries against *C. iria* transcriptome to identify candidate genes. 5 phosphatase homologs, 10 farnesol dehydrogenase homologs and 10 farnesal dehydrogenase homologs were identified through the Blast search (Figure C-3). These homologs were candidate genes that were responsible for farnesyl diphosphate to farnesol, farnesol to farnesal and farnesal to farnesoic acid respectively. However, we cannot identify homologs to farnesoic acid O-methyl transferase and juvenile hormone acid methyltransferase which are the enzymes convert farnesoic acid into methyl farnesoate in insects. At the same time, cytochrome P450-dependent monooxygenases of the same subfamily that convert methyl farnesoate into JH III, homologs to epoxidase, also cannot be found.

3.4. Identification of SABATH genes among Unigenes

Since the homologs to the methyltransferases that convert farnesoic acid into methyl farnesoate in insects cannot be identified in *C. iria*, SABATH, a plant specific methyltransferase family, quickly drawn our attention. To identify SABATH genes, a search against *C. iria* transcriptome based on HMMR with Pfam file (PF03492) was conducted. Seven unigene coding proteins were identified as SABATH methyltransferases with three of them appeared to be in full length and they were renamed CiSABATH1-3 (Table S1). The sizes of these three proteins range from 372 to 393 amino acids in length. Their similarity to AtFAMT, one known FAMT from *Arabidopsis*, are between 39% and 43%. The amino acid

residues that bind the methyl donor SAM among CiSABATHs and AtFAMT are identical. The residues that interact with the carboxyl moiety of the substrate are conserved. However, the residues that interact with the aromatic moiety are different (Figure C-4).

3.5. Biochemical characterization of CiSABATHs expressed in *E. coli*

To obtain catalytic function of each full-length SABATH protein, they were first cloned by RT-PCR into a protein expression vector without any tag and expressed in *E. coli* to produce recombinant proteins. Individual recombinant CiSABATH enzymes in their native forms were assayed with following substrates: IAA, GA₃, SA, BA, JA, LA, GeA and FA (Figure C-5A). Among three CiSABATH enzymes, CiSABATH3 showed highest activity towards FA. It also showed high activity with geranic acid, which was about 95% of the activity with FA and some activity with BA, IAA and LA, which was about 21%, 18% and 24% of the activity with FA (Figure C-5A). CiSABATH1 showed highest activity against both BA and SA but no activity against FA, while CiSABATH2 showed highest activity against JA but no activity against FA. Since CiSABATH3 shows highest activity using farnesoic acid as substrate, CiFAMT was assigned to this protein. Similarly, CiBSMT and CiJAMT were assigned to CiSABATH1 and CiSABATH2 respectively.

3.6. Detailed biochemical properties of CiFAMT

With the identification of CiFAMT in *C. iria*, then, we performed biochemical assays to determine its detailed biochemical properties using a purified CiFAMT containing a N-terminal His-tag. The optimum pH for CiFAMT was determined to be 6.5 (Figure C-5B). It loses activity when pH is above 6.5 and when pH is above 8 it loses all activity. For

temperature stability, its activity is relatively stable when the temperature is below 42 °C. When the temperature is above 42 °C, it loses activity dramatically. When the temperature is above 70 °C, it loses all activity (Figure C-5C). Na⁺, Mg²⁺ can slightly increase the activity of CiFAMT. Potassium ion, ammonium ion, calcium ion and manganese ion slightly inhibit the activity of CiFAMT. CiFAMT loses all activity when ferrous ion, zinc ion, copper ion and ferric ion exist in the reaction system (Figure C-5D). The apparent Km value of CiFAMT is 60.42 μM (Figure C-5E). The product of CiFAMT was verified to be MF.

3.7. Comparison of FAMT gene expression in different tissues at different ages of *C. iria*

To study the expression pattern of CiFAMT in *C. iria*. We collected samples following above-mentioned in chemical characterization part. cDNA was acquired through reverse-transcription PCR and qPCR was performed to measure relative expression level of CiFAMT between samples. qPCR showed that CiFAMT expressed more in roots than in leaves and its expression increased dramatically at 60DAG when comparing to the level at 30DAG (data not shown).

3.8. Phylogenetic Analysis of CiSABATHs with other SABATH members in Monocots

To understand how CiFAMT emerged through evolution, a maximum-likelihood phylogenetic tree was constructed using three CiSABATHs and SABATH proteins from *Arabidopsis thaliana* and five species of monocots including *Oryza sativa*, *Elaeis guineensis*, *Musa acuminata*, *Ananas comosus* and *Zea mays*. (Figure C-6). CiBSMT was closely related to OsBSMT; CiJAMT was closely related to OsJAMT1. CiFAMT was clustered with two

uncharacterized SABATHs from *A. comosus* and together they resolved in clade with AtBSMT, OsBSMT, CiBSMT and many other uncharacterized SABATH. On the other hand, AtFAMT clustered with other uncharacterized SABATH from *Arabidopsis* and together resolved in a clade with AtGAMTs, IAMTs and other SABATH from monocots.

4. Discussion

With the identification of homologs to insect enzyme-coding genes for JH III biosynthesis in *C. iria*, this research showed that the biosynthetic pathway until farnesoic acid might be conserved in insect and plant but not the following last two steps. We then moved on to test whether plant specific enzymes might step into this pathway and verify this assumption through the isolation and characterization of a FAMT in *C. iria*. Based previous knowledge, SABATH methyltransferase genes became our major candidates. Three full-length SABATH genes, CiSABATH1-3, have been isolated and characterized. One of them, CiSABATH3 was determined to encode FAMT and renamed as CiFAMT. The K_m value (60.42 μ M) of CiFAMT is about one and half time of that for AtFAMT from *Arabidopsis* (Yang et al., 2006). CiFAMT and AtFAMT also showed different selectivity towards geranic acid and lauric acid. CiFAMT had about 94% and 24% of activities of farnesoic acid with geranic acid and lauric acid, respectively. While AtFAMT showed about 28% and 18% of activities of farnesoic acid with geranic acid and lauric acid, respectively (Yang et al., 2006). Gene expression data is in accord with biochemical profile, in which high expression of CiFAMT in root with high content of methyl farnesoate and JH III in root. Beside CiFAMT, CiSABATH1

has highest activity against benzoic acid and salicylic acid and CiSABATH2 has highest activity against jasmonic acid and they were renamed CiBSMT and CiJAMT, respectively.

As indicated by our phylogenetic study, CiFAMT shared a common ancestor with two SABATH proteins from *A. comosus* with unknown function which together far from AtFAMT of *Arabidopsis*. CiBSMT and OsBSMT from rice together with other monocot SABATH proteins shared a common ancestor. CiJAMT, OsJAMT from rice and a SABATH protein from *Zea mays* shared a common ancestor. The phylogeny in combination with biochemical and sequence evidences implies that the evolution of FAMT in *C. iria* and *Arabidopsis* is the result of convergent evolution.

Our previous knowledge on plant methyltransferases prompted the idea that this FAMT might belong to SABATH family of methyltransferase (D'Auria et al., 2003). Majority known substrate of SABATH methyltransferases are vital phytohormones including indole-3-acetic acid, gibberellins, salicylic acid and jasmonic acid (Chen et al., 2003; Qin et al., 2005; Seo et al., 2001; Varbanova et al., 2007), or plant volatile compounds including benzoic acid, *p*-methoxybenzoic acid and anthranilic acid (Koeduka et al., 2016; Köllner et al., 2010; Murfitt et al., 2000). Other substrates of SABATH methyltransferases include nitrogen-containing compounds, such as nicotinic acid and theobromine (Hippauf et al., 2010; KatoM and Crozier, 2000; McCarthy and McCarthy, 2007). Through a large-scale in vitro substrate screening, Yang et al. isolated a SABATH protein, AtFAMT, that uses FA as substrate, but methyl farnesoate and final product JH III can only be detected in *C. iria*. This difference implies that these two FAMTs might have different roles in their plant. The high level of JH III in *C. iria* makes it tempting to speculate that the compound is involved in the protection of

the plant against insect herbivores. JH titer is precisely regulated during insect development. A topical application of JH or synthetic analogues onto the final larval stadium resulted in inappropriate retention of juvenile characteristics at the next molt (Sehnal, 1983). The amounts of JH III in *C. iria* are significantly higher than the levels of the JH III found in insects, implying that JH III can exert a developmental impact on the insects that feed on the plant. The effect of *C. iria* on the development of grasshoppers has been mentioned earlier. In a field study, it was observed that the eggs of the dipteran leafminer, *Hydrellia* sp., did not hatch if laid on the leaves of *C. iria* (Meneses and Garcia de la Osa, 1988). In another study, the leaf extract of *C. iria* showed a larvicidal effect on the larvae of mosquito, *Aedes aegypti* (Schwartz et al., 1998). The effects of JH III on other organisms should also be considered. These include defense against nematodes, fungi, and bacteria. JH III in *C. iria* roots may be involved in allelopathy. *C. iria* is an invasive species, and it is intriguing to ask whether JH III in roots contribute to the invasiveness of the plant. In an allelopathic study, the extracts of *C. iria* inhibited the germination of lettuce seeds and shoot growth of rice seedlings (Bede et al., 2000). However, it should be pointed out that most of the experiments were performed in in vitro bioassays. The in vivo biological function of JH III in plants still needs to be established.

Plants produce inexhaustible and diverse chemicals which play important roles in various physiological processes including the interaction between plant and other organisms including insects and microbial pathogens. Many evidences have supported that plant synthesizes a variety of compounds as chemical weapon against herbivores and pathogens. While insects and pathogenic microbes always find their countermeasures over a period of time. To cope with this situation, plant produces other chemicals to deter enemies. The levels

of defense and counter-defense continue to escalate without either side “winning”. This arm-race forces plant produce various chemicals. In this study, a SABATH methyltransferase, CiFAMT, helps plant upgrade its arsenal to a new level by directly targeting insect molting.

Reference

- Bede, J., Teal, P. and Tobe, S., 1999. Production of insect juvenile hormone III and its precursors in cell suspension cultures of the sedge, *Cyperus iria* L. Plant cell reports. 19, 20-25.
- Bowers, W., Fales, H., Thompson, M. and Uebel, E., 1966. Juvenile hormone: identification of an active compound from balsam fir. Science. 154, 1020-1021.
- Chaiprasongsuk, M., Zhang, C., Qian, P., Chen, X., Li, G., Trigiano, R. N., Guo, H. and Chen, F., 2018. Biochemical characterization in Norway spruce (*Picea abies*) of SABATH methyltransferases that methylate phytohormones. Phytochemistry. 149, 146-154.
- Chen, F., D'auria, J. C., Tholl, D., Ross, J. R., Gershenzon, J., Noel, J. P. and Pichersky, E., 2003. An *Arabidopsis thaliana* gene for methylsalicylate biosynthesis, identified by a biochemical genomics approach, has a role in defense. The Plant Journal. 36, 577-588.
- Duffey, S. S., 1980. Sequestration of plant natural products by insects. Annual review of entomology. 25, 447-477.
- Goodstein, D. M., Shu, S., Howson, R., Neupane, R., Hayes, R. D., Fazo, J., Mitros, T., Dirks, W., Hellsten, U. and Putnam, N., 2011. Phytozome: a comparative platform for green plant genomics. Nucleic acids research. 40, D1178-D1186.
- Hippauf, F., Michalsky, E., Huang, R., Preissner, R., Barkman, T. J. and Piechulla, B., 2010. Enzymatic, expression and structural divergences among carboxyl O-methyltransferases after gene duplication and speciation in Nicotiana. Plant molecular biology. 72, 311-330.

- Katoh, K. and Standley, D. M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution*. 30, 772-780.
- KatoM, M. and Crozier, A., 2000. caffeine synthase gene from tea leaves [J]. *Nature*. 406, 956-957.
- Koeduka, T., Kajiyama, M., Suzuki, H., Furuta, T., Tsuge, T. and Matsui, K., 2016. Benzenoid biosynthesis in the flowers of *Eriobotrya japonica*: molecular cloning and functional characterization of *p*-methoxybenzoic acid carboxyl methyltransferase. *Planta*. 244, 725-736.
- Köllner, T. G., Lenk, C., Zhao, N., Seidl-Adams, I., Gershenzon, J., Chen, F. and Degenhardt, J., 2010. Herbivore-induced SABATH methyltransferases of maize that methylate anthranilic acid using *S*-adenosyl-L-methionine. *Plant physiology*. pp. 110.158360.
- McCarthy, A. A. and McCarthy, J. G., 2007. The structure of two N-methyltransferases from the caffeine biosynthetic pathway. *Plant physiology*. 144, 879-889.
- Murfitt, L. M., Kolosova, N., Mann, C. J. and 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*. *Archives of biochemistry and biophysics*. 382, 145-151.
- Nishida, R., 1994. Sequestration of plant secondary compounds by butterflies and moths. *Chemoecology*. 5, 127-138.

- Qin, G., Gu, H., Zhao, Y., Ma, Z., Shi, G., Yang, Y., Pichersky, E., Chen, H., Liu, M. and Chen, Z., 2005. An indole-3-acetic acid carboxyl methyltransferase regulates *Arabidopsis* leaf development. *The Plant Cell*. 17, 2693-2704.
- Sasidharan, S., Chen, Y., Saravanan, D., Sundram, K. and Latha, L. Y., 2011. Extraction, isolation and characterization of bioactive compounds from plants' extracts. *African Journal of Traditional, Complementary and Alternative Medicines*. 8.
- Seo, H. S., Song, J. T., Cheong, J.-J., Lee, Y.-H., Lee, Y.-W., Hwang, I., Lee, J. S. and Do Choi, Y., 2001. Jasmonic acid carboxyl methyltransferase: a key enzyme for jasmonate-regulated plant responses. *Proceedings of the National Academy of Sciences*. 98, 4788-4793.
- Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. 30, 1312-1313.
- Toong, Y. C., Schooley, D. A. and Baker, F. C., 1988. Isolation of insect juvenile hormone III from a plant. *Nature*. 333, 170.
- Varbanova, M., Yamaguchi, S., Yang, Y., McKelvey, K., Hanada, A., Borochoy, R., Yu, F., Jikumaru, Y., Ross, J. and Cortes, D., 2007. Methylation of gibberellins by *Arabidopsis* GAMT1 and GAMT2. *The Plant Cell*. 19, 32-45.
- Yang, Y., Yuan, J. S., Ross, J., Noel, J. P., Pichersky, E. and Chen, F., 2006. An *Arabidopsis thaliana* methyltransferase capable of methylating farnesoic acid. *Archives of biochemistry and biophysics*. 448, 123-132.

- Bede, J., Goodman, W., Tobe, S., 1999a. Developmental distribution of insect juvenile hormone III in the sedge, *Cyperus iria* L. *Phytochemistry* 52, 1269-1274.
- Bede, J., Goodman, W., Tobe, S., 2000. Quantification of juvenile hormone III in the sedge *Cyperus iria* L.: comparison of HPLC and radioimmunoassay. *Phytochemical Analysis: An International Journal of Plant Chemical and Biochemical Techniques* 11, 21-28.
- Bede, J. C., Teal, P. E., Goodman, W. G., Tobe, S. S., 2001. Biosynthetic pathway of insect juvenile hormone III in cell suspension cultures of the sedge *Cyperus iria*. *Plant physiology* 127, 584-593.
- Bhaskaran, G., Sparagana, S. P., Barrera, P., Dahm, K. H., 1986. Change in corpus allatum function during metamorphosis of the tobacco hornworm *Manduca sexta*: regulation at the terminal step in juvenile hormone biosynthesis. *Archives of insect biochemistry and physiology* 3, 321-338.
- Catling, David, 1993. *Rice in deep water*. Springer.
- Chaiprasongsuk, M., Zhang, C., Qian, P., Chen, X., Li, G., Trigiano, R. N., Guo, H., Chen, F., 2018. Biochemical characterization in Norway spruce (*Picea abies*) of SABATH methyltransferases that methylate phytohormones. *Phytochemistry* 149, 146-154.
- Chappell, J., Wolf, F., Proulx, J., Cuellar, R., Saunders, C., 1995. Is the reaction catalyzed by 3-hydroxy-3-methylglutaryl coenzyme A reductase a rate-limiting step for isoprenoid biosynthesis in plants? *Plant Physiology* 109, 1337-1343.

- Chen, F., D'auria, J. C., Tholl, D., Ross, J. R., Gershenzon, J., Noel, J. P., Pichersky, E., 2003. An *Arabidopsis thaliana* gene for methylsalicylate biosynthesis, identified by a biochemical genomics approach, has a role in defense. *The Plant Journal* 36, 577-588.
- D'Auria, J. C., Chen, F., Pichersky, E., 2003. Chapter eleven the SABATH family of MTS in *Arabidopsis thaliana* and other plant species. *Recent advances in phytochemistry*, vol. 37. Elsevier, pp. 253-283.
- Feyereisen, R., Pratt, G. E., Hamnett, A. F., 1981. Enzymic synthesis of juvenile hormone in locust corpora allata: Evidence for a microsomal cytochrome P-450 linked methyl farnesoate epoxidase. *European Journal of Biochemistry* 118, 231-238.
- Gilbert, L. I., Granger, N. A., Roe, R. M., 2000. The juvenile hormones: historical facts and speculations on future research directions. *Insect biochemistry and molecular biology* 30, 617-644.
- Goldstein, J. L., Brown, M. S., 1990. Regulation of the mevalonate pathway. *Nature* 343, 425.
- Goodstein, D. M., Shu, S., Howson, R., Neupane, R., Hayes, R. D., Fazo, J., Mitros, T., Dirks, W., Hellsten, U., Putnam, N., 2011. Phytozome: a comparative platform for green plant genomics. *Nucleic acids research* 40, D1178-D1186.
- Hippauf, F., Michalsky, E., Huang, R., Preissner, R., Barkman, T. J., Piechulla, B., 2010. Enzymatic, expression and structural divergences among carboxyl O-methyltransferases after gene duplication and speciation in *Nicotiana*. *Plant molecular biology* 72, 311-330.
- Holm, L.G., Plucknett, D.L., Pancho, J.V., Herberger, J.P., 1977. The world's worst weeds. The world's worst weeds.

- Iwamura, J., 1979. The constituents of essential oils from *Cyperus polystachyos* Rottb., *Cyperus globosus* Allioni and *Cyperus difformis* L. Journal of the Agricultural Chemical Society of Japan (Japan).
- Katoh, K., Standley, D. M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Molecular biology and evolution 30, 772-780.
- KatoM, M., Crozier, A., 2000. caffeinesynthasegene from tealeaves [J]. Nature 406, 956-957.
- Koeduka, T., Kajiyama, M., Suzuki, H., Furuta, T., Tsuge, T., Matsui, K., 2016. Benzenoid biosynthesis in the flowers of *Eriobotrya japonica*: molecular cloning and functional characterization of *p*-methoxybenzoic acid carboxyl methyltransferase. Planta 244, 725-736.
- Köllner, T. G., Lenk, C., Zhao, N., Seidl-Adams, I., Gershenzon, J., Chen, F., Degenhardt, J., 2010. Herbivore-induced SABATH methyltransferases of maize that methylate anthranilic acid using S-adenosyl-L-methionine. Plant physiology, pp. 110.158360.
- Lichtenthaler, H. K., 1999. The 1-deoxy-D-xylulose-5-phosphate pathway of isoprenoid biosynthesis in plants. Annual review of plant biology 50, 47-65.
- McCarthy, A. A., McCarthy, J. G., 2007. The structure of two N-methyltransferases from the caffeine biosynthetic pathway. Plant physiology 144, 879-889.
- Meneses, C., Garcia de la Osa, J., 1988. Principal weed hosts of *Hydrellia* sp. in the southern rice-growing zone of Sancti Spiritus. Centro Agrícola 15, 90-92.

- Miller, M. A., Pfeiffer, W., Schwartz, T., 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. 2010 gateway computing environments workshop (GCE). Ieee, pp. 1-8.
- Murfitt, L. M., Kolosova, N., Mann, C. J., 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*. Archives of biochemistry and biophysics 382, 145-151.
- Nijhout, H. F., 1998. Insect hormones. Princeton University Press.
- Qin, G., Gu, H., Zhao, Y., Ma, Z., Shi, G., Yang, Y., Pichersky, E., Chen, H., Liu, M., Chen, Z., 2005. An indole-3-acetic acid carboxyl methyltransferase regulates *Arabidopsis* leaf development. The Plant Cell 17, 2693-2704.
- Raikhel, A., Brown, M., Belles, X., 2005. 3.9 Hormonal Control of Reproductive Processes. Comprehensive molecular insect science, 433-491.
- Riddiford, L. M., 1994. Cellular and molecular actions of juvenile hormone I. General considerations and premetamorphic actions. Advances in insect physiology, vol. 24. Elsevier, pp. 213-274.
- Sasidharan, S., Chen, Y., Saravanan, D., Sundram, K., Latha, L. Y., 2011. Extraction, isolation and characterization of bioactive compounds from plants' extracts. African Journal of Traditional, Complementary and Alternative Medicines 8.
- Sccftwartz, A., Paskewitz, S. M., Orth, A. P., Tesch, M. J., Toong, I., Goodman, W. G., 1998. The lethal effects of *Cyperus iria* on *Aedes aegypti*. J Am Mosq Contr Assoc 14, 78-82.

- Seo, H. S., Song, J. T., Cheong, J.-J., Lee, Y.-H., Lee, Y.-W., Hwang, I., Lee, J. S., Do Choi, Y., 2001. Jasmonic acid carboxyl methyltransferase: a key enzyme for jasmonate-regulated plant responses. *Proceedings of the National Academy of Sciences* 98, 4788-4793.
- Shinoda, T., Itoyama, K., 2003. Juvenile hormone acid methyltransferase: a key regulatory enzyme for insect metamorphosis. *Proceedings of the National Academy of Sciences* 100, 11986-11991.
- Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312-1313.
- Toong, Y. C., Schooley, D. A., Baker, F. C., 1988. Isolation of insect juvenile hormone III from a plant. *Nature* 333, 170.
- Varbanova, M., Yamaguchi, S., Yang, Y., McKelvey, K., Hanada, A., Borochoy, R., Yu, F., Jikumaru, Y., Ross, J., Cortes, D., 2007. Methylation of gibberellins by *Arabidopsis* GAMT1 and GAMT2. *The Plant Cell* 19, 32-45.
- Yang, Y., Yuan, J. S., Ross, J., Noel, J. P., Pichersky, E., Chen, F., 2006. An *Arabidopsis thaliana* methyltransferase capable of methylating farnesoic acid. *Archives of biochemistry and biophysics* 448, 123-132.

Appendix C: Figures and Tables

Figure C-1. Proposed pathway of JH III biosynthesis. Intermediate and final products are in bold. Lepidopteran specific pathway is indicated by the dashed line and arrow. Farnesoic acid methyltransferase is labeled in red.

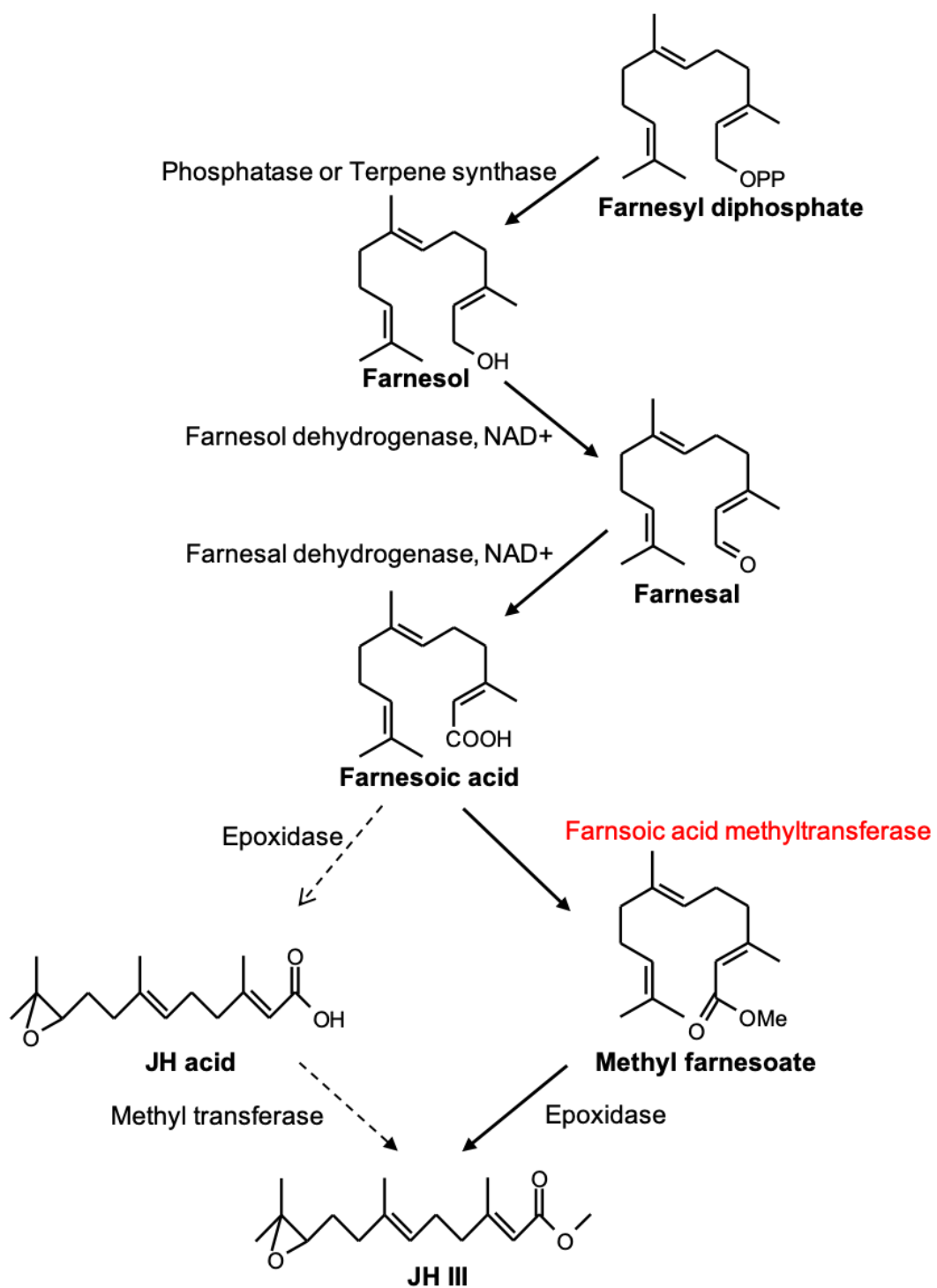
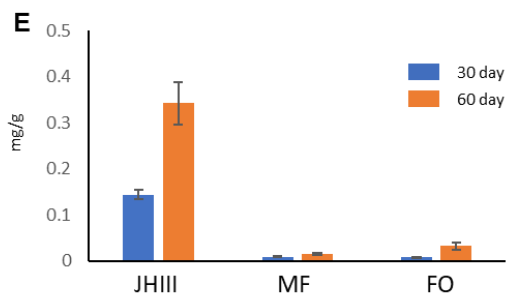
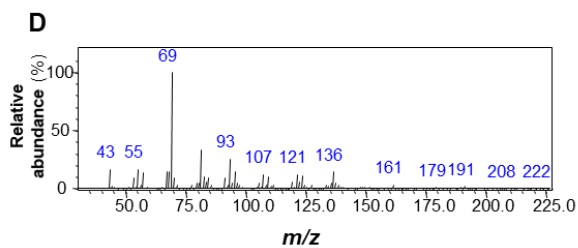
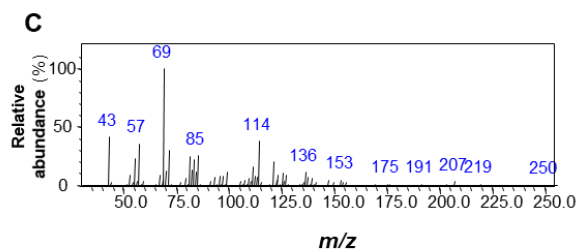
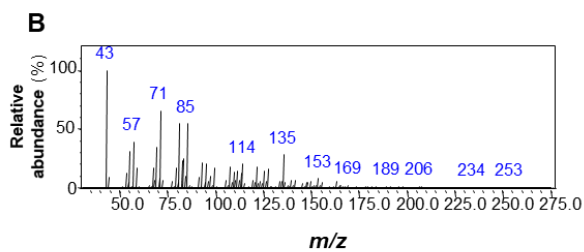
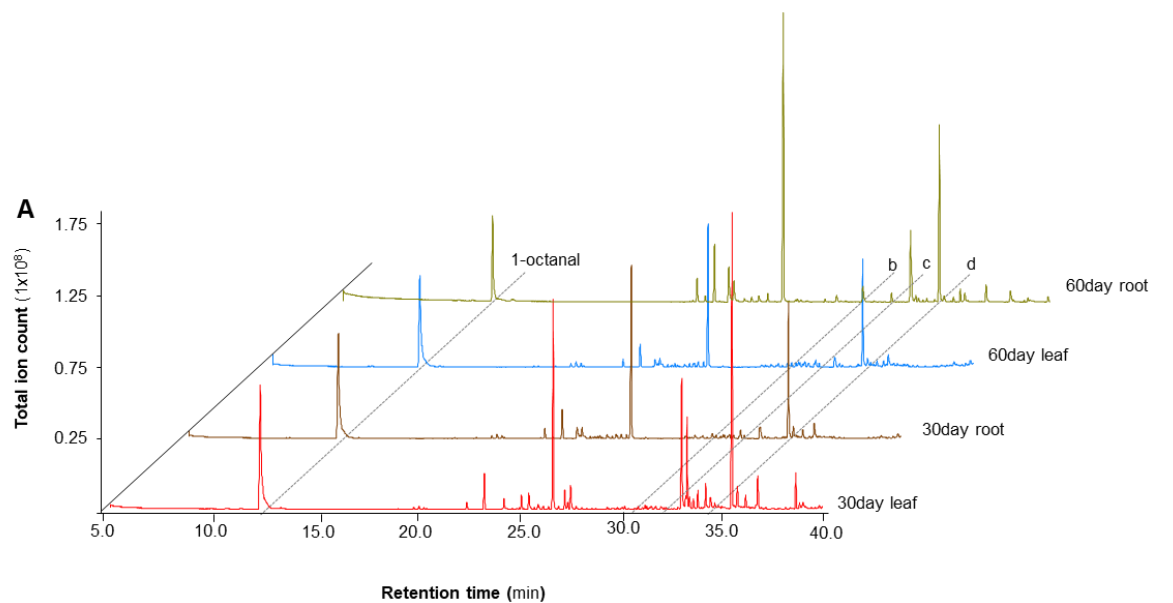


Figure C-2. Identification of JH3 and related chemicals in *C. iria*. (A) GC chromatogram of the organic extraction of *C. iria*. 30DAG leaf in red color; 30DAG root in brown color; 60DAG leaf in blue color and 60DAG root in green color; (B) Mass spectrum of compound b; (C) Mass spectrum of compound c; (D) Mass spectrum of compound d. (E) Quantification of Juvenile hormone (JH) III, methyl farnesoate (MF) and farnesol (FO). Each identity was verified by an authentic standard. DAG, day after germination. 1-octanol was added as internal control.



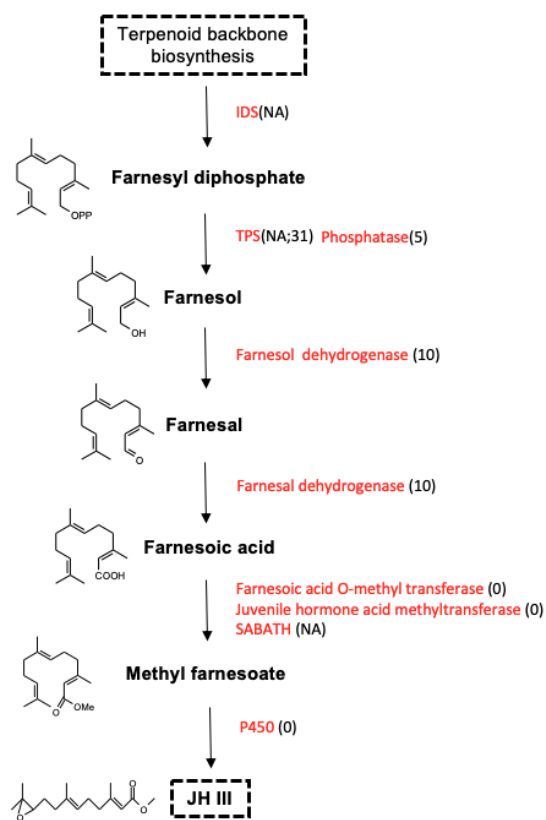


Figure C-3. The number of *C. iria* homologs of those enzyme-coding genes in insect and proposed JH III biosynthetic pathway in plants. Key enzymes are labeled in red; homolog numbers are in the parentheses.

```

CISABATH1 : -----YKVEKFLQITCGACHSSYAANSKLOEKAILKMPLEEAVGEVHKNLLIPADQKMVADLCSSGENTFLVL : 73
CISABATH2 : MAGNKEQQLQRLVRMNSVGETSYAKNSTMOKVIASLTSTRKDLARGYYQ-AHLS--KMVIADLGSSGTGNAILMAS : 77
CISABATH3 : -----YALERDFHVGCTGESSYAANSRLQKKIASGAMFMVKEAV-ESLVSTIPSSSEMLLVFADLCSTCANTFLMFP : 72
AtFAMT : -----MSTSFMTICGEGPNSYREHSKYOGALVIAAKEKINEAISTKLDIDFTSN--LVNIADLCSSGENTFTAVQ : 69
CbSAMT : -----YDROVLHMKGCAGENSYAMNSFLQROVISITKPITEAATALYSGDTVT--RLAADIACSSGENTALFAVT : 71
          &          &          &
          *
CISABATH1 : QIMKILTSKOCHAIR-LQPEQMOLFNDLPSNDFNNIPKS-LDSFEESVRADADAEHGQEQGSVSPHFAVAGLEGSFYGR : 151
CISABATH2 : DALQSVLAVCSELG-HLPPEQIIFNDLPWDFNSIFRSIGDYILSRGGDTQEQ-----LCFISAPGGSFYGR : 145
CISABATH3 : EVQRTVSAACSKLG-RPDEETQILFNDLPGNDFNALLGSLSSYKEKMNESMPDNK-----YTPFYAAAAGSFYDRL : 144
AtFAMT : TLLDAVENKYKESNIEGIEQVFPNDSSNDFNTLFTLPPAR-----LYEASGVGGSFFGRV : 128
CbSAMT : ELIKTVEELRKKVGRENSEFYQIFNDLPGNDFNATFRSL-----PIENDVDG-----VCFINSGGSFYGR : 134
          &&          &&
          &&
CISABATH1 : FPNQSVHFFHSSYSLMWISOVPEKGLDNKSQRDDAHLNKGNIYIGKTSPPLVVKLYQEQFORDFSLFLKLRFTTELVP : 231
CISABATH2 : FPSRSLHVVSESSLEWLSVPRVLOEGSIV-----NKGRIYISEASEPSVFDAYKHQFOAGFSLFLKRCRGVEVVK : 220
CISABATH3 : FPKRSVHVVYSFSSLEWLSOVPPCPNNETEG-----PINETSFNISKRSPPLVSKLYLEQFOKDFSSFLHRSEEL : 221
AtFAMT : LPKNSLHVGVSYSLEFVSKVPKEIKDRDSL----VWNKDLHCSCGSKE-VVKLYLGQYKIDVGSFLTARAQELVSG : 203
CbSAMT : FPRNLTLEHHSYSLMWISOVPEIGIES-----NKGNIYMANTCPOSVLNAMYKQFOEDHALFLRCRAQEVVPGGR : 205
          #          *
CISABATH1 : VLAFLGR-KNQIYSGELSTLYGLSDSLISMIAEGLVEEDKIDSFNLPPYSPSMEEAKAVINAQGLMDIIEIDMFESN : 310
CISABATH2 : VLTFMAR-RGLPSABEDSYWELTAVAMDLYSMGLIEEBVDSFNAPYAPSMDEVVEEQITNEGSFAIKKIELFEASY : 299
CISABATH3 : ALIFPAR-TSVIPVMEPCSVWMSWYEEALRELISCGVVKASVDSEFNIEFFPSLEEVSIIQNDASEEIIQHDVLEVS : 300
AtFAMT : ILLGSCRPTGVOMFETVEGMIDFISSSNEIANQGLDQOQIDTEKLPYAPNVDLKOILEDNKCFTEAFEFKISHAK : 283
CbSAMT : VLTILGR-RSEIRASTCCLLWQLTAMALNOMVSEGLIEEBMDKENIPQYTPSPTEVEAEILKEGSELIDHIEASEIY : 284
          #
CISABATH1 : DPYDDSDHSFVINNIQSGIN-----FAMCIKAVMEPTTIASHFGALIVDELFSRFAINVAKHLAKQ : 370
CISABATH2 : DALLDDQDFEEEHQEVTONNSNHASNKHSLLRKARRYAELTARTNRAVLESMIRNHFGGEIIEELFRRYRDLEEYYCN : 379
CISABATH3 : R--DGLGDISGQDSAEIMKK-----VVMAGRVVMEPLIVSHFGDSIIDPLFSVAPKVLAKYFWKE : 358
AtFAMT : GEYPLDPEYLTS-----AFKVTYGGSVASLFGQDGMKTYELVKEKTQEMLPQI : 332
CbSAMT : SSCTKDGDDGGGSVEEEGYN-----VARCMKAVAPEPLILDHFGCAIIEDEVHRYKLLTIERMSKE : 343
          # #
CISABATH1 : KTKLPVLTLSLKKDM : 386
CISABATH2 : KAELTNMLVALERK-- : 393
CISABATH3 : EYKLACTLSLKK-- : 372
AtFAMT : AKAKPGMOYLIVLRN : 348
CbSAMT : KTKFINVIVSLIRKSD : 359
          #

```

Figure C-4. Sequence alignment of CiSABATHs with AtFAMT and CbSAMT. Conserved residues are in shade with the more conserved the darker. Residues indicated with “&” are SAM/SAH-binding residues. Residues indicated with “*” are residues that interact with the carboxyl moiety of substrate. Residues indicated with “#” interact with the aromatic moiety of substrate and are important for substrate selectivity. CiSABATHs, *Cyperus iria* SABATHs; AtFAMT, *Arabidopsis thaliana* farnesoic acid methyltransferase; CbSAMT, *Clarkia breweri* salicylic acid methyltransferase.

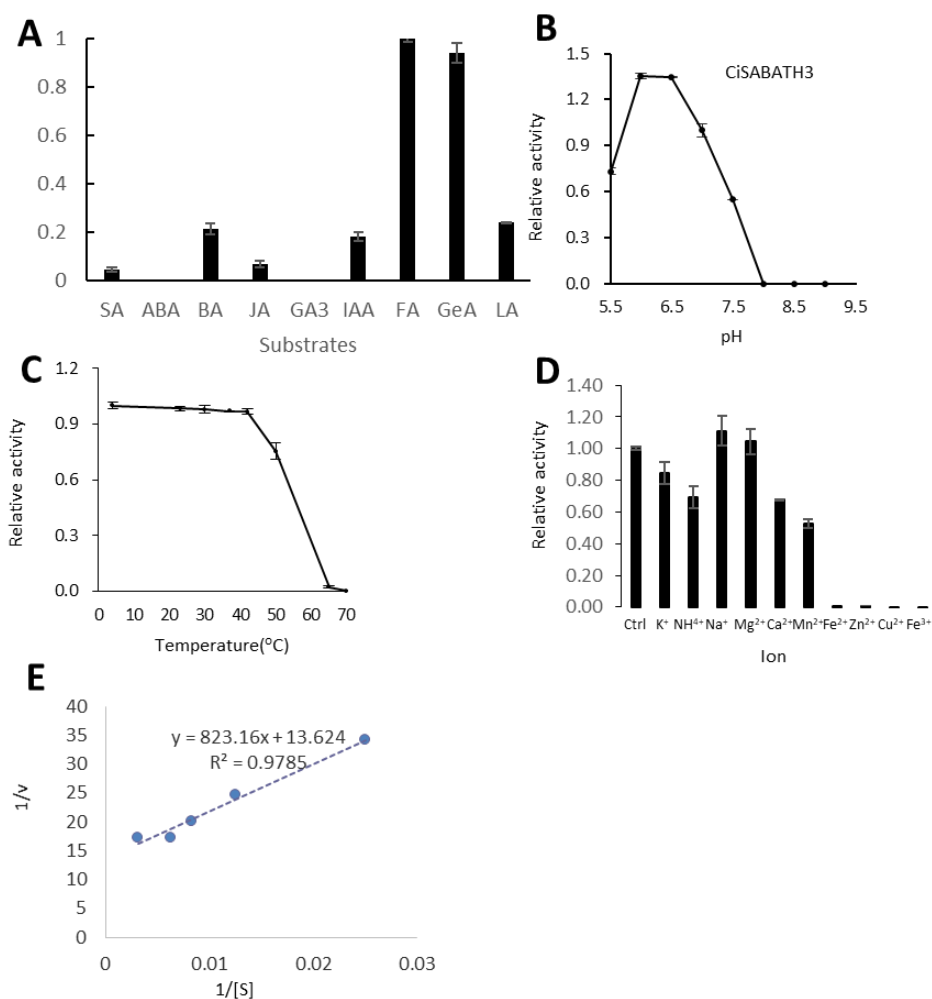
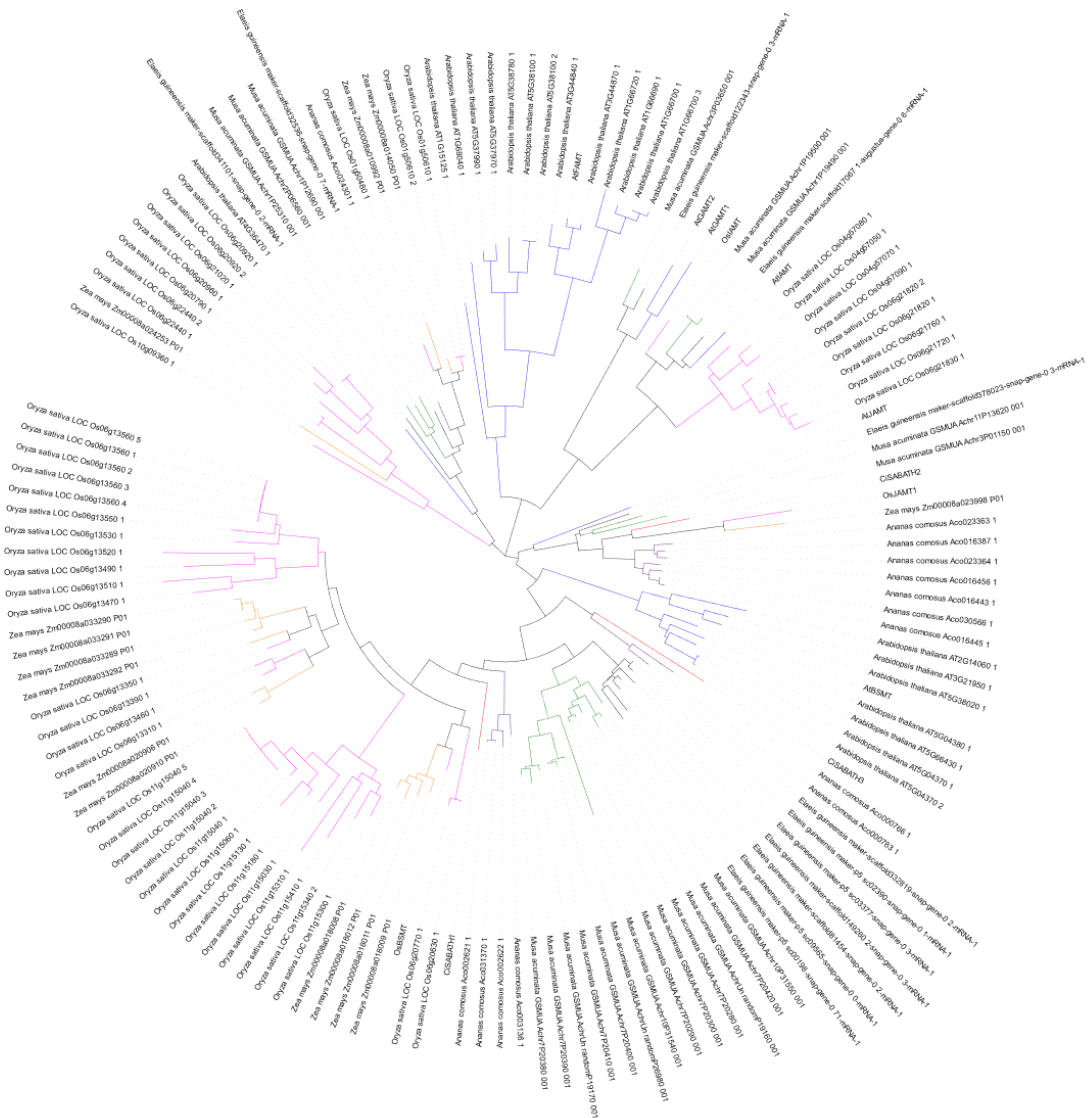


Figure C-5. Biochemical properties of CiSABATH3(CiFAMT) using FA as substrate. (A) Activity of CiSABATH3 with different substrates. The highest activity was arbitrarily set at 1.0. (B) pH effect on CiSABATH3 activity. The activity at pH 7.0 was arbitrarily set at 1.0. (C) Thermostability of CiSABATH3. The activity of CiSABATH3 with incubation at 0 oC for 30 min was arbitrarily set at 1.0. (D) Effects of metal ions on activity of CiSABATH3. The activity of CiSABATH3 without any metal ion added as control (Ctrl) was arbitrarily set at 1.00. (E) Km value of CiSABATH3.

Figure C-6. Phylogenetic tree of the SABATH family of methyltransferase in *C. iria* with SABATHs from selected monocots. The SABATH proteins from each species were color-coded. FAMTs: farnesoic acid methyltransferases; IAMTs: indole-3-acetic acid methyltransferases; BSMTs: benzoic acid/salicylic acid methyltransferases; JAMT: jasmonic acid methyltransferase. Proteins for different species are color coded.



- *Ananas comosus*
- *Arabidopsis thaliana*
- *Elaeis guineensis*
- *Musa acuminata*
- *Oryza sativa*
- *Zea mays*

**Chapter IV. Gibberellin Methyltransferase Gene is Ancient in Seed Plants
and Got Lost in Many Flowering Plants**

Adapted from:

Zhang Chi, Chaiprasongsuk Minta, Fu Jianyu, Chen Feng. Gibberellin Methyltransferase Gene is Ancient in Seed Plants and Got Lost in Many Flowering Plants. Drafted

Abstract

Methylation is the most recently discovered route of catabolism for plant hormone gibberellins (GAs). Previous to the present work, GA methyltransferases (GAMTs) of the SABATH family were identified only in *Arabidopsis*. Here we report that *GAMT* gene arose early in the evolution of seed plants and got lost in many flowering plants. By systematically searching the genome sequences of 260 land plants for *SABATH* genes and subjecting them to phylogenetic analysis, a putative ancient *GAMT* clade specific to seed plants was identified. Next, 24 representative *SABATH*s in the putative *GAMT* clade were selected for biochemical analysis. Nineteen of them were demonstrated to have GA methylating activity. Notably, more than 2/3 of the 260 plants do not contain an apparent ortholog in the *GAMT* clade, best exemplified by all 41 species of Poales among monocots and all 62 species in superasterids among eudicots analyzed. The loss of methylation-mediated GA catabolism in *GAMT*-lacking plants was further verified by the analysis of a representative monocot *Brachypodium distachyon*. It contains 11 *SABATH* genes but none encodes GA-methylating enzyme. While a single copy in most species when present, *GAMT* appears in multiple copies in most species in Brassicaceae, indicating lineage-specific retention of *GAMT* genes after duplication. In eight selected species, *GAMT* genes are mainly expressed in flowers and/or developing seeds, suggesting a conserved role of *GAMT* in the reproduction of *GAMT*-containing seed plants.

Keywords: Plant hormone, gene loss, reproduction, seed plants.

1. Introduction

Gibberellins (GAs) are a class of diterpenoid plant hormones ubiquitous in vascular plants (MacMillan, 2002). The physiological roles of bioactive GAs are best understood in flowering plants, where they are involved in regulating stem elongation, leaf expansion, flowering, fruit senescence and seed germination (Sun and Gubler, 2004; Hedden, 2016). Plants maintain appropriate levels of bioactive GAs in various tissues during growth and development through the regulation of their biosynthesis and catabolism (Hedden and Phillips, 2000; Olszewski et al., 2002). For biosynthesis, regulation may be exerted on spatiotemporal expression of one or more of the three types of GA biosynthetic genes: terpene synthases, cytochrome P450 monooxygenases, and 2-oxoglutarate-dependent dioxygenases (Yamaguchi, 2008). While relatively little is known about the regulation of GA catabolism, multiple routes exist (Figure D-1). The major route is 2 β -hydroxylation catalyzed by GA 2-oxidases (Thomas et al., 1999). Conjugation to form glucosyl esters and glucosides (Schneider et al., 1992) and epoxidation catalyzed by a cytochrome P450 monooxygenase (Zhu et al., 2006) are other known routes of GA catabolism. Methylation of the carboxyl group to form GA methylesters was the most recently discovered mechanism of GA catabolism (Varbanova et al., 2007).

Methylation of GAs is catalyzed by GA methyltransferase (GAMT), which transfers a methyl group from *S*-adenosyl-L-methionine to the carboxyl group of GAs to form GA methylesters (Varbanova et al., 2007). GAMT belongs to the methyltransferase family called SABATH (D'Auria et al., 2003). Other known members of the SABATH family that methylate phytohormones include indole-3-acetic acid methyltransferase (IAMT)

(Chaiprasongsuk, 2018; Qin et al., 2005; Zhao et al., 2009; Zhao et al., 2008; Zhao et al., 2007), salicylic acid methyltransferase (SAMT) (Barkman et al., 2007; Effmert et al., 2005; Fukami et al., 2002; Lin et al., 2013; Negre et al., 2002; Pott et al., 2002; Ross et al., 1999; Tieman et al., 2010; Zhao et al., 2013), and jasmonic acid methyltransferase (JAMT) (Meur et al., 2015; Preuß et al., 2014; Seo et al., 2001; Tieman et al., 2010; Zhao et al., 2013). *IAMT* has been demonstrated to be ancient and conserved in seed plants (Zhao et al., 2008), while *SAMT* and *JAMT* appear to have arisen multiple times during the evolution of seed plants (Tieman et al., 2010; Chaiprasongsuk et al., 2018). *GAMT* genes, on the other hand, so far have been identified only in *Arabidopsis thaliana*, whose genome contains two copies designated as *AtGAMT1* and *AtGAMT2* (Varbanova et al., 2007). The enzymes encoded by these two genes showed difference in substrate specificity: while both are active with multiple GAs, *AtGAMT1* was most active with GA₉ and GA₂₀ while *AtGAMT2* showed highest activity with GA₄ (Varbanova et al., 2007). Both *AtGAMT1* and *AtGAMT2* genes showed highest levels of expression during seed development. Using overexpression and knockout lines, the function of *AtGAMTs* in *Arabidopsis* was demonstrated to be deactivating or removing bioactive GAs during seed maturation (Varbanova et al., 2007). Transgenic tobacco, petunia, and tomato plants overexpressing *Arabidopsis GAMTs* exhibit the GA-deficient phenotype (Varbanova et al., 2007; Nir et al., 2014), supporting the role of *GAMT* in GA catabolism.

Both the biosynthetic and signal transduction pathways of GAs are highly conserved in plants (Vandenbussche et al., 2007; Yamaguchi, 2008; Binenbaum et al., 2018), which is consistent with the essential roles of GAs in regulating many aspects of plant physiology. It is

interesting to note that since its discovery in *Arabidopsis* more than a decade ago (Varbanova et al., 2007), no *GAMT* gene has been reported from any other plant. In our previous studies, the apparent orthologs of *Arabidopsis* *GAMTs* appear to be absent in some plants, such as rice (Zhao et al., 2008). This led to us to ask whether, similar to the situation with *IAMT*, an ancient and specific *GAMT* gene lineage exists in other plants (albeit with multiple losses in select plant taxa), whether *GAMT* activity is unique to *Arabidopsis*, or thirdly whether *GAMT* activity evolved independently multiple times in different plant lineages. To address these competing hypotheses, we began with a comparative genomic approach to analyze the *SABATH* family from a wide range of sequenced land plants and identified a putative *GAMT* clade that is specific to seed plants. Subsequently, we performed catalytic activity analysis, phylogenetic analysis and gene expression analysis to understand the origin, evolution and biological function of *GAMT*.

2. Materials and methods

2.1. Materials and reagents

The pollen cones and ovules of *Ginkgo biloba* trees grown in University of Tennessee Institute of Agriculture campus) and siliques of *Brassica rapa* (grown in University of Tennessee Institute of Agriculture North Greenhouse) were collected and used in this work. The pollen cones from male *G. biloba* and ovules from female *G. biloba* that emerged within three days respectively were harvested and immediately frozen in liquid nitrogen. The siliques of *Brassica rapa* were handled in the same way as those of *G. biloba*. All plant materials were

stored at -80 °C for later use. S-[methyl-¹⁴C]-Adenosyl-L-Methionine (¹⁴C-SAM) were purchased from PerkinElmer (Waltham, Massachusetts, USA). All other chemicals including two GA substrates (GA₃ and GA₄) were purchased from Sigma-Aldrich (St. Louis, MO, USA).

2.2. Sequence retrieval and analysis

All protein models of sequenced plant genomes were downloaded from Phytozome v12.1 (<https://phytozome.jgi.doe.gov/pz/portal.html>), *Brassica* Database (<http://brassicadb.org/brad/index.php>), Citrus Genome Database (CGD, <https://www.citrusgenomedb.org>), Cucurbit Genomics Database (CuGenDB, <http://cucurbitgenomics.org>), Hardwood Genomics Project (HWG, <https://www.hardwoodgenomics.org>), or respective databases referred in published research literature, (Table D-1) into a local server and searched using a HMM method, as previously described (Chaiprasongsuk et al., 2018).

2.3. Phylogenetic reconstruction

All newly identified SABATH methyltransferase sequences with a minimum length of 250 amino acids were used for phylogenetic reconstruction together with previously characterized SABATH protein sequences (Table D-2). Multiple protein sequence alignments were made with MAFFT version 7.369b under L-INS-I strategy (Kato and Standley, 2013). The phylogenetic tree was generated by RAxML v8.2 using the LG+G+F model with 1000 bootstraps (Stamatakis, 2014). The phylogenetic tree based on plant taxonomy was

constructed using phyloT (<https://phylot.biobyte.de>). All phylogenetic trees were visualized by iTOL (Letunic and Bork, 2016).

2.4. Gene cloning and enzyme assays

For genes from *Brassica rapa* and *Brachypodium distich*, plant tissues were transferred into 2 mL microcentrifuge tube and homogenized using TissueLyser II (QIAGEN). Total RNA was isolated with the RNeasy Plant Mini Kit (QIAGEN). An aliquot of 1.0 µg total RNA was used for cDNA synthesis with First Strand cDNA Synthesis Kit (GE Healthcare). Forward and reverse specific primers were used to amplify the complete coding sequence of GAMTs respectively (Table D-4). Each GAMT gene was cloned into pET-32a vector (MilliporeSigma) and confirmed by sequencing. For genes from other than *Brassica rapa* and *Brachypodium distich*, each coding DNA sequence (CDS) was synthesized (GenScript) and ligated into pET-32a vector (MilliporeSigma). Each construct was confirmed by sequencing. To express each GAMT protein, the protein expression construct transformed into *E. coli* strain BL21 (DE3) (Stratagene). Protein expression was induced by isopropyl β-D-1-thiogalactopyranoside (IPTG) at concentration of 0.5 M for 18 h at 25 °C, with cells lysed by sonication. Cell lysate was collected by centrifugation and purification of GAMTs was performed by nickel-nitrilotriacetic acid agarose (Invitrogen) as described by (Zhao et al., 2007). Radiochemical assays were performed with a 50 µL volume containing 50 mM Tris-HCl, pH 8.0, 1mM substrates, 3 µL ¹⁴C-SAM and 1 µL purified enzyme. The samples were incubated at 30 °C for 30 min and followed by extraction with 150 µL EtOAc. After phase separation by 1 min centrifugation at 14,000g, 75 µL of EtOAc phase were transferred to a scintillation vial containing 2mL of cocktail and counted in a scintillation counter (Beckman

Coulter).

2.5. Gene expression data retrieval

The gene expression data of *Ginkgo biloba* was retrieved from <http://gigadb.org/dataset/100209> (Guan et al., 2016). The gene expression data of *Arabidopsis thaliana*, *Camelina sativa* and *Vitis vinifera* were analyzed through <http://bar.utoronto.ca/> (Fucile et al., 2011). The gene expression data of *Picea abies* was retrieved from <http://congenie.org> (Sundell et al., 2015). The gene expression data of *Phalaenopsis equestris* was retrieved from <http://orchidstra2.abrc.sinica.edu.tw> (Chao et al., 2017). The gene expression data of *Musa acuminata* was retrieved from <https://banana-genome-hub.southgreen.fr> (Droc et al., 2013). The gene expression data of *Citrus sinensis* was retrieved from <http://citrus.hzau.edu.cn> (Wang et al., 2014).

2.6. Accession numbers

Sequence data from this article can be accessed with request.

3. Results and discussion

3.1. Comparative analysis of the SABATH family in 260 sequenced land plants

To answer the question about the origin and evolution of the *GAMT* gene, we took a comparative genomics approach. We first compiled the land plants with sequenced genomes in public domains. A total of 260 plants species were identified (Table D-1). Then, the

complete proteome for each of the 260 sequenced plants were downloaded to a local server and the entire dataset was searched for SABATH proteins. A total of 6458 SABATH proteins were identified with an average of 25 proteins per plant genome. The sizes of the SABATH family ranged from 1 (*Apostasia shenzhenica* and *Pogostemon cablin*) to 115 (*Triticum aestivum*). Next, the SABATH proteins were subject to phylogenetic analysis. SABATHs from seed plants were placed into five groups (I to V) (Figure D-2A). Group I contains SABATHs from all major lineages of land plants bryophytes, lycophytes, ferns gymnosperms and angiosperms. More interestingly, *Arabidopsis* GAMT1 and GAMT2, the only two known GAMTs, belong to group I. Group II genes are specific to seed plants. It is interesting to note that several IAMTs identified from *Arabidopsis*, rice and poplar and spruce all belong to group II. Group III is specific to gymnosperms. In contrast, both groups IV and V are specific to angiosperms. Among the four groups that contains SABATHs from angiosperms (I, II, IV and V), group I contain the least number of SABTAH proteins.

With the two known *Arabidopsis* GAMTs belonging to group I, subsequent work presented here focused on this group. The phylogenetic division of SABATHs within group I (Figure D-2B) showed that the *Arabidopsis* GAMTs belong to a clade that is specific to angiosperms. In our current understanding, the ability to synthesize GAs evolved in vascular plants. Nonvascular plants are able to synthesize only the precursors of GAs (Hirano et al., 2007), and they should also not be able to synthesize GA methylesters. We therefore hypothesized that this clade is likely to include *GAMT* genes from other angiosperm species

3.2. Catalytic activity of selected members in the GAMT clade and related non-GAMT enzymes

To determine whether any of the genes other than *AtGAMT1* and *AtGAMT2* in this putative *GAMT* clade specify enzymes with GAMT activity, we undertook biochemical analysis with selected candidate enzymes. First, we chose to analyze the three GAMT genes in *Brassica rapa*, a species that belongs to Brassicaceae as *Arabidopsis*. Full-length cDNAs for these three GAMTs were amplified by RT-PCR and spliced into a protein expression vector (pEXP-5-CT/TOPO). After expression in *Escherichia coli*, recombinant proteins were tested for methyltransferase activity in *in vitro* assays using GA₃ and GA₄, two of the most widely occurring bioactive GAs (MacMillan, 2001), as substrates. Two of the three proteins showed activity towards Gas, with one of them preferred GA₃ as substrate and the other one preferred GA₄ as substrate (Table D-5).

With the validation of GAMT enzymatic activity for proteins encoded by two SABATH genes from *B. rapa*, next we selected a total 21 of putative GAMT genes from 19 non-Brassicaceae species for functional analysis (Table D-5). A full-length cDNA for each of the 21 genes was spliced into a protein expression vector and expressed in *E. coli*. Sixteen of the 21 SABATH proteins showed activity with GA₄ (Table D-5). At the same time, 9 of the 21 recombinant SABATH proteins had MT catalytic activity with GA₃ as a substrate. None of the 16 proteins with GAMT activity showed activity with IAA, JA or SA as substrates, indicating that these GAMTs have strict substrate specificity.

Because the evidence indicated that proteins with GAMTs catalytic activity were present in diverse angiosperm lineages, we next assessed whether GAMT activity evolved

before or after the origin of seed plant evolution. Based on our phylogenetic analysis, seed plant GAMTs formed a monophyletic group and this group clustered with SABATHs from a number of non-seed producing plants (Figure D-2B). In previous studies, we demonstrated that the SABATHs in this group from the moss *Physcomitrella patens* Hedw. does not have GAMT activity (Zhao et al., 2013). Therefore, to examine this question further, we selected SmSABATH1 and SmSABATH4 from *Selaginella moellendorffii* Hieron., which show higher sequenced homology to GAMTs among the *S. moellendorffii* SABATHs and tested the proteins they encode for GAMT activity. Neither of them showed GAMT activity with either GA₃ or GA₄.

3.3. Apparent ortholog of GAMT gene is absent in about 2/3 of the 260 plants

The phylogenetic relationship of the SABATHs in the *GAMT* clade of Group I was compared to the species phylogeny. The *GAMT* gene tree (Figure D-3A) is congruent with the species tree established by APG IV (2016) (Figure D-3B), implying that *GAMT* is a conserved gene in seed plants. On the other hand, a *GAMT* gene was not detected in 185 of the 260 plants (Figure D-4). When mapped to the species tree, the absence of *GAMT* genes occurred in 185 species. In one extreme case, no *GAMT* genes were identified in all 41 species in the order Poales. Among eudicots, *GAMT* genes appeared to be absent in all 22 species in Fabales and all 62 species in Superasterids. The absence of *GAMT* genes in certain plant species could be due to incomplete coverage and/or poor quality of genome sequencing. Nonetheless, their absence in lineages with a large number of plant species having sequenced genomes (e.g., Poales and superasterids) can be concluded with confidence.

3.4. *GAMT*-lacking *Brachypodium distachyon* lacks a *GA*-methylating gene

With the absence of an orthologous *GAMT* gene in close to 70% of the plant species analyzed in this study, we examined whether *GAMT* catalytic activity may have been maintained by convergent evolution of other genes arising from other *SABATH* clades. To test this possibility, we chose *Brachypodium distachyon* (L.) P.Beauv, a monocot in Poales, as a model species. The *B. distachyon* genome contains 11 *SABATH* genes (Table D-3). Full-length cDNA for all 11 *SABATH* genes were expressed in *E. coli*, and their respective recombinant proteins tested with GA₃ and GA₄. None of the 11 *SABATH*s proteins investigated had activity with GA₃ or GA₄, supporting the hypothesis that *GAMT* gene has been lost in the *GAMT*-absent plants.

3.5. Retention after duplication of *GAMT* genes in Brassicaceae

While *GAMT* genes appear to be present in a single copy in most plant species (when *GAMT* genes do occur), it is evident that *GAMT* genes can exist in two or more copies within members of Brassicaceae (Figure D-5). In this family, 17 of the 21 species examined contains two *GAMT* genes, including *A. thaliana*. *B. rapa*, *B. oleracea*, *B. napus* contain 4, 4 and 9 *GAMT* genes, respectively. This suggests that two copies of *GAMT* genes resulted from a whole genome duplication event that occurred within the common ancestor of Brassicaceae, which is known as At- α (Cardinal-McTeague et al., 2016). The genus *Brassica* shows additional copies of *GAMT*, which can be attributed to a whole genome triplication event unique to Brassica (Cheng et al., 2014). *B. napus* is a recent allopolyploid obtained by a cross between *B. oleracea* and *B. rapa* (Chalhoub et al., 2014). Consistent with this evolutionary

history, each orthologous pair of GAMTs has one copy in *B. oleracea*, one copy in *B. rapa*, and two copies in *B. napus* (Figure D-5). *Camelina sativa* (L.) Crantz contains 7 copies of GAMT genes. This is also likely the result of genome amplification, as *C. sativa* is believed to have undergone a genome triplication event (Kagale et al., 2014).

Genome duplication is common in evolution of land plants. This evolutionary outcome is an important mechanism that enables gene duplication and expansion within species. It is interesting that *GAMT* remains a single copy gene in most non-Brassicaceae lineages. The doubling or further amplification of *GAMT* genes in Brassicaceae suggests that some of the Brassicaceae *GAMT* genes may have acquired specificity towards different GAs, as demonstrated for the two *GAMTs* in *Arabidopsis* (Varbanova et al., 2007) and the three *GAMTs* in *B. rapa* (Table D-5).

3.6. Expressed patterns of *GAMT* genes in selected plant species

To gain insight into the biological process in which *GAMT* is involved, we evaluated the expression patterns of *GAMT* genes in eight species, *Ginkgo biloba*, *Picea abies*, *Phalaenopsis equestris*, *Musa acuminata*, *Vitis vinifera*, *Citrus sinensis*, *Arabidopsis thaliana* and *Camelina sativa*, using available open-source expression databases (Figure D-6). In *G. biloba*, only one of two putative *GAMTs* showed activity with GAs and its *bona fide GAMT* gene was expressed mainly in ovules that developed into seeds after successful pollination (Figure D-6). A similar expression pattern was observed in another gymnosperm, *Picea abies*. In *P. equestris*, *GAMT* is mainly expressed in the flower, especially in the labellum. In *M. acuminata*, *GAMT* expression was observed in the fruit, with higher transcript levels detected

during ripening. In grapevine, its *GAMT* gene showed highest level of expression in senesced leaves. It also showed expression in young flowers, roots, and pericarp tissues. In *C. sinensis*, *GAMT* is mainly expressed in flowers. There are seven *GAMT* genes in the *C. sativa* genome. They all expressed mainly during reproductive growth, with five of them showing the highest level of expression in early or early-mid stage of seed development, while the other three genes showed the highest levels of expression in flowers (Figure D-6).

4. Conclusions

By analyzing *SABATH* genes across a wide spectrum of land plants, ranging from basal lineage (liverwort, moss), non-seed vascular plants (lycophyte and ferns) to gymnosperms (6 species) and angiosperms (248 species), and by conducting *in vitro* enzymatic assays, we identified a *GAMT* clade that arose early in the evolution of seed plants. This large-scale comparative genomic analysis of the *SABATH* family further revealed that *GAMT* genes appear to have been lost in many flowering plants, in large lineages including Poales, asterids and Caryophyllales (Figure D-4). By contrast, multiple *GAMT* genes are present in the genomes of Brassicaceous plants, and *GAMT* enzymes from multiple lineages appear to have strict substrate specificity. Furthermore, our analysis of the expression patterns of *GAMT* genes from multiple species suggests that *GAMT* genes have conserved biological functions. In all plant species analyzed, except grapevine, the highest level of expression of *GAMT* genes was detected in reproductive organs (Figure D-6). In grapevine, *GAMT* showed the highest level of expression in senesced leaves. In addition to leaves, the highest level of expression was also detected in reproductive organs. Future investigations should look at the

specific biological function of *GAMT* in diverse plants, and how these functions were compensated in species where *GAMT* gene loss has occurred or multiple *GAMT* genes have evolved. Finally, many major crop species do not contain *GAMT* genes. As such, *GAMT* gene may be a useful new molecular tool for genetic improvement of these *GAMT*-lacking crops.

Reference

Barkman, T. J., Martins, T. R., Sutton, E., Stout, J. T., 2007. Positive selection for single amino acid change promotes substrate discrimination of a plant volatile-producing enzyme. *Molecular biology and evolution* 24, 1320-1329.

Binenbaum J, Weinstain R, Shani E (2018) Gibberellin localization and transport in plants. *Trends in plant science*

Cardinal-McTeague WM, Sytsma KJ, Hall JC (2016) Biogeography and diversification of Brassicales: a 103 million year tale. *Molecular phylogenetics and evolution* 99: 204-224

Chaiprasongsuk M, Zhang C, Qian P, Chen X, Li G, Trigiano RN, Guo H, Chen F (2018) Biochemical characterization in Norway spruce (*Picea abies*) of SABATH methyltransferases that methylate phytohormones. *Phytochemistry* 149: 146-154

Chalhoub B, Denoeud F, Liu S, Parkin IA, Tang H, Wang X, Chiquet J, Belcram H, Tong C, Samans B (2014) Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. *science* 345: 950-953

Chao Y-T, Yen S-H, Yeh J-H, Chen W-C, Shih M-C (2017) Orchidstra 2.0—a transcriptomics resource for the orchid family. *Plant and Cell Physiology* 58: e9-e9

Cheng F, Wu J, Wang X (2014) Genome triplication drove the diversification of Brassica plants. *Horticulture research* 1: 14024

D'Auria JC, Chen F, Pichersky E (2003) Chapter eleven the SABATH family of MTS in *Arabidopsis thaliana* and other plant species. *In* Recent advances in phytochemistry, Vol 37. Elsevier, pp 253-283

Droc G, Lariviere D, Guignon V, Yahiaoui N, This D, Garsmeur O, Dereeper A, Hamelin C, Argout X, Dufayard J-F (2013) The banana genome hub. Database 2013

Effmert, U., Saschenbrecker, S., Ross, J., Negre, F., Fraser, C. M., Noel, J. P., Dudareva, N., Piechulla, B., 2005. Floral benzenoid carboxyl methyltransferases: From in vitro to in planta function. *Phytochemistry* 66, 1211-1230.

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 and cell physiology* 43, 1054-1058.

Fucile G, Di Biase D, Nahal H, La G, Khodabandeh S, Chen Y, Easley K, Christendat D, Kelley L, Provart NJ (2011) ePlant and the 3D data display initiative: integrative systems biology on the world wide web. *PLoS One* 6: e15237

Guan R, Zhao Y, Zhang H, Fan G, Liu X, Zhou W, Shi C, Wang J, Liu W, Liang X (2016) Draft genome of the living fossil *Ginkgo biloba*. *Gigascience* 5: 49

Hirano K, Nakajima M, Asano K, Nishiyama T, Sakakibara H, Kojima M, Katoh E, Xiang H, Tanahashi T, Hasebe M, Banks JA (2007) The GID1-mediated gibberellin perception mechanism is conserved in the lycophyte *Selaginella moellendorffii* but not in the bryophyte *Physcomitrella patens*. *The Plant Cell* 19: 3058-79

Hedden P (2016) Annual plant reviews, the gibberellins, Vol 49. John Wiley & Sons

Kagale S, Koh C, Nixon J, Bollina V, Clarke WE, Tuteja R, Spillane C, Robinson SJ, Links MG, Clarke C (2014) The emerging biofuel crop *Camelina sativa* retains a highly undifferentiated hexaploid genome structure. *Nature communications* 5: 3706

Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution* 30: 772-780

Letunic I, Bork P (2016) Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic acids research* 44: W242-W245

Lin, J., Mazarei, M., Zhao, N., Zhu, J. J., Zhuang, X., Liu, W., Pantalone, V. R., Arelli, P. R., Stewart Jr, C. N., Chen, F., 2013. Overexpression of a soybean salicylic acid methyltransferase gene confers resistance to soybean cyst nematode. *Plant biotechnology journal* 11, 1135-1145.

MacMillan J (2001) Occurrence of gibberellins in vascular plants, fungi, and bacteria. *Journal of plant growth regulation* 20: 387-442

Meur, G., Shukla, P., Dutta-Gupta, A., Kirti, P., 2015. Characterization of *Brassica juncea*–*Alternaria brassicicola* interaction and jasmonic acid carboxyl methyl transferase expression. *Plant gene* 3, 1-10.

Negre, F., Kolosova, N., Knoll, J., Kish, C. M., 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. *Archives of biochemistry and biophysics* 406, 261-270.

Nir I, Moshelion M, Weiss D (2014) The *Arabidopsis* GIBBERELLIN METHYLTRANSFERASE 1 suppresses gibberellin activity, reduces whole-plant transpiration and promotes drought tolerance in transgenic tomato. *Plant, cell & environment* 37: 113-123

Pott, M. B., Pichersky, E., Piechulla, B., 2002. Evening specific oscillations of scent emission, SAMT enzyme activity, and SAMT mRNA in flowers of *Stephanotis floribunda*. *Journal of Plant Physiology* 159, 925-934.

Preuß, A., Augustin, C., Figueroa, C. R., Hoffmann, T., Valpuesta, V., Sevilla, J. F., Schwab, W., 2014. Expression of a functional jasmonic acid carboxyl methyltransferase is negatively correlated with strawberry fruit development. *Journal of plant physiology* 171, 1315-1324.

Qin, G., Gu, H., Zhao, Y., Ma, Z., Shi, G., Yang, Y., Pichersky, E., Chen, H., Liu, M., Chen, Z., 2005. An indole-3-acetic acid carboxyl methyltransferase regulates *Arabidopsis* leaf development. *The Plant Cell* 17, 2693-2704.

Ross, J. R., Nam, K. H., D'Auria, J. C., 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. *Archives of biochemistry and biophysics* 367, 9-16.

Seo, H. S., Song, J. T., Cheong, J.-J., Lee, Y.-H., Lee, Y.-W., Hwang, I., Lee, J. S., Do Choi, Y., 2001. Jasmonic acid carboxyl methyltransferase: a key enzyme for jasmonate-regulated plant responses. *Proceedings of the National Academy of Sciences* 98, 4788-4793.

Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30: 1312-1313

Sun T-p, Gubler F (2004) Molecular mechanism of gibberellin signaling in plants. *Annu. Rev. Plant Biol.* 55: 197-223

Sundell D, Mannapperuma C, Netotea S, Delhomme N, Lin YC, Sjödin A, Van de Peer Y, Jansson S, Hvidsten TR, Street NR (2015) The Plant Genome Integrative Explorer Resource: PlantGen IE. org. New Phytologist 208: 1149-1156

Tieman D, Zeigler M, Schmelz E, Taylor MG, Rushing S, Jones JB, Klee HJ (2010) Functional analysis of a tomato salicylic acid methyl transferase and its role in synthesis of the flavor volatile methyl salicylate. The Plant Journal 62: 113-123

Vandenbussche F, Fierro AC, Wiedemann G, Reski R, Van Der Straeten D (2007) Evolutionary conservation of plant gibberellin signalling pathway components. BMC Plant Biology 7: 65

Varbanova M, Yamaguchi S, Yang Y, McKelvey K, Hanada A, Borochoy R, Yu F, Jikumaru Y, Ross J, Cortes D (2007) Methylation of gibberellins by *Arabidopsis* GAMT1 and GAMT2. The Plant Cell 19: 32-45

Wang J, Chen D, Lei Y, Chang J-W, Hao B-H, Xing F, Li S, Xu Q, Deng X-X, Chen L-L (2014) *Citrus sinensis* annotation project (CAP): a comprehensive database for sweet orange genome. PloS one 9: e87723

Yamaguchi S (2008) Gibberellin metabolism and its regulation. Annu. Rev. Plant Biol. 59: 225-251

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

Zhao N, Ferrer J-L, Moon HS, Kapteyn J, Zhuang X, Hasebe M, Stewart Jr CN, Gang DR, Chen F (2012) A SABATH Methyltransferase from the moss *Physcomitrella patens* catalyzes S-methylation of thiols and has a role in detoxification. *Phytochemistry* 81: 31-41

Zhao N, Ferrer J-L, 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 physiology* 146: 455-467

Zhao N, Guan J, Lin H, Chen F (2007) Molecular cloning and biochemical characterization of indole-3-acetic acid methyltransferase from poplar. *Phytochemistry* 68: 1537-1544

Zhao, N., Yao, J., Chaiprasongsuk, M., Li, G., Guan, J., Tschaplinski, T. J., Guo, H., Chen, F., 2013. Molecular and biochemical characterization of the jasmonic acid methyltransferase gene from black cottonwood (*Populus trichocarpa*). *Phytochemistry* 94, 74-81.

Appendix D: Figures and Tables

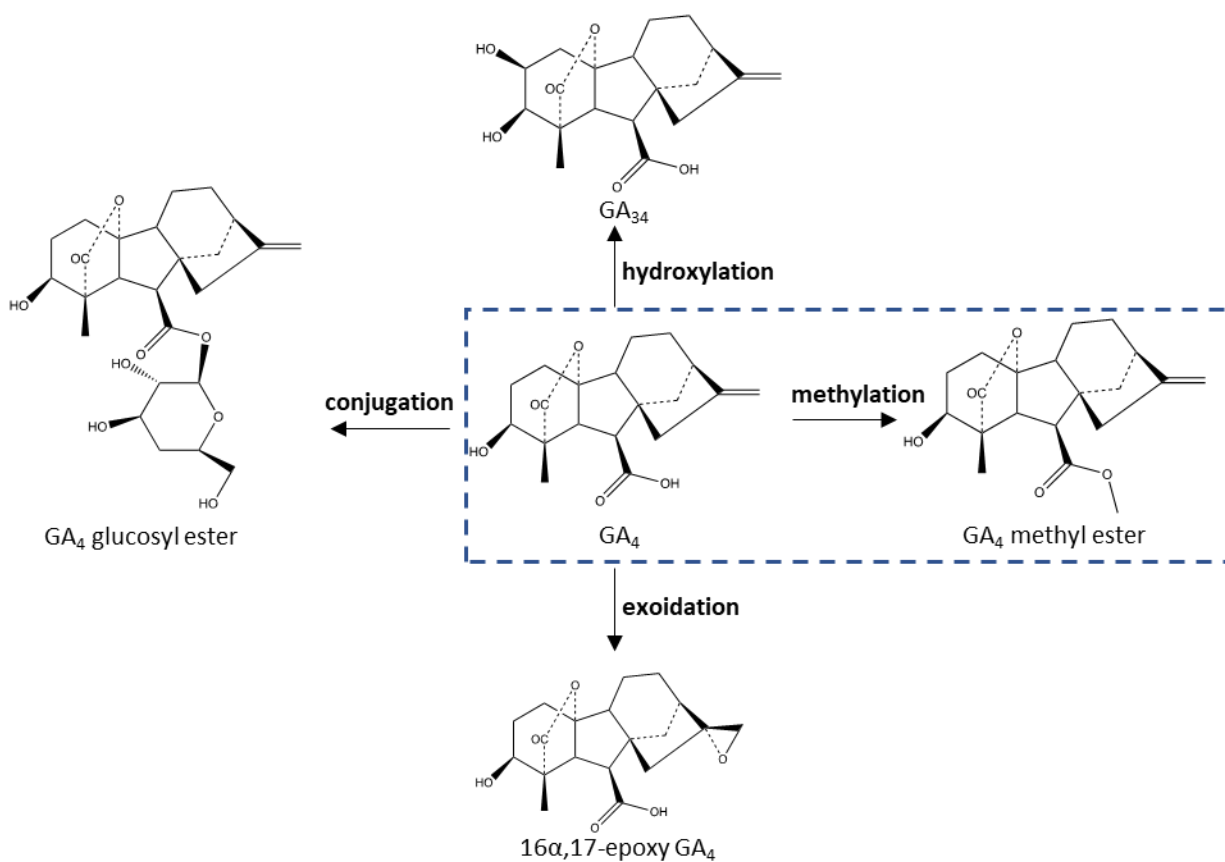


Figure D-1. Known routes of catabolism of gibberellins (GAs). GA₄ is one widely occurring bioactive GA. Methylation is the most recently discovered mechanism.

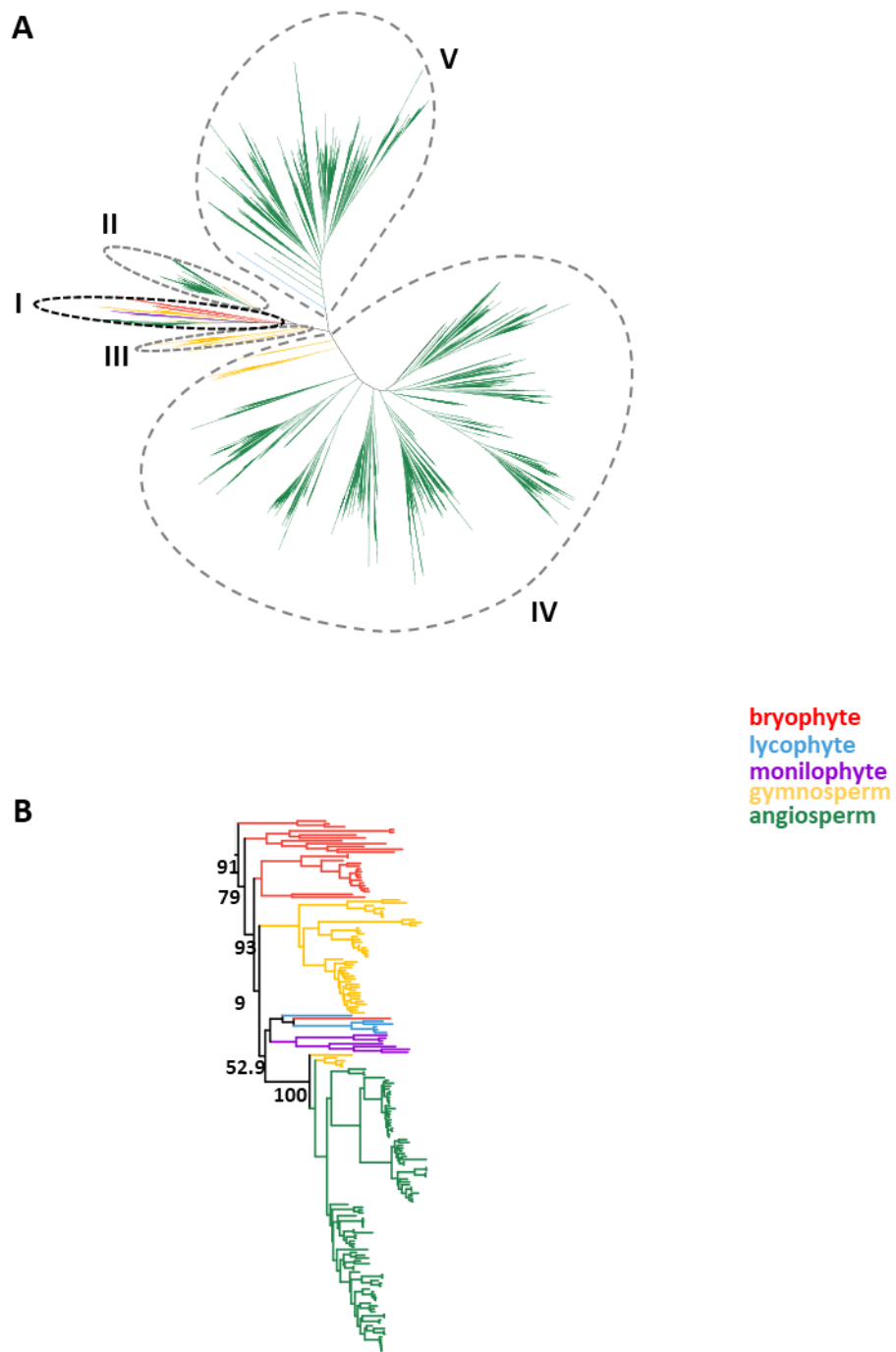


Figure D-2. Phylogenetic analysis of SABATH proteins. (A) Phylogenetic tree SABATH methyltransferases from 260 sequenced plants with five recognized groups; (B) phylogeny of group I is shown with bootstrap value.

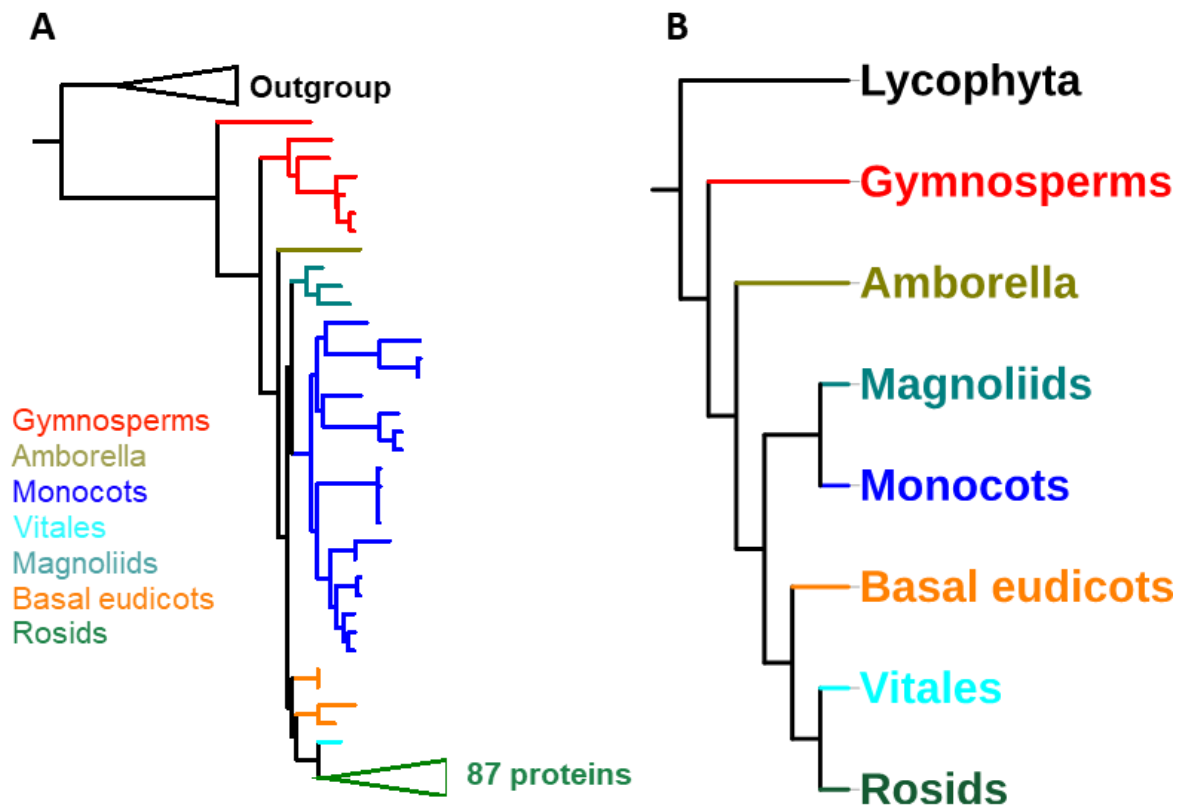


Figure D-3. Phylogenetic tree of the SABATH protein in the putative *GAMT* clade (A) and species tree (B).

Figure D-4. Presence/absence of SABATHs of the *GAMT* clade among families of sequenced plants. Three taxa with complete loss of *GAMTs* were shaded. Presence/absence of SABATH genes in the *GAMT* clade among families of sequenced plants. The phylogeny was redrawn from APG IV (2016). The lineages in gray include that no species from those lineages was analyzed in this study. The two numbers (red and black) includes in the number of species containing *GAMT* gene and the total number of species from that specific lineages analyzed in this study.

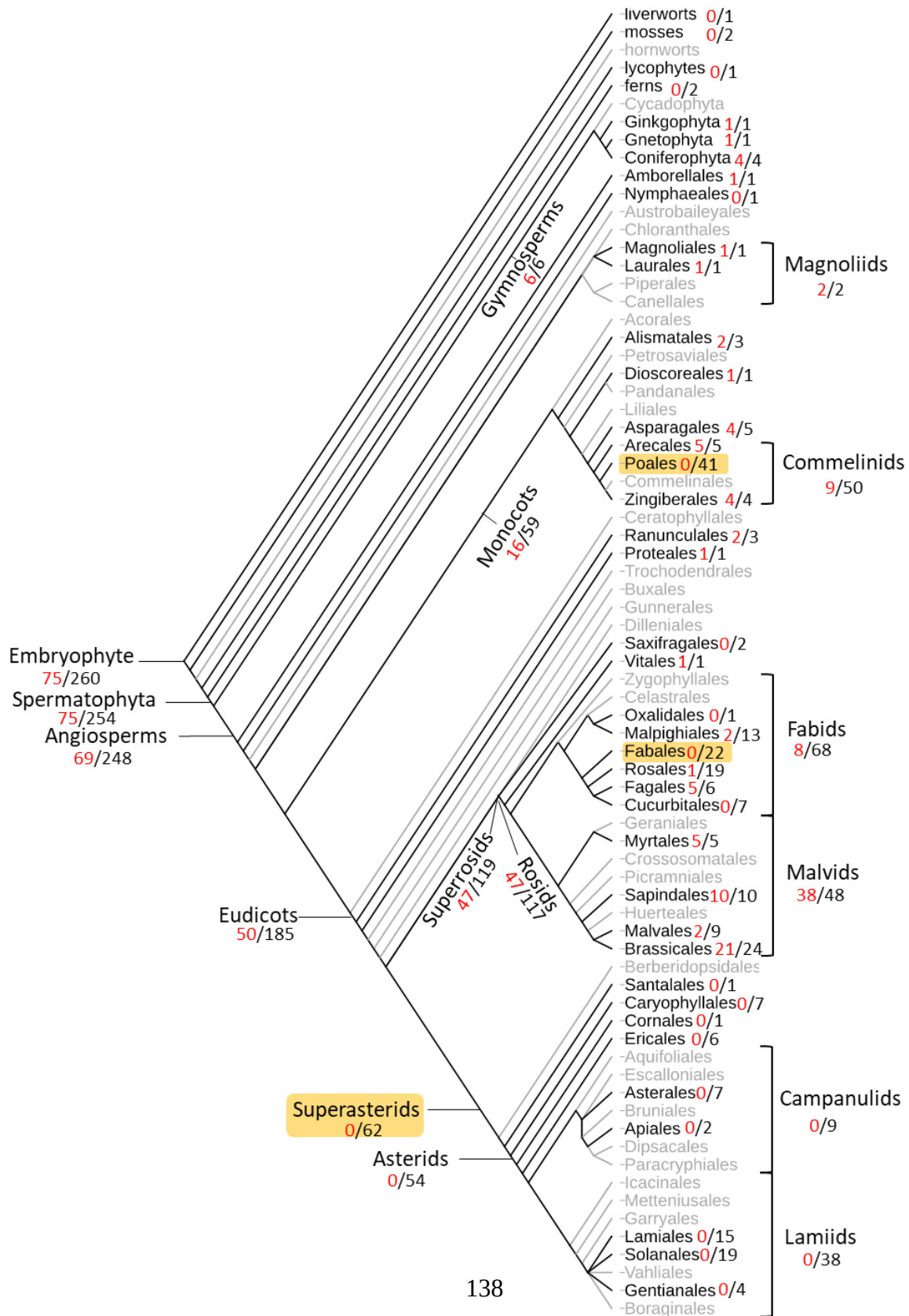




Figure D-5. Phylogenetic tree of SABATHs of the *GAMT* clade in Brassicaceae. The two clades in blue and in yellow depicts two clades that resulted from a whole gene duplication event occurred in the ancestor of Brassicaceae. The two smaller blocks indicate gene duplication due to Brassica-specific whole genome duplication.

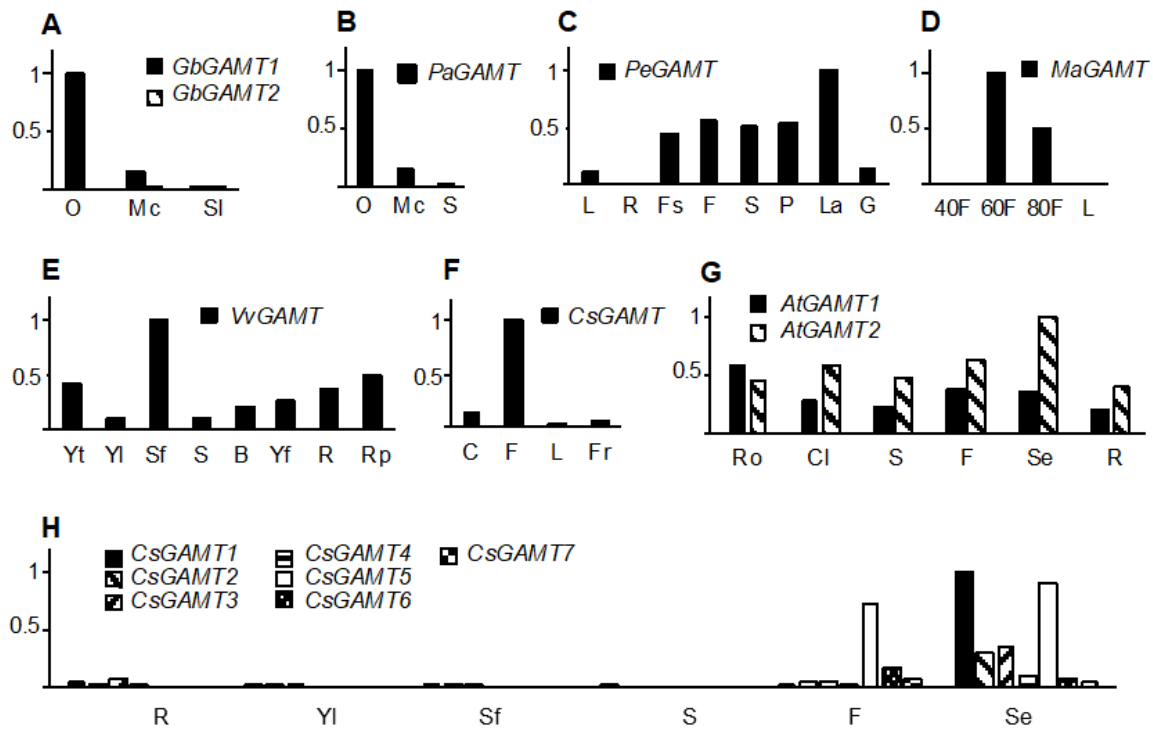


Figure D-6. Expression patterns of *GAMTs* in different selected species. (A) *Gingo biloba*; (B) *Picea abies*; (C) *Phalaenopsis equestris*; (D) *Musa acuminata*; (E) *Vitis vinifera*; (F) *Citrus sinensis*; (G) *Arabidopsis thaliana*; (H) *Camelina sativa*. O: Ovules; Mc: Male cones; Sl: Stem and leaves; S: Stem; L: Leaf; R: Root; Fs: Floral stalk; F: Flower; S: Sepal; P: Petal; La: Labellum; G: Gynostemium; 40F: 40-day-fruit; 60F: 60-day-fruit; 80F: 80-day-fruit; Yt: Young tendril; Yl: Young leaf; Sf: Senescent leaf; B: Bud; Yf: Young flower; Rp: Ripening pericarp; C: Callus; Fr: Fruit; Ro: Rosette; Cl: Cauline leaf; Se: Seed; The highest level of expression was arbitrarily set at 1.0. Enzyme expression profiles were obtained from open-source databases, as described above.

Figure D-6. Gene expression pattern of seven *GAMTs* from *Camelina sativa*. (A) to (G) are corresponding with *CsGAMT1-7* in Figure 6(H). Gene ID was listed above each figure of expression.

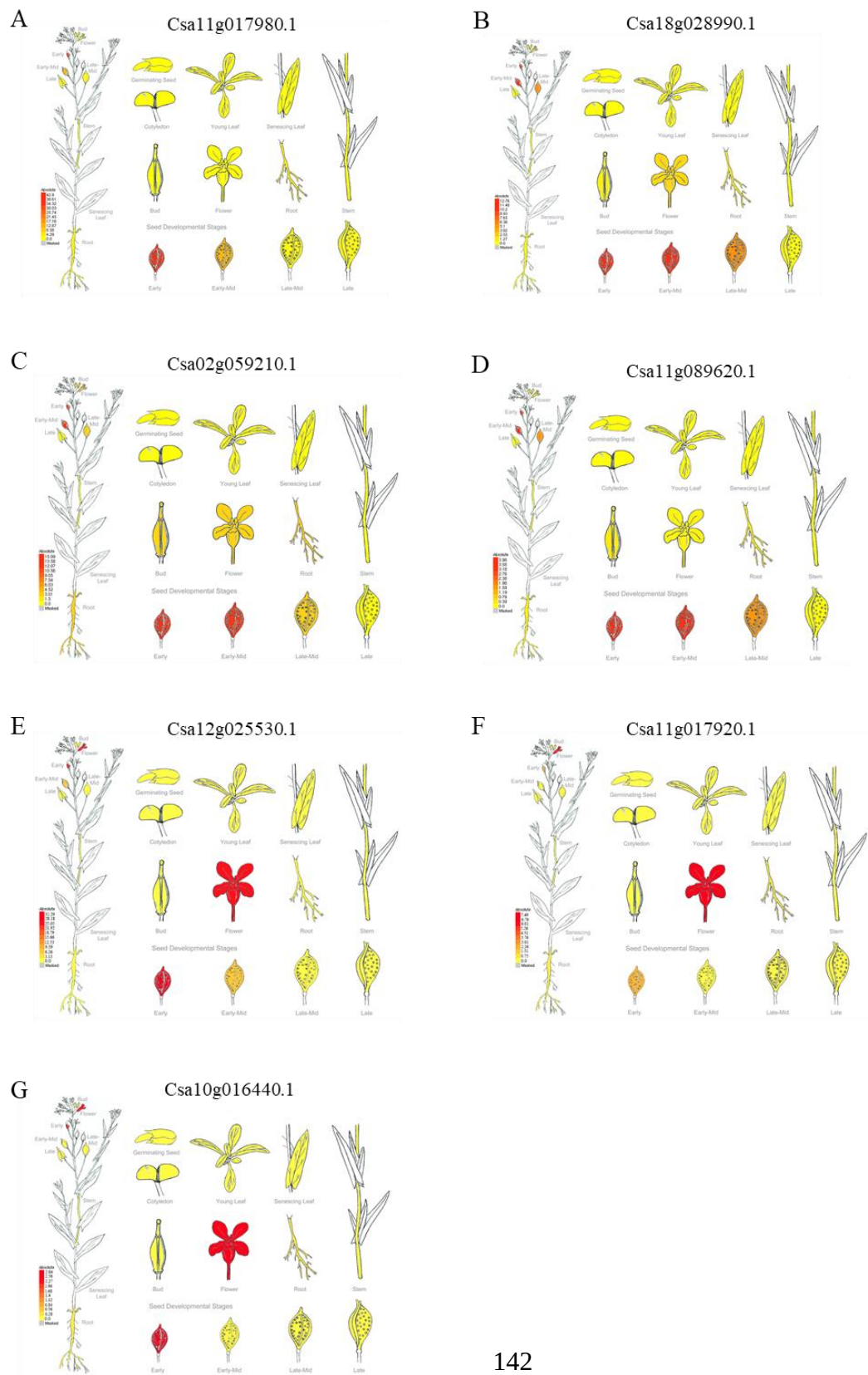


Table D-1. Plant genomes used in phylogenetic analysis. The source of each genome used in this study was listed with species name and database or original literature.

Plant species	Depository	Literature References
<i>Actinidia chinensis</i>	NCBI Genome	
<i>Aegilops tauschii</i>	Phytozome	
<i>Aethionema arabicum</i>	Brassica Database	
<i>Amaranthus hypochondriacus</i>	Phytozome	
<i>Amborella trichopoda</i>	Phytozome	
<i>Ammopiptanthus nanus</i>		(Gao et al., 2018)
<i>Anacardium occidentale</i>	Phytozome	
<i>Ananas comosus</i>	Phytozome	
<i>Apostasia shenzhenica</i>		(Zhang et al., 2017)
<i>Antirrhinum majus</i>		(Li et al., 2019)
<i>Aquilegia coerulea</i>	Phytozome	
<i>Arabidopsis halleri</i>	Phytozome	
<i>Arabidopsis lyrata</i>	Phytozome	
<i>Arabidopsis thaliana</i>	Phytozome	
<i>Arachis duranensis</i>		(Bertioli et al., 2015)
<i>Arachis ipaensis</i>		(Bertioli et al., 2015)
<i>Arachis monticola</i>		(Yin et al., 2018)
<i>Artemisia annua</i>		(Lin et al., 2017)
<i>Artocarpus camansi</i>		(Gardner et al., 2016)
<i>Asparagus officinalis</i>	Phytozome	
<i>Atalantia buxifolia</i>		(Wang et al., 2017)
<i>Avicennia marina</i>		http://evolution.sysu.edu.cn/Sequences.html
<i>Azolla filiculoides</i>		(Li et al., 2018)
<i>Barbarea vulgaris</i>		(Byrne et al., 2017)
<i>Beta vulgaris</i>	Phytozome	
<i>Doroceras hygrometricum</i>		(Xiao et al., 2015)
<i>Bombax ceiba</i>		(Gao et al., 2018)
<i>Boechera retrofracta</i>		(Kliver et al., 2018)
<i>Boechera stricta</i>	Phytozome	
<i>Boehmeria nivea</i>		(Luan et al., 2018)
<i>Brachypodium distachyon</i>	Phytozome	

Table D-1 (continued)

Plant species	Depository	Literature References
<i>Brachypodium hybridum</i>	Phytozome	
<i>Brachypodium stacei</i>	Phytozome	
<i>Brachypodium sylvaticum</i>	Phytozome	
<i>Brassica napus</i>	Brassica Database	
<i>Brassica oleracea</i>	Brassica Database	
<i>Brassica rapa</i>	Brassica Database	
<i>Cajanus cajan</i>		(Varshney et al., 2012)
<i>Calamus jenkinsianus</i>		(Zhao et al., 2018)
<i>Calamus simplicifolius</i>		(Zhao et al., 2018)
<i>Calotropis gigantea</i>		(Hoopes et al., 2018)
<i>Camelina sativa</i>	Brassica Database	
<i>Camellia sinensis</i>		(Wei et al., 2018)
<i>Camptotheca acuminata</i>		(Zhao et al., 2017)
<i>Cannabis sativa</i>		(van Bakel et al., 2011)
<i>Capsella grandiflora</i>		(Slotte et al., 2013)
<i>Capsella rubella</i>	Brassica Database	
<i>Capsicum annuum</i>		(Kim et al., 2014)
<i>Capsicum baccatum</i>		(Kim et al., 2017)
<i>Capsicum chinense</i>		(Kim et al., 2017)
<i>Cardamine hirsuta</i>		(Gan et al., 2016)
<i>Carica papaya</i>	Phytozome	
<i>Castanea mollissima</i>	HWG	
<i>Catharanthus roseus</i>		(Kellner et al., 2015)
<i>Cephalotus follicularis</i>		(Fukushima et al., 2017)
<i>Chenopodium quinoa</i>	Phytozome	
<i>Chrysanthemum indicum</i>		(Song et al., 2018)
<i>Cicer arietinum</i>	Phytozome	
<i>Cicer reticulatum</i>		(Gupta et al., 2016)
<i>Cinnamomum micranthum</i>	NCBI Genome	
<i>Citrullus lanatus</i>		(Guo et al., 2013)
<i>Citrus clementina</i>	Phytozome	
<i>Citrus maxima</i>	Citrus Genome Database	
<i>Fraxinus excelsior</i>	NCBI Genome	

Table D-1 (continued)

Plant species	Depository	Literature References
<i>Citrus medica</i>	Citrus Genome Database	
<i>Citrus cavaleriei</i>	Citrus Genome Database	
<i>Citrus reticulata</i>		(Wang et al., 2018)
<i>Citrus sinensis</i>	Phytozome	
<i>Citrus unshiu</i>		(Shimizu et al., 2017)
<i>Cocos nucifera</i>		(Xiao et al., 2017)
<i>Coffea arabica</i>	NCBI Genome	
<i>Coffea canephora</i>	Phytozome	
<i>Corchorus capsularis</i>	Phytozome	
<i>Corchorus olitorius</i>		(Sarkar et al., 2017)
<i>Cucumis melo</i>	CuGenDB	
<i>Cucumis sativus</i>	Phytozome	
<i>Cucurbita pepo</i>	CuGenDB	
<i>Cynara cardunculus</i>		(Scaglione et al., 2016)
<i>Daucus carota</i>	Phytozome	
<i>Dendrobium officinale</i>		(Yan et al., 2015)
<i>Dianthus caryophyllus</i>		(Yagi et al., 2013)
<i>Dichanthelium oligosanthes</i>		(Studer et al., 2016)
<i>Dioscorea rotundata</i>	Phytozome	
<i>Durio zibethinus</i>		(Teh et al., 2017)
<i>Echinochloa crus-galli</i>		(Guo et al., 2017)
<i>Elaeis guineensis</i>		(Jin et al., 2016)
<i>Eragrostis tef</i>		(Cannarozzi et al., 2014)
<i>Erigeron breviscapus</i>		(Yang et al., 2017)
<i>Eschscholzia californica</i>		(Hori et al., 2017)
<i>Eucalyptus camaldulensis</i>	NCBI Genome	
<i>Eucalyptus grandis</i>	Phytozome	
<i>Eutrema salsugineum</i>	Brassica Database	
<i>Fagopyrum esculentum</i>	NCBI Genome	
<i>Fagopyrum tataricum</i>	NCBI Genome	
<i>Fagus sylvatica</i>	NCBI Genome	
<i>Faidherbia albida</i>		(Chang et al., 2018)

Table D-1 (continued)

Plant species	Depository	Literature References
<i>Fragaria vesca</i>	Phytozome	
<i>Genlisea aurea</i>	NCBI Genome	
<i>Ginkgo biloba</i>		(Guan et al., 2016)
<i>Glycine max</i>	Phytozome	
<i>Glycine soja</i>		(Qi et al., 2014)
<i>Glycyrrhiza uralensis</i>		(Mochida et al., 2017)
<i>Gnetum montanum</i>		(Wan et al., 2018)
<i>Gossypium arboreum</i>	NCBI Genome	
<i>Gossypium barbadense</i>	NCBI Genome	
<i>Gossypium hirsutum</i>	Phytozome	
<i>Gossypium raimondii</i>	Phytozome	
<i>Handroanthus impetiginosus</i>	NCBI Genome	
<i>Helianthus annuus</i>	Phytozome	
<i>Hevea brasiliensis</i>	NCBI Genome	
<i>Hordeum vulgare</i>	Phytozome	
<i>Humulus lupulus</i>	NCBI Genome	
<i>Ipomoea batatas</i>	NCBI Genome	
<i>Ipomoea nil</i>	NCBI Genome	
<i>Ipomoea trifida</i>	NCBI Genome	
<i>Jatropha curcas</i>	NCBI Genome	
<i>Juglans regia</i>	NCBI Genome	
<i>Kalanchoe fedtschenkoi</i>	Phytozome	
<i>Kalanchoe laxiflora</i>	Phytozome	
<i>Lablab purpureus</i>		(Chang et al., 2018)
<i>Lactuca sativa</i>	Phytozome	
<i>Lagenaria siceraria</i>	CuGenDB	
<i>Leavenworthia alabamica</i>	Brassica Database	
<i>Leersia perrieri</i>	Phytozome	
<i>Linum usitatissimum</i>	Phytozome	
<i>Liriodendron chinense</i>	NCBI Genome	

Table D-1 (continued)

<i>Lolium perenne</i>	NCBI Genome	
<i>Lotus japonicus</i>	Phytozome	
<i>Lupinus angustifolius</i>	Phytozome	
<i>Macleaya cordata</i>	NCBI Genome	
Plant species	Depository	Literature References
<i>Malania oleifera</i>		(Xu et al., 2019)
<i>Malus domestica</i>	Phytozome	
<i>Manihot esculenta</i>	Phytozome	
<i>Marchantia polymorpha</i>	Phytozome	
<i>Medicago truncatula</i>	Phytozome	
<i>Erythranthe guttata</i>	NCBI Genome	
<i>Miscanthus sinensis</i>	Phytozome	
<i>Momordica charantia</i>	NCBI Genome	
<i>Moringa oleifera</i>		(Chang et al., 2018)
<i>Morus notabilis</i>	NCBI Genome	
<i>Musa itinerans</i>	NCBI Genome	
<i>Musa acuminata</i>	Phytozome	
<i>Musa balbisiana</i>		(Davey et al., 2013)
<i>Musa schizocarpa</i>	NCBI Genome	
<i>Nelumbo nucifera</i>	NCBI Genome	
<i>Nicotiana attenuata</i>	Phytozome	
<i>Nicotiana benthamiana</i>	NCBI Genome	
<i>Nicotiana obtusifolia</i>	NCBI Genome	
<i>Nicotiana sylvestris</i>	NCBI Genome	
<i>Nicotiana tabacum</i>	NCBI Genome	
<i>Nicotiana tomentosiformis</i>	NCBI Genome	
<i>Ocimum tenuiflorum</i>	NCBI Genome	
<i>Olea europaea</i>	Phytozome	
<i>Oropetium thomaeum</i>	Phytozome	
<i>Oryza barthii</i>	Phytozome	
<i>Oryza brachyantha</i>	Phytozome	
<i>Oryza glaberrima</i>	Phytozome	
<i>Oryza longistaminata</i>	Phytozome	
<i>Oryza meridionalis</i>	Phytozome	
<i>Oryza sativa</i> f. <i>spontanea</i>	Phytozome	

Table D-1 (continued)

<i>Oryza meyeriana</i> var. <i>granulata</i>		(Wu et al., 2018)
Plant species	Depository	Literature References
<i>Oryza punctata</i>	Phytozome	
<i>Oryza rufipogon</i>	Phytozome	
<i>Oryza sativa</i>	Phytozome	
<i>Panax ginseng</i>		(Kim et al., 2018)
<i>Panicum hallii</i>	Phytozome	
<i>Panicum virgatum</i>	Phytozome	
<i>Parasponia andersonii</i>	NCBI Genome	
<i>Cenchrus americanus</i>	NCBI Genome	
<i>Petunia axillaris</i>		(Bombarely et al., 2016)
<i>Petunia integrifolia</i> subsp. <i>inflata</i>		(Bombarely et al., 2016)
<i>Phalaenopsis aphrodite</i>	NCBI Genome	
<i>Phalaenopsis equestris</i>	NCBI Genome	
<i>Phaseolus vulgaris</i>	Phytozome	
<i>Phoenix dactylifera</i>	NCBI Genome	
<i>Phyllostachys edulis</i>		(Fei et al., 2018)
<i>Physcomitrella patens</i>	Phytozome	
<i>Picea abies</i>	NCBI Genome	
<i>Pinus lambertiana</i>	NCBI Genome	
<i>Pinus taeda</i>	NCBI Genome	
<i>Pogostemon cablin</i>	NCBI Genome	
<i>Populus deltoides</i>	Phytozome	
<i>Populus euphratica</i>	NCBI Genome	
<i>Populus pruinosa</i>		(Wang et al., 2017)
<i>Populus tremula</i>		(Lin et al., 2018)
<i>Populus tremuloides</i>		(Lin et al., 2018)
<i>Populus trichocarpa</i>	Phytozome	
<i>Potentilla micrantha</i>		(Giongo et al., 2017)
<i>Primula veris</i>	NCBI Genome	
<i>Primula vulgaris</i>	NCBI Genome	
<i>Prunus avium</i>	NCBI Genome	
<i>Prunus mume</i>	NCBI Genome	

Table D-1 (continued)

<i>Prunus persica</i>	Phytozome	
<i>Pseudotsuga menziesii</i>	NCBI Genome	
Plant species	Depository	Literature References
<i>Punica granatum</i>	NCBI Genome	
<i>Pyrus x bretschneideri</i>	NCBI Genome	
<i>Pyrus communis</i>		(Chagné et al., 2014)
<i>Quercus lobata</i>	NCBI Genome	
<i>Quercus robur</i>	NCBI Genome	
<i>Quercus suber</i>	NCBI Genome	
<i>Raphanus raphanistrum</i>	NCBI Genome	
<i>Raphanus sativus</i>	NCBI Genome	
<i>Rhizophora apiculata</i>		http://evolution.sysu.edu.cn/Sequences.html
<i>Rhododendron delavayi</i>		(Yan et al., 2017)
<i>Ricinus communis</i>	Phytozome	
<i>Rosa chinensis</i>	NCBI Genome	
<i>Rosa multiflora</i>	NCBI Genome	
<i>Rubus occidentalis</i>		(Wai et al., 2018)
<i>Saccharum spontaneum</i>	NCBI Genome	
<i>Salix purpurea</i>	Phytozome	
<i>Salvia miltiorrhiza</i>		(Zhang et al., 2015)
<i>Salvia splendens</i>		(Dong et al., 2018)
<i>Salvinia cucullata</i>		(Li et al., 2018)
<i>Schrenkiella parvula</i>	Brassica Database	
<i>Sclerocarya birrea</i>		(Chang et al., 2018)
<i>Secale cereale</i>	NCBI Genome	
<i>Selaginella moellendorffii</i>	Phytozome	
<i>Sesamum indicum</i>	NCBI Genome	
<i>Setaria italica</i>	Phytozome	
<i>Setaria viridis</i>	Phytozome	
<i>Siraitia grosvenorii</i>		(Xia et al., 2018)
<i>Sisymbrium irio</i>	Brassica Database	
<i>Solanum lycopersicum</i>	Phytozome	
<i>Solanum melongena</i>	NCBI Genome	
<i>Solanum pennellii</i>	NCBI Genome	
<i>Solanum pimpinellifolium</i>	NCBI Genome	

Table D-1 (continued)

<i>Solanum tuberosum</i>	Phytozome	
<i>Sonneratia alba</i>		http://evolution.sysu.edu.cn/Sequences.html
Plant species	Depository	Literature References
<i>Sonneratia caseolaris</i>		http://evolution.sysu.edu.cn/Sequences.html
<i>Sorghum bicolor</i>	Phytozome	
<i>Sphagnum fallax</i>	Phytozome	
<i>Spinacia oleracea</i>	NCBI Genome	
<i>Spirodela polyrhiza</i>	Phytozome	
<i>Taraxacum kok-saghyz</i>		(Lin et al., 2017)
<i>Tarenaya hassleriana</i>	NCBI Genome	
<i>Tectona grandis</i>		(Zhao et al., 2019)
<i>Theobroma cacao</i>	Phytozome	
<i>Thinopyrum intermedium</i>	Phytozome	
<i>Thlaspi arvense</i>		(Dorn et al., 2015)
<i>Trema orientale</i>		(van Velzen et al., 2018)
<i>Trifolium pratense</i>	Phytozome	
<i>Trifolium subterraneum</i>		(Hirakawa et al., 2016)
<i>Triticum aestivum</i>	Phytozome	
<i>Triticum turgidum</i>		(Avni et al., 2017)
<i>Triticum urartu</i>	Phytozome	
<i>Utricularia gibba</i>		(Ibarra-Laclette et al., 2013)
<i>Vaccinium corymbosum</i>		(Gupta et al., 2015)
<i>Vigna angularis</i>		(Kang et al., 2015)
<i>Vigna radiata</i>		(Kang et al., 2014)
<i>Vigna subterranea</i>		(Chang et al., 2018)
<i>Vigna unguiculata</i>	Phytozome	
<i>Vitis vinifera</i>	Phytozome	
<i>Zea mays</i>	Phytozome	
<i>Zizania latifolia</i>		(Guo et al., 2015)
<i>Ziziphus jujuba</i>		(Liu et al., 2014)
<i>Zostera marina</i>	Phytozome	
<i>Zostera muelleri</i>		(Lee et al., 2016)
<i>Zoysia japonica</i>		(Tanaka et al., 2016)
<i>Zoysia matrella</i>		(Tanaka et al., 2016)
<i>Zoysia pacifica</i>		(Tanaka et al., 2016)

Table D-2. Information of characterized SABATH methyltransferases. Enzyme names and their corresponding literatures were listed respectively. NCBI accession number was also provided if available.

Enzyme Name	Accession number	Literature References
Salicylic acid methyltransferase (CbSAMT)	AF133053	(Ross, Nam et al. 1999)
Salicylic acid methyltransferase (AmSAMT)	AF515284	(Negre et al., 2002)
Salicylic acid methyltransferase (SfSAMT)	AJ308570	(Pott et al., 2002)
Salicylic acid methyltransferase (AbSAMT)	AB049752	(Fukami et al., 2002)
Salicylic acid methyltransferase (HcSAMT)	AJ863118	(Effmert et al., 2005)
Salicylic acid methyltransferase (DwSAMT)	EF472972	(Barkman et al., 2007)
Salicylic acid methyltransferase (SISAMT)	GU299532	(Tieman et al., 2010; Ament et al., 2010)
Salicylic acid methyltransferase (PtSAMT)	KC894591	(Zhao et al., 2013)
Salicylic acid methyltransferase (PaSAMT)	MA_3356g0010	Chaiprasongsuk et al., 2018
Salicylic acid methyltransferase (GmSAMT)	Glyma02g06070	(Lin et al., 2013)
Benzoic acid methyltransferase (PtBAMT)	KU758949	Han et al., 2018
Benzoic acid methyltransferase (AmBAMT)	AF198492	(Murfitt et al., 2000; Dudareva et al., 2000)
Benzoic acid:salicylic acid methyltransferase (AtBSMT)	BT022049	(Chen et al., 2003a)

Table D-2 (continued)

Enzyme Name	Accession number	Literature References
Benzoic acid:salicylic acid methyltransferase (AIBSMT)	AY224596	(Chen et al., 2003a)
Benzoic acid:salicylic acid methyltransferase (PhBSMT)	AY233465	(Negre et al., 2003)
Benzoic acid:salicylic acid methyltransferase (NsBSMT)	AJ628349	(Pott et al., 2004)
Benzoic acid:salicylic acid methyltransferase (OsBSMT)	Os02g48770	(Koo et al., 2007; Zhao et al., 2010)
Indole-3-acetic acid methyltransferase (PaIAMT)	MA_98251g0010	Chaiprasongsuk et al., 2018
Indole-3-acetic acid methyltransferase (AtIAMT1)	AK175586	(Qin et al., 2005)
Indole-3-acetic acid methyltransferase (PtIAMT1)	XP_002298843	(Zhao et al., 2007)
Indole-3-acetic acid methyltransferase (OsIAMT1)	EU375746	(Zhao et al., 2008)
Indole-3-acetic acid methyltransferase (SIIAMT1)		(Tieman et al., 2010)
Indole-3-acetic acid methyltransferase (PgIAMT1)		(Zhao et al., 2009)
Jasmonic acid methyltransferase (AtJAMT)	AY008434	(Seo et al., 2001)
Jasmonic acid methyltransferase (SIJAMT)		(Tieman et al., 2010)
Jasmonic acid methyltransferase (PtJAMT)		(Zhao et al., 2013)
Jasmonic acid methyltransferase (FvJAMT)		(Preuß et al. 2014)
Jasmonic acid methyltransferase (BjJAMT)	AY667499.1	(Meur et al., 2015)
Jasmonic acid methyltransferase (PaJAMT1)	MA_128083g0020	Chaiprasongsuk et al., 2018

Table D-2 (continued)

Enzyme Name	Accession number	Literature References
Jasmonic acid methyltransferase (PaJAMT2)	MA_128083g0020	Chaiprasongsuk et al., 2018
Jasmonic acid methyltransferase (PaJAMT3)	MA_9561g0010	Chaiprasongsuk et al., 2018
Jasmonic acid methyltransferase (OsJAMT1)	AK240953	Qi et al., 2016
Farnesoic acid methyltransferase (AtFAMT)	AY150400	(Yang et al., 2006)
Gibberellic acid methyltransferase (AtGAMT1)	At4g26420	(Varbanova et al., 2007)
Gibberellic acid methyltransferase (AtGAMT2)	At5g56300	(Varbanova et al., 2007)
Antranilic acid methyltransferase (ZmAAMT1)	ADI87449.1	(Kollner, Lenk et al. 2010)
Antranilic acid methyltransferase (ZmAAMT2)	ADI87451.1	(Kollner, Lenk et al. 2010)
Antranilic acid methyltransferase (ZmAAMT3)	ADI87452.1	(Kollner, Lenk et al. 2010)
PcCS1		Huang et al., 2016
PcCS2	BK008796.1	Huang et al., 2016
(CisXMT1)	lg044727	Huang et al., 2016
(CisXMT2)	lg047625	Huang et al., 2016
caffeine synthase 1 (CaCCS1)	AB086414	(Mizuno et al., 2003)
xanthosine methyltransferase (CaXMT1/CaCmXRS1)	AB048793	(Ogawa et al., 2001)
3,7-dimethylxanthine N-methyltransferase (CaDXMT1)	AB084125	(Uefuji et al., 2003)
7-methylxanthine N-methyltransferase (CaMXMT1/CTS1)	AB048794	(Ogawa et al., 2001)

Table D-2 (continued)

Enzyme Name	Accession number	Literature References
7-methylxanthine N-methyltransferase (CaMXMT2)	AB084126	(Uefuji et al., 2003)
theobromine synthase 2 (CaCTS2)	AB054841	(Mizuno et al., 2003)
caffeine synthase (CsTCS1)	AB031280	(Kato et al., 2000)
caffeine synthase (CsTCS2)	AB031281	(Kato et al., 2000)
caffeine synthase (TcBCS1)	AB096699	(Yoneyama et al., 2006)
TcCS1	EC766748	Huang et al., 2016
TcCS2	EC778019	Huang et al., 2016
theobromine synthase (CpPCS1)	AB207817	(Yoneyama et al., 2006)
theobromine synthase (CpPCS2)	AB207818	(Yoneyama et al., 2006)
theobromine synthase (CiICS1)	AB056108	(Yoneyama et al., 2006)
theobromine synthase (CiICS2)	AB207816	(Yoneyama et al., 2006)
loganic acid methyltransferase (CrLAMT)	EU057974	(Murata et al., 2008)
loganic acid methyltransferase (LojLAMT)	AMB61019	
norbixin carboxyl methyltransferase (BonBMT)	AJ548847	(Bouvier et al., 2003)
cinnamate/ <i>p</i> -coumarate carboxyl methyltransferase 1 (ObCCMT1)	EU033968	(Kapteyn et al., 2007)
cinnamate/ <i>p</i> -coumarate carboxyl methyltransferase 2 (ObCCMT2)	EU033969	(Kapteyn et al., 2007)
cinnamate/ <i>p</i> -coumarate carboxyl methyltransferase 3 (ObCCMT3)	EU033970	(Kapteyn et al., 2007)
<i>p</i> -methoxybenzoic acid carboxyl methyltransferase (EjMBMT)	BAV54103	(Koeduka et al., 2016)

Table D-2 (continued)

Enzyme Name	Accession number	Literature References
gw1.47.108.1 (PpSABATH1)	Pp3c3_23910V3.1	Zhao et al. 2012

Table D-3. SABATH gene family in *Brachypodium distachyon*.

Gene name and gene ID of each SABATH member from *B. distachyon* was listed.

Gene name	Brachy ID
<i>BdSABATH1</i>	<i>Bradi1g44620</i>
<i>BdSABATH2</i>	<i>Bradi1g43080</i>
<i>BdSABATH3</i>	<i>Bradi1g42760.1</i>
<i>BdSABATH4</i>	<i>Bradi2g47550</i>
<i>BdSABATH5</i>	<i>Bradi2g60180</i>
<i>BdSABATH6</i>	<i>Bradi2g60170.1</i>
<i>BdSABATH7</i>	<i>Bradi3g05550</i>
<i>BdSABATH8</i>	<i>Bradi3g57790</i>
<i>BdSABATH9</i>	<i>Bradi4g00890</i>
<i>BdSABATH10</i>	<i>Bradi4g08320.1</i>
<i>BdSABATH11</i>	<i>Bradi4g16077</i>
<i>BdSABATH12</i>	<i>Bradi4g16110</i>

Table D-4. Gene-specific primers used in this study.

Gene IDs	Primers	Sequence (5'→3')
GbGAMT1	Gb1-F	ATGGATTGCTCGACTAGCGTGGTGTCA
Gb_37441	Gb1-R	CTATTTTCTGACAGCAAACACAACAATAACGCCA
GbGAMT2	Gb2-F	ATGGATGGTCCTACCATAGCTGCATC
Gb_37750	Gb2-R	TTATTCCTCAGCAAATCTGGTCATTACATC
BrGAMT1	Br1-F	ATGGATTCTTCTCGGAGCCTC
Bra026440	Br1-R	TCAAACCCTAATCGCTGAGACAG
BrGAMT2	Br2-F	ATGGAGTCACCAAGTCTTCCAGTAAC
Bra028937	Br2-R	TCATTCAATTCGGATGGCAGA
BrGAMT3	Br3-F	ATGCAAGGCGGAGACGGTGACGTCAG
Bra010433	Br3-R	TTAATACCAAACCCGAATCGCTGAAAC

Table D-5. Activities of GAMTs from selected plant species against GA₃ and GA₄

Division	Species	Enzyme Name	GA ₄	GA ₃
Core eudicots	<i>Brassica rapa</i>	BrGAMT1	+	+
		BrGAMT2	+	+
		BrGAMT3	-	-
	<i>Vitis vinifera</i>	VvGAMT	+	-
	<i>Ziziphus jujuba</i>	ZjGAMT	-	-
	<i>Hevea brasiliensis</i>	HbGAMT	+	+
	<i>Quercus robur</i>	QrGAMT	+	+
	<i>Fagus sylvatica</i>	FsGAMT	+	-
	<i>Corchorus olitorius</i>	CoGAMT	+	+
	<i>Atalantia buxifolia</i>	AbGAMT	-	-
	<i>Citrus clementina</i>	CcGAMT	+	-
	<i>Sisymbrium irio</i>	SiGAMT1	+	+
		SiGAMT2	+	+
Basal eudicots	<i>Macleaya cordata</i>	McGAMT	+	-
Monocots	<i>Zostera marina</i>	ZmGAMT	-	-
	<i>Phalaenopsis</i>	PaGAMT	+	+
	<i>aphrodite</i>			
	<i>Phoenix dactylifera</i>	PdGAMT	+	+
	<i>Musa acuminata</i>	MaGAMT	+	-

Table D-5 (continued)

Division	Species	Enzyme Name	GA ₄	GA ₃
Basal Magnoliophyta	<i>Amborella trichopoda</i>	AtGAMT	+	+
Gymnosperms	<i>Ginkgo biloba</i>	GbGAMT1	+	+
		GbGAMT2	-	-
	<i>Gnetum montanum</i>	GmGAMT	+	-
	<i>Pseudotsuga menziesii</i>	PmGAMT	+	-
	<i>Pinus taeda</i>	PtGAMT	+	-

Chapter V. Conclusions and Perspectives

1. Conclusions

In the previous chapters, comparative genomic approaches were employed to identify SABATH genes from *C. salebrosum*, *C. iria* and GAMT genes from a large-scale collection of plant genomes. *C. salebrosum* and *C. iria* were used to study the biosynthesis of methyl cinnamate and that of JH III, respectively. The expression patterns of newly identified SABATH genes in multiple tissues were analyzed using a semiquantitative or quantitative RT-PCR approach and genes in full length were cloned. Two *C. salebrosum* genes, three *C. iria* genes and 21 GAMT genes from 17 species were identified with MT activities and biological functions of these genes were studied with respect to plant interactions with environment and homeostasis of GAs. The evolution of the functions of the SABATH gene family was investigated. Moreover, the cinnamic acid methylating enzyme in *Conocephalum salebrosum* was surveyed using phylogenetic and enzyme structure analysis to understand the relationship of CAMT genes between bryophytes and seed plants. This research resulted in the following conclusions.

First, nine SABATH genes were identified in the transcriptome of *Conocephalum salebrosum*. Sequence alignment showed that six of the nine genes appeared in full-length. Three of the six full-length CsSABATH genes showed relatively high expression in the thallus. Two genes including CsCAMT and CsSAMT displayed the highest level of catalytic activity towards cinnamic acid and SA, respectively. CsCAMT might have the defense function. As shown in chapter II, methyl cinnamate was detected as major volatile in *Conocephalum salebrosum* and CsCAMT was in high expression in thallus. Multiple studies have shown that methyl cinnamate has direct inhibiting effects on microbes and herbivores

(Davis and Landry, 2012; Fujiwara et al., 2017; Gilles et al., 2010; Padalia et al., 2017).

CsCAMT and CsSAMT evolved independently in liverworts from their counterparts in seed plants. Both phylogenetic analysis and structure modeling supported that the occurrence of CsCAMT and ObCCMT was a result of convergent evolution.

Second, seven SABATH genes were identified in the transcriptome of *Cyperus iria*. Sequence alignment showed that three of the seven genes appeared in full-length. All three full-length CiSABATH genes including CiFAMT, CiJAMT, and CiBSMT were identified to have the highest level of catalytic activity with farnesoic acid, JA and BA/SA, respectively. CiFAMT is the key enzyme that methylates farnesoic acid into methyl farnesoate, the direct precursor of JH III, thus involved in plant defense. qRT-PCR showed that CiFAMT expressed mainly in the root which is in accord with the high content of methyl farnesoate in the root. JH III is a vital hormone that maintains the normal life cycle of insects and can have a dramatic effect on insect development even in very low quantity. Rice flat sedge showed larvicidal effects on larvae laid by different species of insects (Meneses and Garcia de la Osa, 1988; Sccftwartz et al., 1998). CiFAMT evolved from unknown monocot SABATH. The discovery implied the independent evolution of JH III biosynthetic pathway in plants and insects.

Third, GAMT genes were identified from 75 species of land plant genomes. 24 representative GAMT candidates were selected and 21 of them showed the highest level of catalytic activity with GA3 or GA4. GAMT genes are conserved among different species and involved in seed development. A putative GAMT clade was inferred in a large-scale plant SABATH phylogenetic analysis. Enzymatic assay of selected proteins showed that they were

indeed GAMT. Gene expression pattern addressed that GAMTs mainly expressed through seed development. GAMT is lost in a large number of flowering plants. Although GAMT sequences are relatively conserved and they exist in all gymnosperms and basal angiosperms, their orthologs are clearly missing in many major plant orders including all that belong to Poales, Fabales, and superasterids.

Overall, in this dissertation, I used comparative functional genomics approach to identify the SABATH families in *Conocephalum salebrosum*, *Cyperus iria* and GAMTs from a large collection of sequenced plant genomes. The biochemical function and evolution of members of SABATH family were studied in order to increase the knowledge of evolutionary aspects of the SABATH family. However, further studies are required to performed to identify the biochemical function of a large amount uncharacterized SABATH proteins, the biological function of SABATH genes in plant development and interaction with the environment, as well as the evolution of SABATH family in the plant kingdom.

2. Perspectives

Despite the progress made, there are still many unsolved questions regarding the function and evolution of SABATH families in plants. Further efforts can be made from the following perspectives.

2.1. Biochemical functions of SABATH genes

Conocephalum salebrosum and *Cyperus iria* contain at least nine and seven SABATH genes, respectively. The biochemical functions of two CsSABATH and three CiSABATH

have been identified. However, the substrates for the remainder of the SABATH genes detected cannot be detected. Revealing the biochemical functions of SABATH proteins lays the foundation for understanding their biological roles. Gene expression analysis showed all identified genes expressed at least in one tissue suggesting that these genes have a role in the plant. It will be interesting to examine how they involve in the specific physiological process of plants. It will also be interesting to examine whether some of these genes can be induced under stress. Moreover, identification of the complete family of SABATH genes can provide a comprehensive understanding of the evolution of the SABATH family. Phylogenetic analysis suggested that there were several cases of convergent evolution in SABATH family including CAMT and FAMT. It will be interesting to test whether convergent evolution is widespread in other SABATH MTs.

2.2. Biological functions of SABATH genes

The biological functions of SABATH genes in plants need to be further addressed since there are still a lot of unanswered questions. First, the majority of SABATH genes in *Conocephalum salebrosum* and *Cyperus iria* are biochemically uncharacterized. Biological research through mutants and metabolic correlation analysis may provide a hint about the actual substrate of SABATH *in vivo*. Second, many methylated products are constantly made by plants (for example methyl cinnamate and methyl farnesoate), thus it will be interesting to study how they are regulated by SABATH and whether there are extra copies to maintain the production. Three, although SA is known to exist in liverwort, it is still interesting to verify whether CsSAMT has a similar function as other SAMT in seed plants (Ament et al., 2010).

Fourth, is CiFAMT playing a role in allelopathy? Previous studies have discussed that methyl farnesoate might be synthesized by *C. iria* for allelopathy since *C. iria* is an invasive plant and methyl farnesoate mainly accumulates in the root. Via gene expression pattern we know that GAMT is involved in seed development and transgenic plants have shown the phenotype of GA deficiency (Varbanova et al., 2007), but how GA methylation works with other regulating mechanisms including conjugation and the direct switch above GAMT are still unknown.

2.3. Evolution of SABATH genes

Further understanding of the evolution of SABATH genes will help to discover the substrates of uncharacterized SABATH and understand the role of SABATH in plant growth, development, and interactions with the environment. Several achievements have been made regarding SABATH family evolution in this dissertation. A large-scale phylogenetic tree has been constructed including 6458 SABATH sequences that cover both vascular plants and non-vascular plants. From the phylogenetic tree, several conserved clades were observed other than the GAMT clade described in Chapter IV. It is interesting to study these conserved clades since they might have similar biochemical and/or biological functions. There were also many taxonomically specific SABATH group which brought the question of whether they were corresponding to some specialized functions. However, some questions are still unclear. What's the relationship between SABATH in non-vascular plants and vascular plants? Are they functionally conserved or totally unrelated? Is there some functional divergence even in the conserved clade? How did IAMT and GAMT, two ancient groups, evolve? Is there any

enzyme with promiscuous activities from non-conserved clades? Is there any other MT holding the highest activity towards more than one substrate like BSMT? Why has JAMT evolved independently, on multiple occasions in seed plants? Is there any *SABATH* gene in a large biosynthesis cluster for specialized product biosynthesis? Can other phytohormones with carboxyl groups (for example ABA) also be regulated by *SABATH*? As plant genome sequencing is becoming more and more economically feasible, we will gain more information about the phylogeny about *SABATH* especially in non-seed plants which will help us answer some basic question about original evolution of *SABATH*. Biochemical and biological advances will certainly help with the evolution study since the more known parts of the phylogeny the better assumption and speculation we can make when facing the unknown parts of the phylogeny.

References

- Ament, K., Krasikov, V., Allmann, S., Rep, M., Takken, F. L., Schuurink, R. C., 2010. Methyl salicylate production in tomato affects biotic interactions. *The Plant Journal* 62, 124-134.
- Davis, D. R., Landry, J.-F., 2012. A review of the North American genus *Epimartyria* (Lepidoptera, Micropterigidae) with a discussion of the larval plastron. *ZooKeys*, 37.
- Fujiwara, G. M., Annies, V., de Oliveira, C. F., Lara, R. A., Gabriel, M. M., Betim, F. C., Nadal, J. M., Farago, P. V., Dias, J. F., Miguel, O. G., 2017. Evaluation of larvicidal activity and ecotoxicity of linalool, methyl cinnamate and methyl cinnamate/linalool in combination against *Aedes aegypti*. *Ecotoxicology and environmental safety* 139, 238-244.
- Gilles, M., Zhao, J., An, M., Agboola, S., 2010. Chemical composition and antimicrobial properties of essential oils of three Australian Eucalyptus species. *Food Chemistry* 119, 731-737.
- Meneses, C., Garcia de la Osa, J., 1988. Principal weed hosts of *Hydrellia* sp. in the southern rice-growing zone of Sancti Spiritus. *Centro Agrícola* 15, 90-92.
- Padalia, R. C., Verma, R. S., Chauhan, A., Goswami, P., Singh, V. R., Verma, S. K., Darokar, M. P., Singh, N., Saikia, D., Chanotiya, C. S., 2017. Essential Oil Composition and Antimicrobial Activity of Methyl cinnamate-Linalool Chemovariant of *Ocimum basilicum* L. from Indi Rajendra Chandra Padalia, Ram Swaroop Verma, Amit Chauhan, Prakash Goswami, Ved Ram Singh, Sajendra Kumar Verma, Mahendra Pandurang Darokar, Alka

kurmi, Nandan Singh, Dharmendra Saikia and Chandan Singh Chanotiya. Records of Natural Products 11, 193.

Sccftwartz, A., Paskewitz, S. M., Orth, A. P., Tesch, M. J., Toong, I., Goodman, W. G., 1998. The lethal effects of *Cyperus iria* on *Aedes aegypti*. J Am Mosq Contr Assoc 14, 78-82.

Varbanova, M., Yamaguchi, S., Yang, Y., McKelvey, K., Hanada, A., Borochoy, R., Yu, F., Jikumaru, Y., Ross, J., Cortes, D., 2007. Methylation of gibberellins by *Arabidopsis* GAMT1 and GAMT2. The Plant Cell 19, 32-45.

Vita

Chi Zhang was born in Dalian, China on September 30, 1989. After completing his early education in Dalian, China, he entered the Sun Yat-sen University, Guangzhou, China in 2008, where he obtained his B.S. degree in 2012. He graduated M.S. degree in molecular ecology from Institute of Applied Ecology, Chinese Academy of Sciences in 2015. Soon after, he moved to Knoxville, Tennessee to join Dr. Feng Chen's lab as a Ph.D. student at University of Tennessee in August 2015. During the next four years, he studied biochemistry and evolution of SABATH methyltransferase family in plants.