An Investigation of Flavanone *O*-and *C*-methylation in *Eucalyptus*

Krishna Somaletha Chandran
ORCID: 0000-0001-9084-0617

Doctor of Philosophy

MAY 2021

Faculty of Science
School of Biosciences

Submitted in total fulfilment for the degree of Doctor of Philosophy

THE UNIVERSITY OF MELBOURNE

ABSTRACT

Flavonoid compounds are well recognized for their diverse health-promoting properties in humans. Methylation of flavonoids catalysed by plant methyltransferases is an important modification technique that alters the biochemical properties of the compounds, thereby enabling extension of their promising medicinal effects from *in vitro* to *in vivo*. Some *Eucalyptus* trees accumulate high levels of *O*-and *C*-methylated flavanones in their leaves. These compounds have potential medicinal and antimicrobial use. Despite the potential use of *Eucalyptus* species as a commercial source of these flavonoids, little is known about the underlying mechanism of *in planta* methylation and biosynthesis of these compounds. Through the isolation and characterization of flavanone *C*-and *O*- methyltransferases from *Eucalyptus*, this research aims to provide a detailed insight into the underlying mechanism of different types of *in planta* methylation and the process by which these compounds are modified.

An integrated analysis of transcriptome and metabolome coupled with in silico methods resulted in a cDNA clone (EnOMT1), catalysing the position-specific *O*-methylation of a single hydroxyl group in the flavanone pinocembrin. The biochemical characterization showed that EnOMT1 is a regiospecific methyltransferase with strict positional specificity to the 7-hydroxyl of flavonoids. The reduction/absence of activity of EnOMT1 with B-ring substituted flavonoids such as luteolin, apigenin, and quercetin demonstrates the pronounced effect B-ring hydroxyl configuration has on the activity of the enzyme.

Several *Eucalyptus* species belonging to subgenus *Eucalyptus* (monocalypts) show accumulation of different complements of methylated flavonoids. To further understand the genetic basis of this differential accumulation, three homologous genes of EnOMT1 were isolated and characterised from various species. A homologue (ExOMT1) from a species that does not accumulate any methylated flavanones showed 94% identity to EnOMT1; however, the homologue showed no methylation activity with flavanones *in vitro*. Gene expression analysis using qPCR showed a comparably lower expression of ExOMT1 to that of EnOMT1.Therefore, the observed difference in activity is likely due to the amino acid difference of the EnOMT1 sequence to ExOMT1. Homology modelling along with site-directed mutagenesis of EnOMT1 further identified residues important for activity. The residues ― namely Trp252, Val253, and Asn256 ― play critical roles for EnOMT1 activity. SAM binding residues — namely, Trp148, Phe161, Met165, Asn169, Gln194, Phe216, Asp217, Arg218, Asp237, Met238, Phe239, Lyts251, Trp252, Val253, and Trp257 — were

also predicted with a higher degree of confidence

Genetic basis of flavonoid *C*-methylation is an understudied area and the limited published information on plant CMT sequences makes the research progress difficult. The work presented in chapter 4 provides interesting insights into flavanone *C*-methylation in *Eucalyptus*. Extensive flavanone profiling of ten species from subgenus *Eucalyptus* was conducted to identify a *C*-methyl flavanone rich species. Two known *C*-methyl flavanones, cryptostrobin (monomethylated at position C6) and desmethoxymatteucinol (dimethylated at C6 and C8), were identified in *E diversifolia* in relative abundance. The sequenced transcriptome of *E diversifolia* was analysed for the identification of secondary metabolite pathway genes putatively involved in *C*-methylation. Six putative CMT genes were cloned successfully and analysed for activity. Despite our attempts, no *in vitro* *C*-methylation was achieved on the flavanone substrates tested. The data presented in chapter 4 provide a solid platform of knowledge for future research on flavanone *C*-methylation.

The results presented in this thesis represent an important step towards obtaining a greater understanding of flavonoid biosynthesis in *Eucalyptus* and provide clear insights into the reaction mechanism as well as the active site residues of the enzyme.

DECLARATION

I hereby certify that this thesis comprises my original work towards the degree of Doctor of Philosophy. Due acknowledgement has been made in the text to all other material used. The thesis is fewer than 100,000 words in length exclusive of tables, bibliographies and appendices.

Krishna S Chandran

Melbourne, May 2021

PREFACE

The chapters included in this thesis contain original, unpublished material not submitted for publication. I am the principal contributor to, and primary author of, all chapters presented in this thesis. My supervisors have been recognised for their invaluable contribution to all chapters and all other contributors are mentioned in the acknowledgements.

The mass spectrometry was carried out with the assistance of Metabolomics Australia - the University of Melbourne node. This research was supported by funding from Australian Research Council Linkage Projects and Gretals Australia Pty. Ltd. Further financial support was provided by the Melbourne Research Scholarship (Stipend and Fee offset) for my PhD study at the University of Melbourne and student scholarships from the Botany foundation (The Sophie Ducker Scholarship) and Science awards (Norma Hilda Schuster (nee Swift) Scholarship -2019; Dawson Bursary 2019, 2020).

ACKNOWLEDGEMENTS

This research would never have been possible without the support of many people. I am delighted to express my sincere gratitude to all the wonderful individuals who constantly supported and guided me throughout this research.

Firstly, my sincere thanks to my supervisors Dr. Jason Goodger, Prof. Ian Woodrow for their continuous guidance, insights, and trust during the last four years. Their research enthusiasm and positive attitude towards everything was most inspirational and gave me a sense of the researcher I want to be in the future. To my co-supervisor Dr. John Humphries— words cannot express the gratitude I have towards you for always being the support pillar of my professional life.

The financial support from the Melbourne Research Scholarship for my PhD study at the University of Melbourne and the scholarships from Botany foundation and Science awards is greatly acknowledged. A special thanks to my committee chair Prof. Staffan Persson for his supportive and instructive feedback. I would also like to acknowledge my committee members A/Prof Alex Johnson and A/Prof Joshua Hazlewood for their willingness to share their knowledge. Thanks for all the patient advice and guidance throughout the research process. On a personal note, I would like to thank Dr. Shyama Fernando for her technical and non-technical advice during my PhD journey and also for the sincere care. Many thanks to the entire plant physiology group for creating a healthy research environment. A big, heartfelt gratitude to all my fellow PhD students, particularly Miss Neeta Lohani, Dr. Chinchu Mohan, and Dr. Ayana Bhaskaran for their constructive criticisms. I also acknowledge Dr. Melissa Bain for being my proofreading angel.

In truth, I could not have achieved my current level of success without a strong support group. I owe my greatest gratitude to Acha, Amma and Kunjava. I extend my sincere appreciation to my friends, particularly Arjun Radhakrishnan, Vipin Prem, Martino, Ghas, Sneha and Matt for making my time in Australia memorable. Thanks to Athira, Neethu, Sree, Yadu, Linju, Sujai, Kichu, Avinash, and Tanvi for all their emotional support. I am grateful to Mr Mithul Mohan for helping me with the tedious job of formatting and proof-reading.

Finally, I would like to thank all the forerunners in science and technology who made this research possible today.

TABLE OF CONTENTS

CHAPTER 3

CHARACTERIZATION OF A FLAVANONE 7- *O*-METHYLTRANSFERASE FROM *EUCALYPTUS*......89

CHAPTER 4

LIST OF FIGURES

LIST OF TABLES

LIST OF ABREVIATIONS

5-HF	5-hydroxy ferulic acid
AN	Alpinetin
AtOMT1	*Arabidopsis thaliana O*-methyltransferase1
ATP	Adenosine triphosphate
BAMT	Benzoic acid methyltransferase
BIA	Benzylisoquinoline alkaloid
BMT	Bergaptol *O*-methyltransferase
CCoAOMT	Caffeoyl-CoA OMT
ChOMT	Chalcone OMT
CMT	*C*-methyltransferase
CNMT	Coclaurine *N*-methyltransferase
CoA	Coenzyme A
COMT	Caffeic acid OMT
CS	Cryptostrobin
CVOMT	Chavicol OMT
DMM	Desmethoxymatteucinol
DMPC	Dimethylpinocembrin
DW	Dry weight
ESI	Electro-spray ionisation
GC-FID	Gas chromatography-flame ionisation detection
GO	Gene Ontology
HCA	Hydroxycinnamic acid
HI4′OMT	2,7,4′-trihydroxyisoflavanone *4′-O*-methyltransferase
HMM	(+)-6a-hydroxymaackiain 3-OMT
HPLC	High performance liquid chromatography
IAMT	indole acetic acid methyltransferase
IEMT	isoeugenol OMT
IOMT	Isoflavone OMT
KEGG	Kyoto Encyclopedia of Genes and Genome
LAMT	Loganic acid OMT

LB	Luria-Bertani
LC-MS	Liquid chromatography- Mass spectrometry
m/z	Mass to charge ratio
MIA	Monoterpene indole alkaloids
MTs	Methyltransferases
NMT	*N*-methyltransferase
OMT	*O*-methyltransferase
OOMT	Orcinol *O*-methyltransferase
PavNMT	Pavine NMT
PC	Pinocembrin
PDB ID	Protein data bank identifier
PFOMT	Phenylpropanoid flavonoid OMT
PFOMT	Phenylpropanoid/flavonoid OMT
PS	Pinostrobin
PTFE	Polytetrafluoroethylene
PVP	Polyvinylpyrrolidone
R7OMT	Reticuline 7-*O*-methyltransferase
RP-HPLC	Reverse phase high pressure liquid chromatography
RT	Retention time
SAH	*S*-adenosyl-L-homocysteine
SAM	*S*-adenosyl-L-methionine
SAMT	Salicylic acid methyltransferase
SER	Surface entropy reduction
SMT	*S*-methyltransferase
THI	2,7,4′- trihydroxyisoflavanone
Trik	Triketone
UV-A	Ultraviolet-A
XNMT	Xanthosine NMT
CDS	Coding sequence

CHAPTER 1

PLANT METABOLITE DIVERSIFICATION FACILITATED BY S-ADENOSYL-L-METHIONINE DEPENDENT METHYLTRANSFERASES: A FUNCTIONAL AND STRUCTURAL PERSPECTIVE

1.1 Introduction

Plants accumulate a vast number of metabolites which are often diversified by various modification mechanisms like glycosylation, methylation, acylation and prenylation. Methylation is catalysed by methyltransferases (MTs) which to date have been shown to exclusively utilise *S*-adenosyl-L-methionine (SAM) as a cofactor. Generally, these enzymes are named according to the nucleophile in the plant metabolite backbone they act on such as *O*-methyltransferase *N*-methyltransferase, *C*-methyltransferase and *S*-methyltransferase, attacking oxygen, nitrogen, carbon, and sulphur atoms, respectively. The term plant *S*-adenosyl-L-methionine dependent MTs used in this chapter denotes all of the SAM dependant MTs acting on plant primary and secondary metabolites.

The functional divergence of SAM-dependent MTs is attributed to the capability of their flexible substrate-binding core to accepting a vast number of structurally dissimilar compounds. In animals, they take part in biochemical reactions involved in the activation, catabolism, and metabolism of many biological compounds (namely, neurotransmitters, nucleic acids, hormones, phospholipids and proteins; Friedel et al. 1989), and they play a crucial role in controlling gene expression, cellular signalling, and metabolite degradation (Lu, 2000). In plants, these enzymes contribute to the biosynthesis of a wide repertoire of plant metabolites with diverse chemical scaffolds ranging from macromolecular DNA, RNA and proteins to small molecule metabolites. MTs take part in both primary and secondary metabolism, including cell signalling, epigenetic regulation, post-translational control of protein function, defence compound metabolism, membrane component and pigment formation. However, this review will focus in detail on the plant MTs that act on some of the major shikimate derived phenolic metabolites namely, phenylpropanoids, flavonoids, and alkaloids. A consolidated list of fully characterised MTs catalysing these substrates is given in Appendix A. Hereafter, the term plant MTs will denote the MTs working on these three classes of plant metabolites.

The first section of this chapter examines the broader structural classes of SAM-dependent MTs and their general catalytic mechanism. The sections that follow review the bioactivities of fully characterised plant MTs and their classification. Focusing on structurally characterised plant MTs to date, structural features that determine their function are discussed in the final sections of this review. Together, this chapter provides a comprehensive knowledge on contributions of plant specific MTs towards ornamentation of plant secondary metabolites.

1.2 Overview of SAM dependant methylation

SAM is a structural hybrid of adenosine and methionine (**Figure 1.1**) which acts as a universal methyl donor. It is the second most common small molecule cofactor after ATP (Schubert et al., 2003). Even though SAM acts as a universal methyl donor, its methyl group is in fact derived from *N*-methyltetrahydrofolate, mediated by a cobalamin-independent methionine synthase (Dixon et al., 1996). SAM-dependent MTs catalyse the methyl group transfer from SAM to an acceptor compound resulting in its methylated derivative and the by-product, *S*-adenosyl-L-homocysteine (SAH).

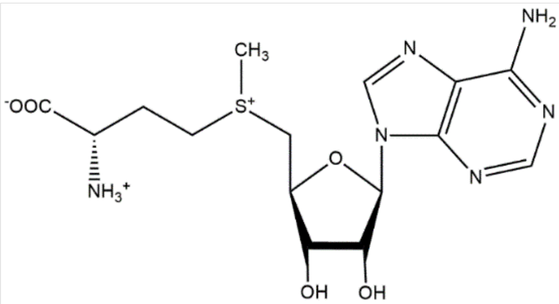


Figure 1.1: Structure of *S*-adenosyl-L- methionine.

All MTs characterised to date progress through a direct transfer of a methyl group to a substrate in an S_N2 reaction without the formation of a methylated enzyme intermediate (Bauerle et al., 2015; Schubert et al., 2003; Woodard et al., 1980). This necessitates the activation of a catalytic nucleophile on the substrate molecule by deprotonation which is achieved by various mechanisms depending on the chemistry of the target atom (Klimasauskas and Lukinavicius, 2007). The SAM binding region of MTs is a characteristic α/β/α Rossmann fold, which consists of at least three β-sheets and at least two α-helices in alternating order (Rao and Rossmann, 1973). Although there is poor conservation of SAM-binding residues, the overall fold is fairly maintained, and the enzyme binds SAM at equivalent positions of the fold. The glycine-rich

E/DXGXGXG (Motif I) and the acidic group between the second β-sheet and adjacent α-helix (Motif II) at the N-terminal region of the core fold are the only highly conserved residues (Martin and McMillan, 2002).

SAM-dependent MTs are classified into five classes (I-V) based on their tertiary structure (**Figure 1.2**). Class I is the largest group and includes all DNA MTs, some protein MTs, and *O*-, *N*-, *C*-, *S*-, *As*-, and *Se*-centred nucleophiles of a wide range of plant metabolites (Struck et al., 2012). All plant MTs fall under class I SAM MTs. The structure of class I members is characterised by seven stranded β-sheets flanked by an α-helix resulting in a typical Rossmann fold (Cheng et al., 1993). Even with only 10% primary amino acid similarity, class I members show remarkable tertiary structure consistency (Martin and McMillan, 2002). Most act as monomers, but some variants form homodimers (Vidgren et al., 1994) or tetramers (Huang et al., 2000).

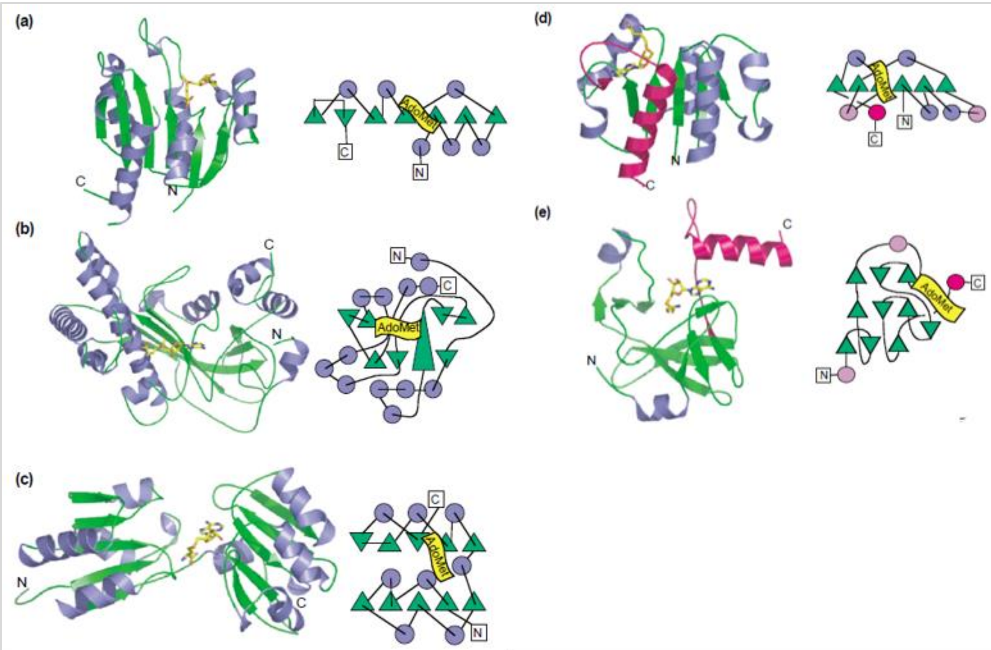


Figure 1.2: Tertiary structure of five SAM-dependent MTs. a) Class I, b) Class II, c)Class III, d)Class IV, e) Class V. Reprinted from Schubert et al. (2003) with permission from Elsevier. In each case a representative topology diagram (right) is shown where grey circles represent α-helices and green triangles represent β-sheets. Magenta circles denote the novel knot structure at the C terminus in class IV family. Pale pink circles represent helices that are not conserved between family members.

Class II MTs have a central antiparallel long β sheet flanked by helices (**Figure 1.2**b). In contrast to class I MTs, SAM forms hydrogen bonds with the conserved RXXGY motif and binds to a shallow groove along the edges of β strands (Jarrett et al., 1998). The SAM recognition site, topology, as well as SAM interaction, is very different from that of class I MTs (Dixon et al., 1996). Class III MTs have two α/β/α-domains each with four helices and five strands, an active site and a SAM binding site in the cleft between the two domains (**Figure 1.2**c; Schubert et al., 1998). Class IV SAM-dependant MTs are characterised by the presence of an unusual C-terminal knot and lacks the classic Rossmann fold features. The active site is near the homodimer interface and the last thirty C-terminal residues fold back and form a knot (**Figure 1.2**d). SAM binds in this knot region (Michel et al., 2002). Class V has eight β- strands which form three small sheets (**Figure 1.2**e), and it has a C- terminal knot like class IV, but it otherwise shows considerable differences in topology compared to class IV (Jacobs et al., 2002). Other than the five major classes, there are also some recently discovered lesser-known MTs. These include the radical SAM superfamily of MTs that works through a highly reactive radical generated via cleavage of SAM (Nicolet, 2020; Zhang et al., 2012), and methyl halide transferases (in plants and microorganisms) which catalyses the methylation of chloride, bromide and iodide ions (Bayer et al., 2009). Having summarised the general characteristics of SAM MTs, I will now move on to describe in greater detail the plant MTs that catalyse phenylpropanoids, flavonoids, and alkaloids.

1.3 Ornamentation of shikimate-derived phenolic metabolites by plant methyltransferases

The past decade has witnessed a rapid increase in methyltransferase research and a surge in the biological database with sequences of enzymes catalysing the methylation of plant phenolic compounds. Numerous transcripts have been detected to code for enzymes with high homology to plant MTs, some of which have been cloned but showed no detectable activity *in vitro* (Lang et al., 2019; Liscombe and Facchini, 2007; Morris and Facchini, 2019). For a better understanding, the MTs are grouped according to the biosynthetic pathway they are involved in (**Figure 1.3**). A consolidated list of plant MTs that have been fully characterised is shown in Appendix A. This section will focus on the contribution of these MTs towards the diversification and ornamentation of plant secondary metabolites as well as their biosynthesis.

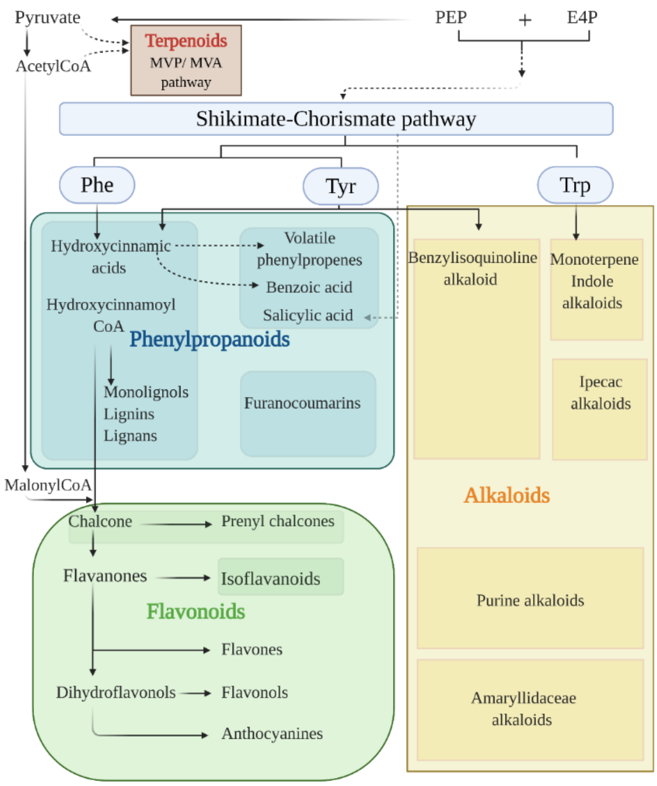


Figure 1.3: Schematic showing the overall biosynthetic outline (minimal) of various compounds listed in this review.

1.3.1 Phenylpropanoid methylation: *O*- and carboxyl- methyltransferases

i. Hydroxycinnamic acid and its CoA esters

Phenylpropanoid methylation is a widely researched area because of its involvement in the biosynthesis of lignin in vascular plants. Such research largely relates to timber production, and crop lodging to a lesser degree. Phenylpropanoids that have a C6-C3 skeleton are collectively referred to as hydroxycinnamic acid (HCA). Two major HCA compounds discussed in this section, caffeic acid and 5-hydroxy ferulic acid (5-HF), are the precursor compounds of monolignols coniferyl alcohol and sinapyl alcohol, respectively (Zhong et al., 1998). *In vitro* methylation studies have shown that HCA-OMTs have a broader substrate acceptance range where they catalyse *meta*-specific (*3*, *5-OH*) methylation of caffeoyl- and 5-HF derivatives containing aldehydes, acids, and alcohols. HCA-OMTs show kinetic preference for 5-HF derivatives over caffeoyl derivatives, with a two-fold increase in catalytic efficiency ($kcat/K_m$), while alcohols and aldehyde derivatives are kinetically preferred over the free acids (Zubieta et al., 2002). Regardless of the nature of the HCA molecule the enzyme accepts, they are generally termed as COMT (for **C**affeic acid **OMT**, not to be confused with catechol OMT). Seven conserved amino acid motifs have been proposed based on the analysis of fifty-six plant MTs (Joshi and Chiang, 1998). Among these, motif A, B and C are conserved in all plant MTs and are thought to be involved in SAM binding. The remaining (motif I, J, K and L) are specific for HCA-OMTs; however, whether these four motifs have any importance in HCA substrate utilization is not known. All characterised HCA-OMTs analysed in this study contain these seven motifs (Joshi and Chiang, 1998).

The other set of enzymes actively studied in the methylation research field are the ones catalysing CoA-esters of HCA compounds, termed as CCoAOMTs (for **C**affeoyl-**CoA OMT**). Unlike HCA-OMTs, CCoAOMTs are smaller, divalent cation dependant (Ferrer et al., 2005), and do not show the presence of all seven conserved motifs (Joshi and Chiang, 1998). They belong to a distinct family of plant MTs with higher sequence identity to mammalian catechol OMTs and bacterial MTs (Ferrer et al., 2005). These enzymes have a high specificity towards CoA-derivatives – i.e., acting only on CoA esters of caffeic acid and 5-HF with a preference for caffeoyl-CoA. Conversely, several variants in this family methylate the *meta*-position of caffeic acid esters and flavonoids (Ibdah et al., 2003; Widiez et al., 2011). An unusual *para*-methylation of flavanones and dihydroflavonols has also been identified in a characterised CCoAOMT-like enzyme from *A. thaliana* (Wils et al., 2013).

ii. Volatile phenylpropenes

Monolignols also serve as intermediates in the synthesis of volatile phenylpropanoids, the main constituents of floral scent, herb or fruit flavour, and OMTs that act on these create some of the diversity of plant scents found in nature (Atkinson, 2018; Dudareva and Pichersky, 2000). Several OMTs with activity toward phenylpropenes, have been characterized in various species including *Clarkia breweri* (Wang and Pichersky, 1999), *Antirrhinum majus* (Murfitt et al., 2000), *Ocimum basilicum* (Gang et al., 2002), *Rosa chinensis* (Wu et al., 2003), *Vanilla planifolia* (Pak et al., 2004), *Pimpinella anisum* (Koeduka et al., 2009), *Malus domestica* apple (Yauk et al., 2015) and *Dacus carota* (Yahyaa et al., 2019), and the methyl acceptors include eugenol, isoeugenol, chavicol, isochavicol (*t*-anol), orcinol, phloroglucinol. Methyl esters of salicylic acid, benzoic acid and jasmonic acid are also major constituents of floral scent volatiles.

The data accumulated to date suggest that substrate discrimination by phenylpropene OMTs is based on the double-bond position in the propenyl side chain and/or the functional groups on the benzene rings (Koeduka, 2014). *Clarkia breweri* isoeugenol OMT (IEMT) accepts both eugenol and isoeugenol as substrates but cannot use chavicol and isochavicol which lack a methoxy group on the benzene ring, whereas chavicolOMT (CVOMT) and eugenol OMT (EOMT1) from *O. basilicum* (basil) is highly specific for chavicol and eugenol, respectively. In contrast, isochavicol/isoeugenol *O*-methyltransferase (PaAIMT) from *P. anisum* is the only enzyme reported to show a differential preference based on double-bond position in the propenyl side chain (Koeduka et al., 2009). PaAIMT prefers propenyl phenylpropenes (isochavicol and isoeugenol) over allyl phenylpropenes (chavicol and eugenol).

In contrast, OMTs catalysing the formation of methyl esters of salicylic acid, benzoic acid and jasmonic acid transfer the methyl group to the respective carboxyl group instead of a hydroxyl. Since these OMTs are distinct in terms of their sequence and structure, they are considered a separate family that also includes cinnamic acid methyltransferase, alkaloid indole acetic acid OMT and purine alkaloid NMTs from coffee and tea (D Auria et al., 2003; Yoneyama et al., 2006). Almost all characterised carboxyl OMTs accept both benzoic acid as well as salicylic acid; but the exception is benzoic acid methyltransferase from snapdragon *A. majus* which shows a strict specificity (Negre et al., 2002). Carboxyl OMTs are named according to their substrate affinity as salicylic acid methyltransferase (SAMT) and benzoic acid methyltransferase (BAMT). So, in case of equal catalytic efficiency for both salicylic and benzoic acid, the enzymes were called BSMTs (Pott et al., 2004). An orthologue of SAMT from *A. thaliana* was found to methylate the carboxyl group on the plant hormone jasmonic acid (Seo et al., 2001).

Evolutionary analysis of the volatile phenylpropene MTs provide an interesting perspective on the distinct strategies taken by plants for aroma/flavour diversification. The mechanisms include gene duplication followed by divergence, repeated evolution, and convergent evolution. For example, the oricinol OMTs from *Rosa hybrida*, OOMT1 and OOMT2, show gene duplication and divergence where OOMT1 retains the original function of catalysing the earlier steps in dimethoxytoluene and trimethoxybenzene synthesis and OOMT2 apparently evolved to perform a new function which is the later steps in biosynthesis (Gang, 2005; Lavid et al., 2002). Another prominent example of gene duplication followed by divergence is the eugenol OMT and chavicol OMT from *O. basilicum* whose substrate preference is dependent only on a single residue (Gang et al., 2002). The products of these enzymes methyleugenol and methylchavicol, synthesised via parallel pathways originating from phenylalanine, account for two different chemotypes in basil, and site-directed mutagenesis of a single residue (F260S in CVOMT and S261F in EOMT1 which in fact is a single nucleotide C→ T conversion) was shown to swap their respective substrate preferences (Gang et al., 2002). Similarly, through site-directed mutagenesis and phylogeny analysis, Wang and Pichersky (1999) demonstrated that substrate discrimination by IEMT and COMT from *C. breweri* is dependant only on seven amino acid differences. Thus, they suggested that IEMT evolved recently from COMT within the order *Myrtales* via duplication.

Repeated evolution is a special case of convergent evolution where a new enzyme with similar or identical function independently evolves from orthologous or paralogous genes with in the same family (Pichersky and Gang, 2000). Three enzymes *C. breweri* IEMT, *O. basilicum* EOMT1, and *R. chinensis* RcOMT1 catalysing para *O*-methylation of eugenol and isoeugenol are good examples of repeated evolution. The three plant species are distantly related and these OMTs are likely to have evolved from different progenitors. *C. breweri* IEMT clearly evolved from *C. breweri* COMT (Wang and Pichersky, 1999), EOMT1 sequence is distinct from COMT and RcOMT1 is closely related to an alkaloid OMT yet they are all specific for eugenol or isoeugenol (Gang, 2005). Lastly, phylogenetic analysis of eugenol/isoeugenol MTs from *D. carota* and *P. anisum* both showed they are not related to either *C. breweri* IEMT or *O. basilicum* EOMT, suggesting their independent origin through convergent evolution (Yahyaa et al., 2019). It is interesting to note that some characterised phenylpropene OMTs, like RcOMT2 from *R chinensis* and a multifunctional OMT isolated from *V. planifolia*, have high sequence similarity to COMTs and prefer caffeic acid over phenylpropenes. Therefore, it has been suggested that these enzymes are likely to be COMTs with expanded catalytic capability (Gang, 2005; Pak et al., 2004; Wu et al., 2003).

iii. <u>Furanocoumarins</u>

Bergaptol 5-*O*-methyltransferase (BMT) is the only characterised methyltransferase involved in furanocoumarin biosynthesis. Characterisation of the first BMT gene isolated from elicitor-treated cells of *A. majus* showed that the enzyme is highly specific for bergaptol and does not accept its 8-hydroxylated isomer xanthotoxol (Hehmann et al., 2004). The presence of a distinct xanthotoxol 8-OMT has been reported (Hauffe et al., 1986), but it has not yet been fully characterised. A recent study elucidated the crystal structure of BMT from *P. praeruptorum* (Zhao et al., 2020), and it showed that the overall structure resembles *M. sativa* caffeic acid

OMT with a relatively spacious active site and a substrate shape selectivity that is due to the rigid α-helices in the active site pocket.

1.3.2 Flavonoid methylation: *O*- and *C*- methyltransferases

In general, flavonoids are products of the phenylpropanoid pathway. The branch point from lignin synthesis is via the formation of a chalcone, which is debatably considered a flavonoid. The flavonoid skeleton has three rings that are conventionally named A, B, and C (**Figure 1.4**A). Several *O*-methylated and *C*-methylated flavonoids compounds have been identified so far. Reviews that focus on occurrence and bioactivities of several methylated flavonoids have been published by Koirala et al. (2016) and Wen et al. (2017). In 2010, Kim et al. (2010) reviewed thirty biochemically characterised plant flavonoid OMTs, focusing on aspects such as regioselectivity, mechanism, and biological application. Over the past decade, a vast number of MTs acting on flavonoids have been isolated and characterised (Appendix A).

A clear distinction cannot be drawn between OMTs acting on compounds in the two classes of phenylpropanoids and flavonoids. As mentioned in section **1.3.1. i**, several variants of COMT and CCoAOMT accept flavonoid compounds (Kopycki et al., 2008; Lee et al., 2008; Lücker et al., 2010). For instance, OMT isolated from *C. americanum* (U16793) catalyses methylation of the *3′-OH* of quercetin as well as caffeic acid. *A. thaliana* AtOMT1 has an 80% sequence identity with caffeic acid OMT, but was found to preferably catalyse methylation of flavonol *3′-OH* (Muzac et al., 2000). Similarly, *O. sativa* ROMT-9 with a 73% identity to maize COMT preferably catalyses methylation of the *3′-OH* of eriodictyol, luteolin, quercetin and dihydroquercetin (Kim et al., 2006) Surprisingly, all the variants of COMT and CCoAOMT (accepting both phenylpropanoids and flavonoids) act on the *meta*-hydroxyl (*3′- OH*) on the flavonoid B-ring which is derived from caffeic acid.

i. <u>Flavonoid methylation</u>

Flavonoid MTs are divalent-cation-independent enzymes. A few exceptions include *M. crystallianum* PFOMT (Kopycki et al., 2008), *Vitis vinifera* anthocyanin OMT (Lücker et al., 2010), and *O. sativa* ROMT (Lee et al., 2008) where catalysis requires the divalent cation Mg^{2+}. All classes of flavonoids have been shown to exhibit *O*-methylation in all positions occupied by a hydroxyl group. cDNA clones of MTs have been isolated and characterised for enzymes that catalyse methylation of almost all classes of flavonoids including, flavones (Christensen et al., 1998), flavonols (Gauthier et al., 1996), flavanones (Rakwal et al., 2000), anthocyanins

(Hugueney et al., 2009), isoflavonoids (Zubieta et al., 2001), chalcones (Zhao et al., 2020) and stilbenes (Chiron et al., 2000).

Enzymatic methylation of flavonoid A, B, and C rings is catalysed by distinct enzymes. A great degree of this regiospecificity is shown by OMTs which methylate flavonol C-ring *3-OH*. For example, biosynthesis of poly-methylated flavonol glucosides in *C. americanum* was found to show a series of five step-wise methylations, each characterised by distinct methyltransferase (Gauthier et al., 1996). Flavonoid OMTs that accept compounds other than flavonoid subclasses have also been shown to be strictly position-specific to the B-ring *meta*-hydroxyl (Ibdah et al., 2003). Even in the case of sequential methylations, where hydroxyls at two different positions were methylated by the same enzyme, the transfer was confined to hydroxyls in a specific ring. (Cacace et al., 2003; Zhou et al., 2006). In contrast, some exceptions have been seen in flavonoid OMT variants such as *M. crystallianum* PFOMT, which catalyses methylation of *6-OH* in the A ring and *3′-OH* in the B-ring of the flavonoid core (Kopycki et al., 2008), as well as *Citrus reticulata* OMT, which catalyses the methylation of B-ring *3′*-, *5′-OH* and A-ring *7-OH* of flavonoids (Liu et al., 2020).

Flavonoid OMTs usually accept aglycones, but enzymes accepting glycoside derivatives have also been identified, as observed for glycosylated anthocyanin OMTs in *Petunia hybrida*, *Solanum lycopersicon* (Gomez Roldan et al., 2014), and *Avina sativa* (Knogge and Weissenbock, 1984). Partially methylated glucosides were accepted as substrates by an OMT isolated from *C. americanum*, and in this case glycosylation was shown to be a prerequisite for subsequent methylation (Gauthier et al., 1996; Gomez Roldan et al., 2014).

ii. <u>Isoflavonoid methylation</u>

Isoflavonoids are flavonoid compounds with the B-ring attached to the third carbon on the C-ring (**Figure 1.4**B). OMTs identified so far catalyse methylation of the *7*- and *4′-OH* of isoflavonoids (Wu et al., 1997; Zubieta et al., 2001) and the *3-OH* in the case of 6a-hydroxymaackiain (**Figure 1.4**C; Liu et al., 2006).

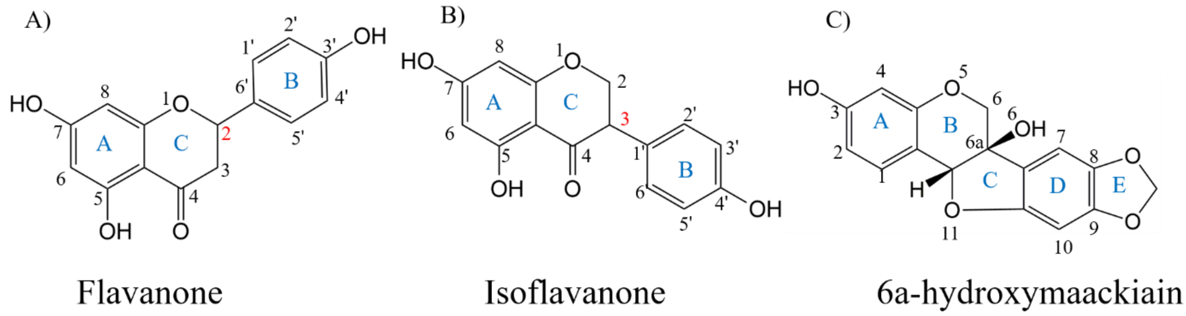


Figure 1.4: Schematic showing the B-ring position in flavonoid compounds.(A) flavonoids, (B)isoflavonoids and (C) 6a-hydroxymaackiain.

Information gained from cloning and characterisation of MTs accepting isoflavonoids has completely transformed the previously assumed pathway of medicarpin biosynthesis. Formononetin, an isoflavonoid, is the precursor of the phytoalexin medicarpin. Formononetin was believed to be derived from daidzein by 4′-*O* methylation. However, it was found that plant protein extracts failed to demonstrate this activity *in vitro*. Indirect proof of an alternative pathway came with the characterisation of *M. sativa* daidzein *7-O*-methyltransferase (D7OMT also called MsIOMT) (He et al., 1998). *In vitro* assays showed that MsIOMT accepted daidzein and converted it into *7-O*-methylated isoformononetin but surprisingly its overexpression in a cell culture of alfalfa resulted in the production of formononetin (*4′-O*-methylated daidzein) and medicarpin but not isoformononetin (He and Dixon, 2000). The 7-OMT activity was suggested to be an artefact of *in vitro* conditions (He et al., 1998). Subsequent *in planta* localization studies by Liu and Dixon (2001) showed co-localization of MsIOMT with an isoflavone synthase which produces 2,7,4′- trihydroxyisoflavanone (THI). The authors proposed that daidzein is not the *in vivo* substrate of IOMT, but of THI (Liu and Dixon, 2001). Based on this result, further *in vitro* assays were carried out and the new pathway via THI was verified when *M. truncula*ta 2,7,4′-trihydroxyisoflavanone *4′-O*-methyltransferase (HI4′OMT) accepted THI as a substrate but showed no activity with daidzein (Liu et al., 2006). MtHI4′OMT and MsIOMT have only ~49% sequence identity and their crystal structures further proved the structural basis of their differential substrate preference (Liu et al., 2006; Zubieta et al., 2001). On the other hand, MtHI4′OMT showed an 83 % sequence similarity to orthologous isoflavonoid OMT (+)-6a-hydroxymaackiain 3-OMT (HMM) from *Pisum sativum* involved in pisatin synthesis (Wu et al., 1997), and it was consequently suggested that in *P. sativum* duplication of an ancestral gene encoding the general biosynthetic gene HI4′OMT led to the formation of HMM (Akashi et al., 2006; Liu et al., 2006).

iii. Chalcone methylation

MTs that act on chalcones include *M. sativa* chalcone OMT (ChOMT) (Zubieta et al., 2001) and prenylated chalcone desmethylxanthohumol OMT from *Humulus lupulus* (hop) (Nagel et al., 2008). However, the hop genes, HlOMT1 and HIOMT2, show litttle similarity to ChOMT, and a phylogenetic analysis clustered them together with alkaloid OMTs (Nagel et al., 2008). HlOMT1 was found to require a divalent cation, and HlOMT2 showed a broad substrate specificity and catalysed 4'-methylation of xanthohumol, which is not yet been isolated from hops.

iv. *C*-methylation in flavonoids

Many *C*-methylated flavonoid compounds have been isolated and identified from various plants including *Eucalyptus* (Dao et al., 2010; Goodger et al., 2016; Sutthivaiyakit et al., 2013; Wollenweber et al., 2000); however, the steps involved in their biosynthesis is not clear. There is limited sequence information available on plant CMTs: sterol-methyltransferase from *A. thaliana*, maize, and soybean (Grebenok et al., 1997), and γ-tocopherol methyltransferases from *A. thaliana* and *Capsicum annum* (Koch et al., 2003). Evolutionary studies have shown that the sterol-CMTs are likely ancestral precursors of alkaloid NMTs (Cázares-Flores et al., 2016).

No sequence data is available for CMTs acting on flavonoids or any closely related phenolic compounds in plants. Therefore, the mechanism of *C*-methylation in plants is still unknown. Two theories have been proposed for the incorporation of *C*-methyl into the flavonoid pathway: a direct transfer to the flavanone backbone and an indirect transfer earlier into a chalcone. The common occurrence of direct transfer of *C*-methyl group to phenolic compounds by orthologous bacterial CMTs (Tengg et al., 2012) supports the hypothesis that the *C*-methyl group could be incorporated into flavanone aglycones or *O*-methylated flavanones through a direct enzymatic transformation (Goodger et al., 2019). In such a case, enzymes specific for the 6- and 8- carbon or a bispecific CMT could explain the formation of all identified *C*-methylated flavanones in *Eucalyptus* (Goodger et al., 2019; Marsh et al., 2019). On the other hand, the second theory of indirect transfer earlier into a chalcone was supported by the finding of a polyketide synthase mediated reaction in *Pinus strobus* where a non-physiological substrate methylmalonyl-CoA instead of malonyl-CoA was incorporated during the diketide or triketide conversion step of chalcone synthesis (**Figure 1.5**; Schröder et al., 1998). This model explains the formation of mono *C*-methylated flavanones, but it does not necessarily explain the synthesis of di-*C*-methylated flavonoids extracted from various species. According to this model 6,8-dimethylated flavanone synthesis would require methylmalonyl-CoA use in the second as well as a third condensation reaction. Further evidence substantiating this pathway or the presence of methylmalonyl-CoA in *C*-methylated flavonoid containing plants are not yet available.

A recently published mini-review proposed a new mechanism of *C*-methylation in bacteria where a SAM-dependant-CMT acted as a methyltransferase as well as a terpene cyclase (Piechulla et al., 2020). Briefly, in the teleocidin B biosynthesis pathway, methylation catalysed by the CMT (TleD) resulted in a subsequent carbocation (carbonate ion) formation in the geranyl group of teleocidin A1 and this promoted the cationic reaction cascade to form a cyclic monoterpenoid (Yu et al., 2016). The schematic of the mechanism is given in **Figure 1.6**. The crystal structure of TleD was elucidated and showed striking similarities to plant phenolic methyltransferase in its overall structure (Yu et al., 2016). One such feature includes an additional N-terminal helix of the monomer chain acting as a backwall of the neighbouring monomer's catalytic site. In addition, the enzyme formed a hexamer in solution which was essentially three groups of N-terminal domain-swapped TleD homodimers.

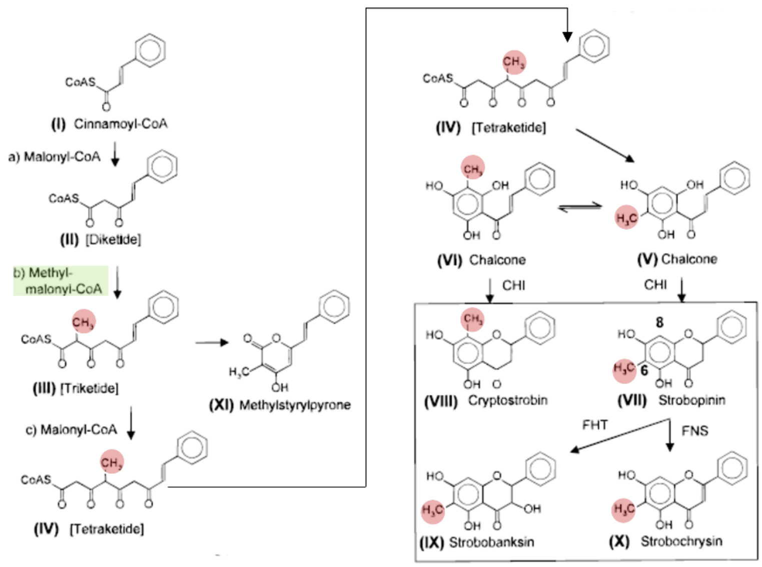


Figure 1.5: Schematic of the flavanone *C*-methylation pathway suggested by Schröder et al. (1998) showing the incorporation of *C*-methyl into the flavonoid pathway via an indirect transfer earlier into a chalcone. Reprinted with permission of American Chemical Society.

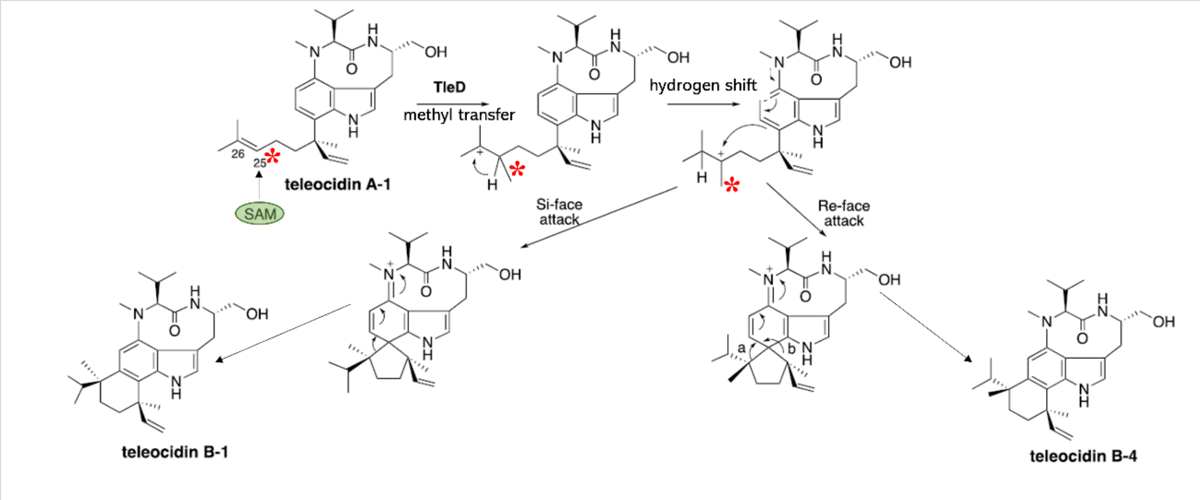


Figure 1.6: Methylation mechanism of SAM dependant CMT- TleD. (Awakawa et al., 2021; Awakawa et al., 2014). The methylation catalysed by TleD initiate a cationic reaction cascade to form a cyclic monoterpenoids. Site of methylation is denoted by red star. Reprinted with permission.

It is interesting to note that all plant CMT sequences isolated so far have *C*-methylation occurring in the terpene or isoprene sidechains of the methyl acceptor (sterols are triterpenoids and tocopherol has a terpene side chain). In flavonoid biosynthesis, the polyketone intermediate also resembles a long terpene hydrocarbon chain which is subsequently cyclized to generate the flavanone precursor chalcone. Based on the above-summarised studies, the possibility of a mechanism analogous to bacterial *C*-methylation followed by cyclization seems very likely in the case of plant flavonoid *C*-methylated compound biosynthesis. It is tempting to speculate that evolutionary variations might have enabled the terpenoid multifunctional methyl-cyclase enzyme to take part both in flavonoid as well as terpenoid biosynthesis.

1.3.3 Alkaloid methylation: *O*- and *N*-methyltransferases

Alkaloids are broadly defined as low molecular weight heterocyclic nitrogenous compounds (Morris and Facchini, 2019). In contrast to flavonoids, alkaloid compounds are extensively used as pharmaceuticals. Some such pharmacologically active compounds include morphine (analgesic), papaverine (muscle relaxant), codeine, and sanguinarine (anticancer) (Singh et al., 2019). Interestingly, all known alkaloid biosynthetic pathways employ one or more *O*- or *N*-methyltransferases (NMTs). Characterised MTs in these categories can be grouped under the different subclasses upon which they act; namely, benzylisoquinoline alkaloids (e.g. scoulerine, norcoclaurine, pavine, morphine, codeine), monoterpenoid indole alkaloids (e.g. catharanthine, vindoline, ajmaline, vinblastine), Amaryllidaceae alkaloids (e.g. galanthamine, lycorine, and crinine), tropane alkaloids (e.g. putrescine), and purine alkaloids (e.g. caffeine, xanthosine, trihydroxanthosine, 1,3,7-trimethyluric acid).

i. Monoterpene indole alkaloids (MIA)

In the monoterpene indole alkaloid pathway, MTs mostly catalyse the terminal steps (De Luca et al., 2012); for example, 16-hydroxytabersonine-16-*O*-methyltransferase (*Cr*16OMT) in *C. roseus* (Levac et al., 2008), noribogane 10-OMT in *Tabernanthe iboga* (Farrow et al., 2018), and 10-hydroxycamptothecin OMT from *Camptotheca acuminate* (Salim et al., 2018). In contrast, a carboxyl OMT gene encoding loganic acid methyltransferase (*Cr*LAMT) acting much earlier in the biosynthesis of vindoline precursors (monoterpene secoiridoid pathway) was discovered later from *C. roseus* leaf epidermis (Murata et al., 2008). In addition to 16-OMT and *Cr*LAMT, one more methyltransferase acting in the biosynthesis of monoterpene indole alkaloids has been identified from *C .roseus*. Liscombe and colleagues (2010) characterised a recombinant enzyme that catalyses aromatic nitrogen methylation of a synthetic substrate 2,3- dihydrotabersonine, which was later confirmed to be the NMT responsible for conversion of 16-methoxy-2,3-dihydro-3-hydroxytabersonine (Liscombe and O'Connor, 2011).

ii. Purine alkaloids

The biosynthesis of purine alkaloids has been well studied, particularly those involved in caffeine biosynthesis. Caffeine is synthesised from the purine alkaloid xanthosine by a series of NMTs (Misako and Kouichi, 2004). Cloned representatives of this group include derivatives from *Coffea arabica*, *Camellia. sinensis*, *Theobroma cacao,* and *Citrus sinensis*. The NMT genes in *C. arabica*, *Ca*XNMT and *Ca*DXNMT, show more than 80% sequence identity to each other but show only 40 % similarity to the orthologous genes in *Camellia sinensis*, suggesting the independent evolution of the caffeine biosynthetic pathways in coffee and tea. All the purine alkaloid NMTs show higher similarity to plant phenylpropene OMT genes, like salicylic acid OMT, than to other alkaloids NMTs (Misako and Kouichi, 2004). The crystal structure of four purine alkaloid MTs have been elucidated from *Coffea canephora* (McCarthy and McCarthy, 2007) and recently from *C. assamica var. kucha* (Zhang et al., 2020).

iii. Isoquinoline alkaloids

Benzylisoquinoline alkaloid (BIA) biosynthesis is one of the most studied metabolic pathways and has a large number of characterised enzymes. Several plant species, for example opium poppy (*P. somniferum*), have emerged as model systems for alkaloid biosynthesis. Like phenylpropanoids and flavonoids, the BIA are also amino acid (tyrosine) derivatives. All MTs reportedly catalyse different pathways of BIA synthesis and they have been isolated from opium poppy *P. somniferum* (Singh et al., 2019), Japanese goldthread *C. japonica* (Morishige et al., 2000) and meadow rue *T. flavum* (Frick and Kutchan, 1999).

The vast majority of the alkaloid MTs are active in the final biosynthetic stage. Some exceptions exist, such as putrescine NMT, which catalyses the first step in the pathway leading to tropane alkaloids and nicotine. Anthranilate NMT from *R. vulgaris* also acts early in acridone alkaloid biosynthesis. Similarly, in MIA synthesis, loganic acid OMT acts on an earlier step in the synthesis.

In BIA synthesis, the order in which the regiospecific MTs act differentiates the pathways. A survey of the existing literature provides ample evidence that reticuline is the branch point from which different classes of BIA pathways emerge (Morris and Facchini, 2019). An exception to this is papaverine biosynthesis in which 4′-*O*-methylation of (S)-coclaurine occurs before coclaurine NMT (CNMT) diverges the reticuline synthesis pathway from its otherwise usual biosynthetic sequence 6-OMT→ CNMT → 4′-OMT. An alternate pathway for papaverine biosynthesis that branches out from (S)-reticuline was suggested with the characterisation of reticuline 7-*O*-methyltransferase (R7OMT) from *P. somniferum* (Han et al., 2010), but virus-induced gene silencing studies of R7OMT and CNMT genes presented confirmative evidence that branching from (S)-coclaurine is the preferred route (Desgagné-Penix and Facchini, 2012). The progress through both pathways requires a 3′-OMT, but no such gene has yet been characterised.

Another set of MTs that has gained much attention in recent years is that involving the isoquinoline-type alkaloid metabolites (Diamond and Desgagné-Penix, 2016) from the Amaryllidaceae plant family, designated as Amaryllidaceae alkaloids (e.g. galanthamine, lycorine, and crinine). Unlike other established classes of alkaloids, these compounds are not widely extracted for human use. Galanthamine, which is a neuroprotectant, is the only compound currently being used pharmaceutically (Heinrich and Teoh, 2004). Among Amaryllidaceae alkaloid MTs, norbelladine 4′-OMT (NpN4OMT) from daffodil was the first enzyme to be identified and characterised (Kilgore et al., 2014). Since then, several NpN4OMT homologous genes have been characterised from several Amaryllidaceae species (Hotchandani et al., 2019; Li et al., 2019; Reis et al., 2019; Sun et al., 2018). It is interesting to note that recombinant enzymes from *Lycoris* (*L. aurea and L. radiata*), which have a high similarity to daffodil N4OMT, are bifunctional enzymes that catalyse methylation of both the *3′*- and 4*′-OH* of norbelladine, among which *L. aurea* OMT preferentially catalyse norbelladine *3′-OH* methylation (Li et al., 2019; Sun et al., 2018). Recently, three new alkaloids of this class with *3′-OH* methylation were identified, and their biosynthetic pathway was anticipated to involve norbelladine and intermediates with a *3′-O*-methylation and N-methylation (not necessarily in any order) (Ka et al., 2020). Therefore, it can be presumed that the N4OMT variant itself might be involved in the synthesis of *3′-O*-methylated derivatives of norbelladine across different species. The lack of any identified biosynthetic enzymes or genes in this biosynthetic pathway makes it difficult to predict the progression sequence. Further enzymatic characterisation needs to be done to verify these predictions.

cDNA clones of monoterpene isoquinoline type alkaloid OMTs from *P. ipecacuanha* elucidated major steps in the pathway of ipecac alkaloid biosynthesis and they showed more sequence similarity to flavonoid OMTs than other BIA-OMTs. The major difference from the other isoquinoline alkaloid pathways is the methylation of the 7-hydroxyl group of the isoquinoline skeleton before 6-*O*-methylation. The three ipecac alkaloid OMTs isolated were suggested to be responsible for the formation of all methyl derivatives of ipecac alkaloids in *P. ipecacuanha* (Nomura and Kutchan, 2010).

Reading through the above summary of characterised MTs, organised according to the type of compounds/biosynthetic pathway, one realises there are many evident variations in each class which show similarities to enzymes from other groups. Therefore, the next section will describe various attempts to systematically classify MTs.

1.4 Classification of plant methyltransferases

The approaches to classify plant MTs more efficiently have changed enormously over time with the identification of an increasing number of plant small molecule MTs. Many plant MTs act on the oxygen atom of hydroxyl/carboxyl groups of phytochemicals, but some can also attack carbon, nitrogen, and sulphur atoms. Therefore, the simplest approach to categorization of plant small molecule MTs is according to the site where methylation occurs: *O*-methyltransferase, *C*-methyltransferase, *N*-methyltransferase, and *S*-methyltransferase (Ibrahim and Muzac, 2000). By sequence comparison of 26 plant cDNA clones, Ibrahim et al. (1998) proposed a classification system of plant small molecule OMTs in which enzymes were arranged according to the nature of their methyl acceptors. Unfortunately, this system did not necessarily consider the phylogenetic origin and the catalytic properties of the enzymes, as it was based purely on the molecular and biochemical (substrate specificity) properties of plant MTs sequenced thus far. With the discovery of more unique MTs sequences and further crystal structure studies, this system appeared inadequate (D Auria et al., 2003). This called for a modified classification system considering all aspects including evolutionary relationship and tertiary structure.

An improved classification system was proposed by Joshi and Chiang (1998)), whereby they grouped fifty-six plant MTs into two categories, PL-OMT I and PL-OMT II (PL for plant), based on the molecular weight of the protein, metal binding and spatial separation between conserved motifs. PL-OMT I comprised enzymes with a molecular weight of 23 to 27 kDa, a requirement for Mg^{2+} for their activity, and had a spatial separation of 19 and 24 amino acids between the conserved motifs A, B, and C, respectively. This group contained enzymes that specifically utilized CoA-esters as substrates [formerly referred to as CCoAOMTs (Davin and Lewis, 1992)] and had higher sequence similarity with the mammalian catechol OMTs than to the rest of the plant OMTs. On the other hand, PL-OMT II comprised enzymes with a molecular weight of 38 to 43 kDa that do not require Mg^{2+} for their activity and with a spatial separation of 52 and 30 amino acids between motif A, B, and C, respectively. PL-OMT II contained the rest of the plant MTs that catalyse a wide variety of substrates including CoA derivatives. Not all candidates used in this classification were functionally characterised, and it failed to accommodate variants of SAM MTs that were subsequently discovered.

The revelation of the crystal structure of isoflavone *O*-methyltransferase and chalcone *O*-methyltransferase from *Medicago sativa* (alfalfa) provided new insights into the field of plant small molecule methylation (Zubieta et al., 2001), and structural modelling of related OMTs proved to be a better tool to understand evolutionary relationships within the enzyme group. The application of these tools enabled Noel et al. (2003) to devise a system that considered sequence similarity as well as the tertiary architecture of plant small molecule OMTs in which they were grouped into three families named Type 1 through 3. Type 1 methylate hydroxyls of phenylpropanoid based compounds like flavonoids, isoflavonoids, aurones, and chalcones. Type 2 methylation is specific for CoA-derivatized phenylpropanoid compounds like caffeoyl and feruloyl CoA and appears to be less diverse than Type 1. Both Type 1 and Type 2 are exclusively plant OMTs. The Type 3/SABATH family constitutes a larger family of MTs that are homodimers with an approximate subunit mass of 40 kDa and members have no significant sequence similarity to the above-mentioned Type 1 and 2 plant MTs. Type 3 was initially designated to enzymes that transfer a methyl moiety to a carboxyl group (also termed as OMTs by the author) resulting in the formation of methyl esters. Nonetheless, sequence similarity searches in databases resulted in the identification of a much larger number of similar putative proteins which also included alkaloid N-MTs (NMT). To distinguish this group from other existing MTs, it was designated as SABATH family [named after three of the first genes Salicylic acid carboxyl MTs (**SA**MT) from *Clarkia breweri (*Ross et al., 1999), benzoic acid carboxyl methyltransferase (**BA**MT) from snapdragon (Murfitt et al., 2000), and **Th**eobromine synthase from *Coffea arabica (Kato et al., 2000)*] (D Auria et al., 2003). Although this categorising was not an attempt to classify plant MTs *per se*, rather an attempted grouping based on phylogenetic analysis of only twelve representative MTs and structural studies of three among them, this system is extensively cited as a classification system of plant MTs. Despite this, later classification ventures with additional characterised candidates have regrouped the genes described by Noel and his colleagues (Lam et al., 2007).

The above-mentioned classification systems, for example the ones suggested by Joshi and Chiang (1998) and Noel et al. (2003), while useful on their own, contradict each other due to conflicting nomenclature; they thus can create confusion. Various authors quote different class names interchangeably. For instance, most publications refer to Joshi’s (1998) PL-OMTI, which contains caffeoyl-CoA OMTs, as Class I OMT (Lam et al., 2007; Liu et al., 2019) but Liu et al. (2016) referred to PL-OMTI as Group II. Furthermore, some refer to Noel’s (2003) Type 1 OMTs as Class I which has phenylpropanoid OMTs instead of CoA OMTs (Levac et al., 2008).

The latest attempt to classify plant OMTs (Lam et al., 2007) can be considered as a refined version of Joshi and Chiang’s classification. Addressing the evolutionary relationships, sixty-one candidates [nine belong to PL-OMTI (called Class I) and the remaining to PL-OMTII (called Class II)] were grouped into two distinct monophyletic lineages A and B. Even though the study claims to have used all biochemically characterised candidates, out of the sixty-one candidates used, *Z. mays* (M73235) and *M. piperita (*AY337458) MTs are putative genes whose substrates have not been confirmed and two are *N*-MTs (*Accession no:* U54767, AB073908). A significant variation from the previous classification systems was the inclusion of carboxylic acid OMTs (previously considered as a distinct Type 3 group by Noel and his colleagues) along with caffeoyl-CoA OMTs in lineage A. All remaining OMTs, including hydroxycinnamic acid, flavonoid, and alkaloid OMTs, were in lineage B. Lineage B further assembled into three distinct clades. The authors attempted to find a trend behind this assemblage based on substrates used, but there were some anomalies such that several OMTs that acted on structurally similar substrates were assigned to different clades. Placing of OMTs that acted on CoA-derivatives as well as flavonoids remained a problem. The OMTs from *S. longipes* (L22203) (Zhang and Chinnappa, 1997) and *M. crystallinum* (AY145521) (Ibdah et al., 2003) utilize both caffeoyl-CoA esters and flavonoids, and they showed sequence similarity to CoA exclusive enzymes, therefore they were grouped with CCoAOMTs. Many OMTs which accepted both caffeoyl-CoA, caffeic acid as well as flavonoid compounds were later identified (Lee et al., 2008; Tsai et al., 2006; Widiez et al., 2011). Serious shortcomings of this system arose with the characterisation of *V. vinifera* anthocyanin OMT (Hugueney et al., 2009). On a sequence level, the enzyme shows all the characteristics of a CoA derivative OMT. In addition to the original study proving anthocyanins and flavonols as substrates, further *in vitro* feeding studies with caffeoyl-CoA and transient expression in tobacco (Lücker et al., 2010) proved that this Mg^{2+} dependant enzyme prefers flavonoids over CoA esters as substrates. Since these do not fit into any of the currently described distinct families, several writers have coined broad terms such as CCoAOMT-like family, COMT superfamily, and COMT-like subfamily (Lücker et al., 2010) to accommodate the variants. Taken together, these results suggest that distinctions based on substrate specificity or segregation by a structural group of substrates are an inadequate basis by which to classify plant OMTs.

It is noteworthy that none of these classification systems have addressed *C*-MTs because of the considerable number of mismatches or at times even the absence of conserved motifs (Lam et al., 2007). In conclusion, a comprehensive and consistent system is still in demand. Along with fully characterised plant methyltransferase sequences in databases, more tertiary structure studies and knowledge about their cellular locations is necessary to establish a reliable and clear classification system for plant small molecule OMTs.

1.5 The architecture of plant methyltransferases

The tertiary structure of an enzyme is important to understand the chemistry behind substrate specificity and catalysis. Protein X-ray crystallography is a tool that has been used extensively for determining the structure of various plant MTs. Structural characterisation of the first SAM-dependant methyltransferase, rat catechol OMT, gave a clear indication of the Mg^{2+} dependant positioning of the substrate for deprotonation and subsequent methylation (Vidgren et al., 1994). The secondary and tertiary structure of the enzyme was found to be comparable to DNA MTs with a high identity at the cofactor binding core. The theory that all MTs share the same tertiary structure was disproved years later by the crystallographic structure elucidation of the two plant MTs: chalcone *O*-methyltransferase (ChOMT) and isoflavone *O*-methyltransferase (IOMT)- from *Medicago sativa* (alfalfa) (Zubieta et al., 2001). Since then, the structure of several phenolic plant MTs has been elucidated and the crystal structure data has revolutionised our understanding of the diverse nature of transmethylation mechanisms. Liscombe and colleagues (Liscombe et al., 2012) have extensively reviewed the structural features of a broad range of small molecule MTs from bacteria, plants and animals, involved in transferring a methyl group to *O*-, *C*-, *N*-, *S*-atoms, and specifically included twelve structurally characterised plant MTs, one of which belonged to a rare halide methyltransferase from *Arabidopsis thaliana.* In addition, a recent review by Morris and Facchini (2019) has summarised the structural and functional features of *O*- and *N*-MTs of benzylisoquinoline alkaloid biosynthesis. Hence, the following sections of this chapter will focus on a general overview of methyltransferase tertiary and quaternary structure followed by a comparative analysis of the structural features that determine their substrate specificity and catalysis. The comparative review will cover the twenty-three structurally characterised phenolic *O*- and *N*-MTs reported so far accepting phenylpropanoid, flavonoid and alkaloid substrates (**Table 1.1**). This knowledge can help with interpretation of the evolutionary strategies utilized by these enzymes, particularly in response to physiological and ecological pressures, that enable them to accommodate compounds with different chemical scaffolds without creating additional burdens on plant machinery.

Table 1.1: Structurally characterised plant MTs accepting phenolic compounds

PDB ID	Name	Abbreviation	Group	Organism	Reference
3P9C	Caffeic acid OMT- LpOMT1	*Lp*OMT1	OMT Phenylpropanoid	*L. perenne*	(Louie et al., 2010)
1KYZ	Caffeic acid OMT	MsCOMT	OMT Phenylpropanoid	*M. sativa*	(Zubieta et al., 2002)
4PGG	Caffeic acid OMT	SbCOMT	OMT Phenylpropanoid	*S. bicolor*	(Green et al., 2014)
1SUI	Caffeoyl CoA 3-OMT	MsCoAOMT	OMT Phenylpropanoid	*M. sativa*	(Ferrer et al., 2005)
5KVA	Caffeoyl CoA OMT	SbCoAOMT	OMT Phenylpropanoid	*S. bicolor*	(Walker et al., 2016)
5XOH	Bergaptol OMT	PpBMT	OMT Phenylpropanoid	*P. praeruptorum*	(Zhao et al., 2020)
3B5I	Indole acetic acid OMT	IAMT	OMT Phenylpropanoid	*A. thaliana*	(Zhao et al., 2008)
1ZHF	6a-hydroxymaackiain OMT	HI4'OMT	OMT Flavonoid	*M. trunculata*	(Liu et al., 2006)
1FP2	Chalcone OMT	ChOMT	OMT Flavonoid	*M. sativa*	(Zubieta et al., 2001)
1FP1	Isoflavone OMT	IOMT	OMT Flavonoid	*M. sativa*	(Zubieta et al., 2001)
3C3Y	Phenylpropanoid/flavonoid OMT	*PFOMT*	OMT Flavonoid	*M. crystallianum*	(Kopycki et al., 2008)
5ICE	Norcoclaurine 6-OMT	Tf6OMT	OMT Alkaloid	*T. flavum*	(Robin et al., 2016)
1M6E	Salicylic acid OMT	*SAMT*	OMT Phenylpropanoid	*C. breweri*	(Zubieta et al., 2003)
6I6K	Scoulerine 9-OMT	PSMT1	OMT Alkaloid	*P. somniferum*	(Cabry et al., 2019)
6NEI	Scoulerine 9-OMT	TfS9OMT	OMT Alkaloid	*T. flavum*	(Valentic et al., 2020)
6C8R	Loganic acid OMT	LAMT	OMT Alkaloid	*C. roseus*	(Petronikolou et al., 2018)
2EFJ	1,7-dimethylxanthine NMT	DXNMT	NMT Alkaloids	*C. canephora*	(McCarthy and McCarthy, 2007)
2EG5	Xanthosine NMT	XNMT	NMT Alkaloids	*C. canephora*	(McCarthy and McCarthy, 2007)
6LYH	1,3,7-trimethyluric acid NMT	CkTcS N9MT	NMT Alkaloids	*C. assamica var. kucha*	(Zhang et al., 2020)
6LYI	1,3,7-trimethyluric acid NMT	CkTbS	NMT Alkaloids	*C. assamica var. kucha*	(Zhang et al., 2020)
6GKV	Coclaurine NMT	CjCNMT	NMT Alkaloid	*C. japonica*	(Bennett et al., 2018)
5KOK	Pavine NMT	PavNMT	NMT Alkaloid	*T. flavum*	(Torres et al., 2016)
6P3N	Tetrahydroprotoberberine NMT	GfNMT	NMT Alkaloid	*G. flavum*	(Lang et al., 2019)

To date, there are only four structurally studied flavonoid OMTs and three of these (ChOMT, IOMT, and HI4'OMT) are involved in isoflavonoid biosynthesis. The substrate of ChOMT is the chalcone, isoliquiritigenin to 2'-methoxychalcone (Zubieta et al., 2001). Via this conversion, ChOMT controls the biosynthetic flux through subsequent flavonoid pathways by restricting the cyclization of isoliquiritigenin to the flavanone liquiritigenin - consequently facilitating isoflavonoid biosynthesis. IOMT is involved in the biosynthesis of medicarpin and has been shown to have unusual properties, depending on the experimental system used. Specifically, it has been shown to methylate the B-ring 4'-hydroxyl of isoflavone substrates using *in vivo* experiments, but in contrast *in vitro* it acts on the A-ring 7-hydroxyl of isoflavone substrates (Zubieta et al., 2001). Owing to the instability of its *in vivo* isoflavone substrate, the crystal structure of IOMT was developed with a similar substrate daidzein, an isoflavone. Under *in vitro* conditions, IOMT converts daidzein to isoformononetin (He and Dixon, 1996). HI4'OMT catalyses methylation of the 4′-OH of 2,7,4'- trihydroxyisoflavanone. The fourth structurally studied flavonoid OMT, PFOMT, is a rather peculiar enzyme which accepts both phenylpropanoids as well as flavonoids as substrates and unlike other flavonoid OMTs it is cation dependant (Kopycki et al., 2008).

Structurally characterised plant MTs that are involved in the hydroxycinnamic acid pathway include the various caffeic acid OMTs (MsCOMT and LpOMT1, SbCOMT) and the CoA ester specific OMTs. Caffeic acid *O*-methyltransferase is an enzyme that is ubiquitous in lignin-producing plants. Although they are termed caffeic acid OMTs, they can methylate a variety of phenylpropanoids on their hydroxyl groups on the 3rd as well as 5th position (Zubieta et al., 2002). The crystal structure of caffeic acid OMT (MsCOMT) from *Medicago sativa*, with its two active substrates ferulic acid and 5-hydroxyconiferaldehyde (Zubieta et al., 2002), was the first to be crystallized in this category.

Of the OMTs transferring a methyl group to a carboxyl group, crystal structures have been elucidated for a salicylic acid methyltransferase (SAMT) and an indole acetic acid methyltransferase (IAMT), which are involved in the floral scent volatile biosynthesis. (Zhao et al., 2008; Zubieta et al., 2003). SAMT isolated from petals of *Clarkia breweri* converts salicylic acid to methyl salicylate, a floral scent volatile ester. IAMT from *A. thaliana* acts on the plant hormone indole acetic acid.

Plant alkaloid MTs are currently receiving the most attention in terms of structure studies, largely due to the medicinal properties or potential of the products they produce. This includes various *O*-MTs (hydroxy, carboxy) as well as *N*-MTs. Due to the pharmacological relevance of BIA compounds, there has been a recent influx in the structural repository of enzymes acting in BIA biosynthesis. Structures in this category include scoulerine 9-OMT from *P. somniferum* (Cabry et al., 2019) and *T. flavum* (Valentic et al., 2020), (S)-norcoclaurine 6-OMT and pavine NMT from *T. flavum* (Robin et al., 2016; Torres et al., 2016), coclaurine NMT from *C. japonica* (Bennett et al., 2018), and tetrahydroprotoberberine NMT from *G flavum* (Lang et al., 2019). Furthermore, four different NMTs acting in caffeine biosynthesis have been structurally characterised so far, namely XNMT, DXNMT (McCarthy and McCarthy, 2007), CkTcs and N9OMT (Zhang et al., 2020).

It is surprising that most crystal structures generated with the enzyme's physiological substrates have always resulted in imprecise forms and in many cases bound analogues or SER variants gave closed (viable for methylation) structures. For example, analysis of the ligand-binding sites of PSMT was not possible with the weakly diffracting structure produced with either the native enzyme or the complex with SAM (3.5Å and 3 Å, respectively). Therefore, using a surface engineering approach, surface entropy reduction (SER) variants, with Lys114 and Lys115 converted to alanine (conversions had a negligible effect on enzyme activity), were used to develop high-resolution structures (Cabry et al., 2019). Furthermore, the native substrates of PavNMT and CjCNMT were not bound at the active site in any crystallisation trials (Bennett et al., 2018; Torres et al., 2016). The reason behind this is not clear, although it may be due to the instability of the natural substrate under the crystallographic condition, as was the case for IOMT (Zubieta et al., 2001), or perhaps due to the non-native conditions required for crystallographic studies. This raises doubt regarding the accuracy of depicted interactions extending to its natural substrate. In the latest study by Lang et al (2019), however, the GfNMT crystal structure provided the most reliable view of interactions as it produced high-resolution crystals complexed with natural substrate and is consistent with previously proposed theories.

1.5.1 Overall tertiary structure is moderately conserved for biochemically diverse substrates

Architectural analysis of MTs that act on distinct phenolic substrate classes suggests that there is overall conservation of protein tertiary structure. For example, even though amino acid residues of ChOMT and IOMT do not sequentially align, their overall tertiary structure looks strikingly similar (Zubieta et al., 2001). In both MTs, their quaternary structure is a homodimer. These results are in agreement with those reported by Zubieta (2002), where the tertiary structure of MsCOMT was also found to be very similar to that of the flavonoid OMTs ChOMT and IOMT. Liu et al. (2006) compared IOMT architecture with another isoflavanone OMT from *M. trunculata* and showed that even with a sequence similarity of only 51%, the overall structure is conserved, but with a shift in the topology of the N-terminal domain with respect to the C-terminal (Liu et al., 2006). Like phenylpropanoid OMTs, most of the alkaloid MTs are also homo-dimeric proteins with a similar tertiary fold.

In general, the common tertiary structure features of MTs include a small N-terminal domain at the dimer interface and a large C-terminal catalytic domain responsible for SAM binding and substrate methylation. The position of the active site is the cavity formed between the C-terminal domain of a monomer and the N-terminal domain helices of the other monomer (**Figure 1.7**).

The predominant theory that all plant OMTs have a similar fold (based on the earlier structures of flavonoid and MsCOMT) was disproved by the structural elucidation of mechanistically and structurally distinct, divalent cation dependant, *M. sativa* CCoAOMT (Ferrer et al., 2005). In structural, as well as evolutionary perspectives, CCoAOMTs are more closely related to the mammalian catechol OMT than to other plant OMTs. Unlike other phenolic OMTs, these enzymes lack an N-terminal dimerization domain and are strictly *meta*-position specific (Ferrer et al., 2005). Further 3D structure research has shown that there are some variants like *M. crystallianum* PFOMT that are promiscuous in substrate specificity accepting other phenylpropanoids and flavonoids and with a structure more similar to CCoAOMTs than to other flavonoid OMTs (Kopycki et al., 2008).

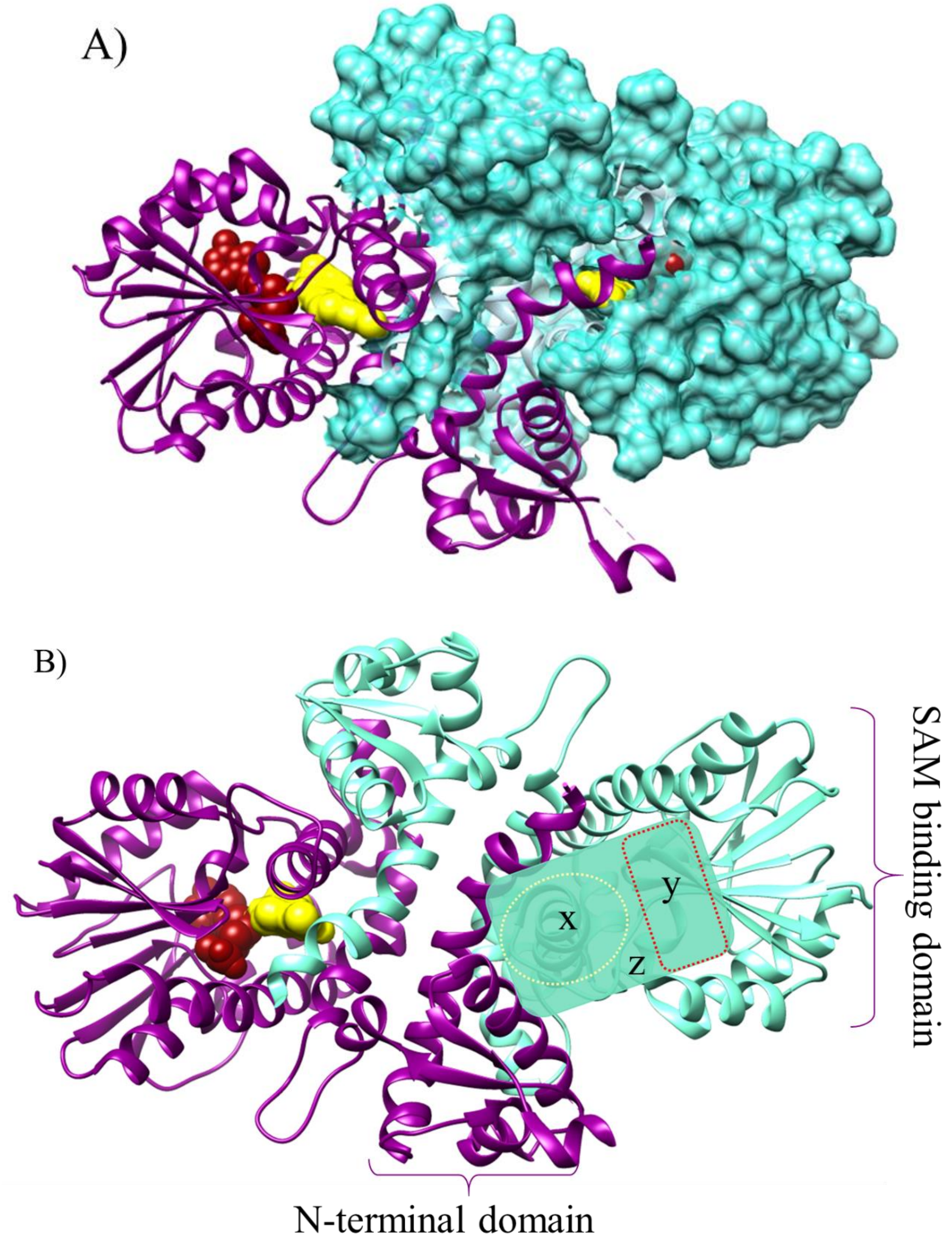


Figure 1.7: Overall architecture of plant MTs bound to methyl acceptor (yellow)and SAM (red). Monomer-I is coloured in cyan, and monomer II is coloured in purple. A) Monomer-I is depicted as space-filling model emphasizing the dimerization interface forming the back wall of the active site of the monomer-II. B) Ribbon diagram highlighting general structural features: x-methyl acceptor binding site; y-SAM binding site; z-active site pocket. Figures created based on the crystal structure of PSMT1 (PDB-6I6K) using Chimera software by K. Chandran.

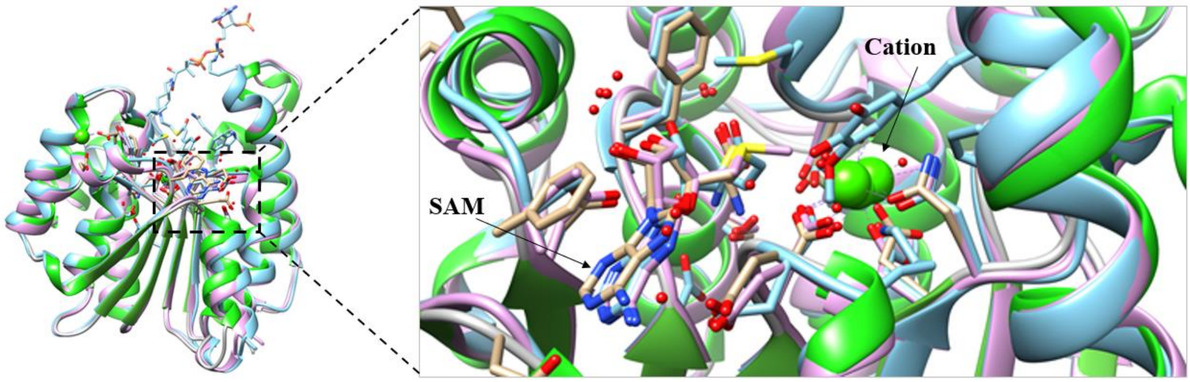


Figure 1.8: Tertiary structure of CCoAOMTS Cation dependant OMTs MsCoAOMT from *M. sativa* (cyan;), SbCoAOMT from *S. bicolor* (pink); PFOMT from *M. crystallinum* (green) superimposed showing conservation of overall tertiary structure as well as SAM and metal-binding residues (Ferrer et al., 2005; Kopycki et al., 2008; Walker et al., 2016). Figure created with Chimera software using structures deposited in protein data bank by K. Chandran.

Even though the CCoAOMTs are different from other phenolic MTs, tertiary structure superimposition of structurally characterised CCoAOMT enzymes shows that they align almost perfectly when superimposed with very good conservation of SAM and divalent cation positions (**Figure 1.8**). Similarly, there exists high fold similarity within the rest of the OMTs and NMTs catalysing a specific group of compounds (i.e. HCA, flavonoids and, alkaloids). Nevertheless, this observation is based purely on a limited number of published structural models and is highly likely to change with additional structural investigations.

There is a vast divergence in the structural scaffold and chemistry of methyl acceptors of plant MTs. Despite the similar fold, plant MTs can differentiate between substrates. MTs make use of subtle changes in tertiary architecture to accommodate substrates of different chemical scaffold/origin to facilitate ornamentation leading to this vast diversity. Some such noticeable features among various classes of plant MTs are described below in detail.

1.6 Structural determinants of substrate recognition

Active site architecture through structure studies has contributed extensively to our understanding of plant methyltransferase substrate specificity. In general, substrate recognition depends on van der Walls interactions, hydrogen bonding, steric effects, and aromatic interactions between enzyme residues and substrates (Lang et al., 2019).

1.6.1 SAM co-substrate binding domain - Rossmann fold elaborations

Despite the minor conservation of specific amino acid residues, all structurally characterised plant SAM dependant MTs have a similar C-terminal architectural core - an α/β Rossmann fold, with only occasional modifications. In the majority of plant MTs, irrespective of the nature of various classes of substrates, SAM binding follows a general conserved mechanism (Fauman et al., 1999; Valentic et al., 2020). The core Rossmann fold is formed by the alternating α-helices and β-strands of the polypeptide folding to form a seven stranded β-sheet flanked by three α-helices on either side making an αβα layered sandwich (**Figure 1.9**). The α-strand at the N-terminal (αZ) of the fold incorporates into the centre of the relatively planar β-sheet, creating a central switch point in topology with the seventh strand antiparallel to all other strands due to a reversed β hairpin (6↑ 7↓ 5↑ 4↑ 1↑ 2↑ 3↑) (**Figure 1.9**A; Fauman et al., 1999). The Rossmann fold takes up the majority of the active site, and central topology flipping results in a deep cleft which in turn acts as the SAM binding site. In plant-specific MTs, elaboration of this core fold is observed often by additional structures like short and long helices, N-terminal extensions, or β-hairpins. **Figure 1.9** shows a summary of the Rossmann fold architecture of plant MTs accepting different classes of substrates. It is evident from the comparison of plant methyltransferase 3D-structures that the N-terminal end of the core fold, which interacts and helps in SAM binding does not accommodate any sizeable amendments. Plant OMTs, however, seat discrete structures within the core Rossmann fold to accept diverse methyl acceptors (Liscombe et al., 2012). The substrate binds in the pocket formed by the C-terminal end of Rossmann fold, and the back wall formed by the neighbour monomer's dimerization domain helix. Therefore, functional evolution can be brought about by the structural appendages in the Rossmann fold like N-terminal extensions (PDB:1M6E) (Zubieta et al., 2003), insertion loop between β5 and αE (PDB:3C3Y) (Kopycki et al., 2008), absence or presence of additional helices (PDB:1FP2) (Zubieta et al., 2001), and an active site gate (Torres et al., 2016).

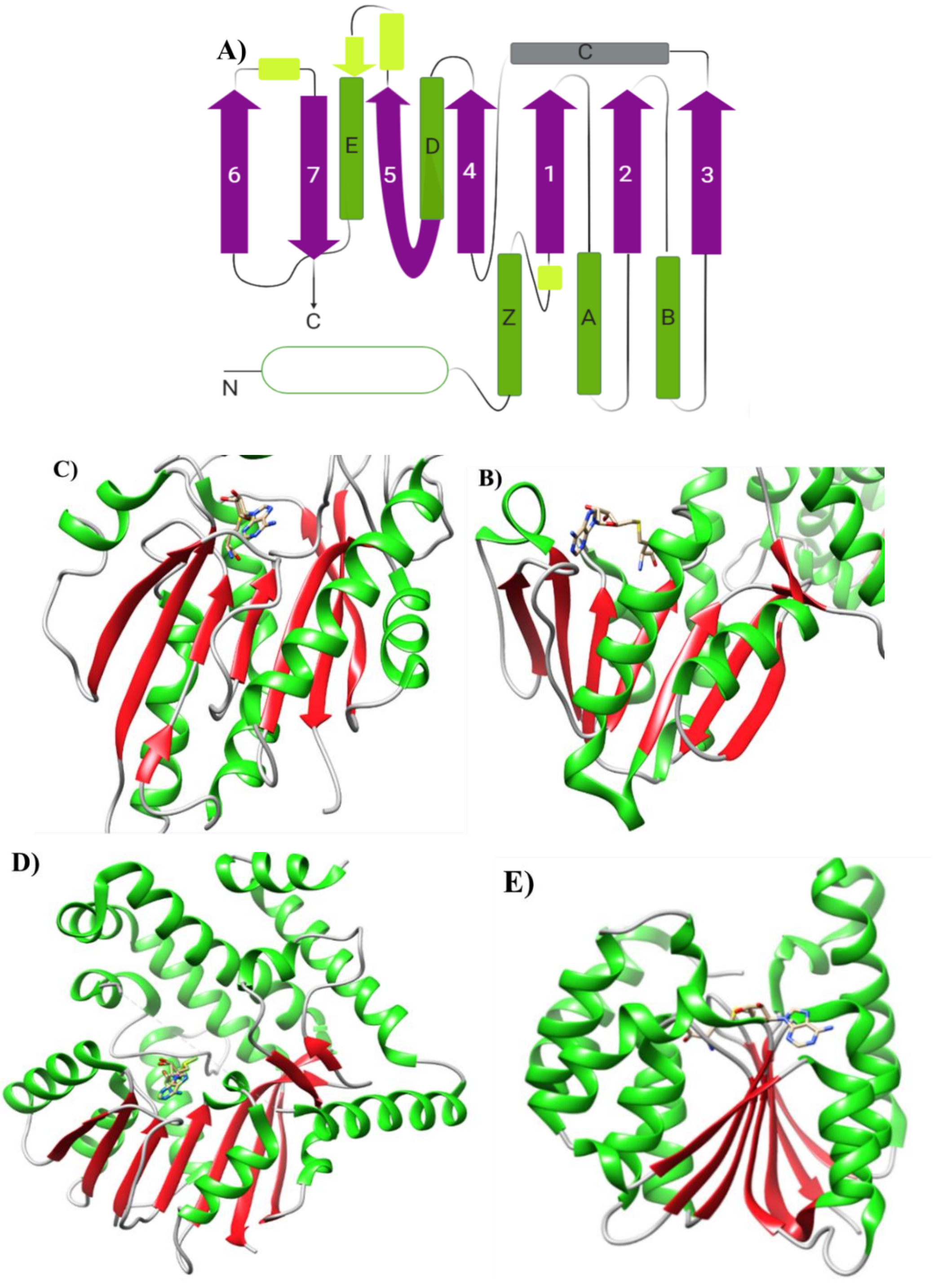


Figure 1.9: Rossmann fold architecture in plant MTs. A) Representative topology diagram of plant MTs Rossmann fold where rectangles and the arrows depict α-helices and β-sheets, respectively. β-sheets are numbered 1 through 7 from N-terminal→C-terminal direction. The αC (shaded grey) is absent in the majority of plant MTs except for metal dependant ones. The regions shaded in yellow are the areas that accommodate structural elaborations. N-terminal section of the protein, which is not the part of the Rossmann fold is depicted by the capsule-shaped region. B) Flavonoid OMT core fold (PDB-1FP2). C) Carboxy-OMT core fold (PDB-1MXE). D) Alkaloid-NMT core fold (PDB-6P3N). E) metal dependant OMT/CCoAOMT core fold (PDB-3C3Y). Figures created with Chimera software based on structures deposited in protein data bank by K. Chandran

For example, in ChOMT and IOMT, αC is absent and a helix is inserted between β5 and αE (Schubert et al., 2003). In the case of NMTs, a short helix is present between β4 and αD (Lang et al., 2019) and additional helices exist between the antiparallel strands β6 and β7 (Bennett et al., 2018; Torres et al., 2016). All structurally characterised divalent cation dependant plant OMTs have a well-conserved Rossmann fold without any modifications.

Proper positioning of the co-substrate for catalysis is mediated by a series of hydrogen bonding (direct or water-mediated) and van der Waals interactions via residues distributed across the various conserved motifs A, B, and C (Joshi and Chiang, 1998). The conserved residues of the plant MTs that are functionally important, more generally the motifs, are located typically in the C-terminal regions of β-strands or the adjoining loops (Fauman et al., 1999; Joshi and Chiang, 1998; Schubert et al., 2003). A major identity feature of the SAM binding pocket is a glycine-rich region [termed Motif A, with a consensus (V/I/L)(V/L)(D/K)(V/I)**GG**XX(**G**/A)] positioned at the end of sheet β1 spanning the following loop and the adjacent helix αA. A universal structural and sequence comparison of the SAM binding plant MTs has established that any substitutions in the glycine-rich regions are with residues possessing small side chains similar to glycine to prevent any steric hindrance (Kozbial and Mushegian, 2005). Another highly conserved region in plant MTs involves the acidic residues, aspartic-acid (D) and alanine (A) in the motif B [consensus (V/I/F)(A/P/E)X(A/P/G)**DA**XXX K(W/Y/F)] which spans the second beta-sheet β2 and the following turn. As evident from crystal structure studies, residues in motif A interact with the carboxypropyl portion of SAM as well as forming a water-mediated hydrogen bond with Met of SAM (Zubieta et al., 2003). Motif B conserved residues have a proven role in the positioning of SAM via hydrogen bonding to both hydroxyls of the ribose moiety. Furthermore, contact with the amino and sulfonium groups in the methionine fragment of SAM is achieved through variable residues C-terminal to β4 (Fauman et al., 1999; Vidgren et al., 1994).

The study of the crystal structure of various MTs generated both open and closed conformations (Morris and Facchini, 2019). An open conformation involves one that is not feasible for methylation. The modelling showed the apoenzyme (just the enzyme) always had an open conformation and so did the holoenzyme with just SAM/SAH. Only the ternary structure with SAM as well as the methyl acceptor produced a closed structure (**Figure 1.10**).

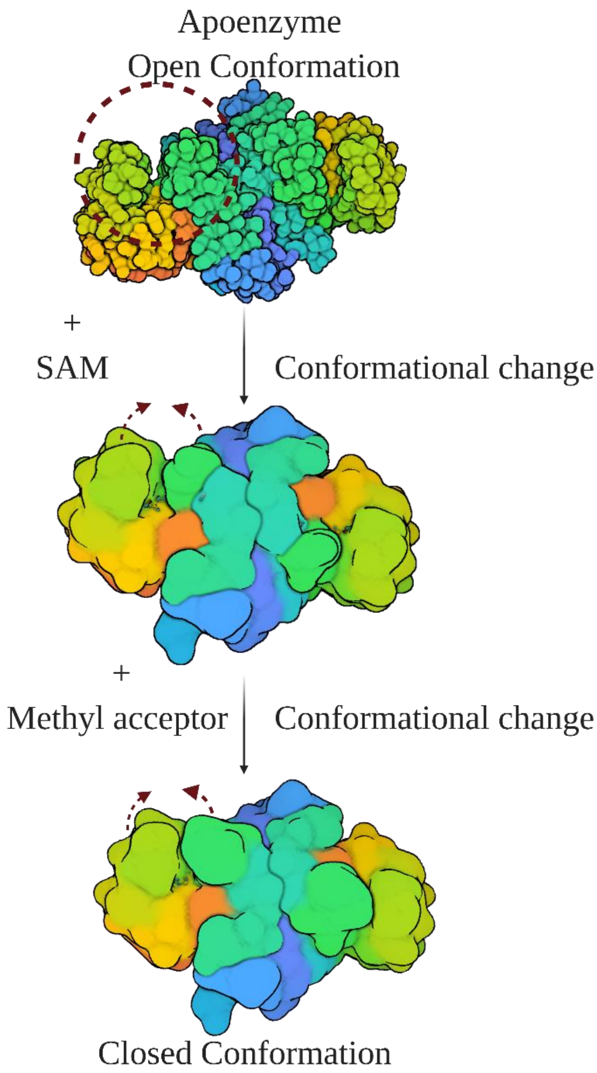


Figure 1.10: Active site assemblage facilitated by a series of conformational changes upon SAM and substrate binding. Image created with BioRender.com by K. Chandran.

The C-terminal of the SAM-binding-domain plays a central role in catalysis by contributing to the conformational changes and allosteric interaction with the substrate domain (Morris and Facchini, 2019). Employing isothermal titration calorimetry, Walker et al. (2016) confirmed the existence of sequential binding in which SAM binds before the substrate and hence, emphasized the role of the SAM-binding-domain in subsequent conformational changes (Walker et al., 2016). The comparison of apoenzyme and the SAH/SAM binary complex of subsequent crystal structures of plant MTs further supported this conclusion. Particularly, in LpOMT1 the functional binding pocket assemblage is assisted by a ≈17° inward rotation of the SAM domain (Louie et al., 2010). The homodimeric MTs facilitate the substrate entry/product exit through an open conformation and use a closed conformation for catalysis (Cooke et al., 2009; Louie et al., 2010; Robin et al., 2016; Valentic et al., 2020). Contradictory to this, SAM/SAH cofactor binding was enough to bring about a conformational change in the C-terminal domain of Tf6OMT, which left the enzyme in a closed state (Robin et al., 2016).

The interconnection between the SAM binding core and the oligomerization region is substantially flexible. In MTs a hinge movement translates through the dimer interface to bring about the required change in shape (Cooke et al., 2009). Compared to homodimer OMTs, NMTs show only minor hinge movements upon SAM binding; however, in the case of metal dependant CCoAOMT, the conformational hinge movement is not relevant as it lacks an oligomerization domain.

1.6.2 N-terminal dimerization domain: Obligatory dimerization in plant MTs

The currently determined crystal structures of plant MTs collectively confirm the influence that variations in the N-terminal region have on substrate selectivity and catalysis. Many bacterial as well as some animal MTs are known to form a dimer (Liscombe et al., 2012). The presence of a region in the N-terminal taking part as the back wall of the active site was first proposed by Zubieta et al. (2001) and has been considered so far plant-specific. Approximately the first 150-160 residues form the dimerization domain, and specific residues within this domain contribute to substrate binding at the interface (Cabry et al., 2019; Torres et al., 2016; Valentic et al., 2020; Zubieta et al., 2001).

The dimerization domain is regarded as a unique feature of plant MTs and most of the plant MTs are obligatory dimers due to the partial contribution of the dimerization region to the active site. Unlike animal OMTs, all structurally characterised plant OMTs show dimerization both in solution as well as in crystal structures (Valentic et al., 2020). In biochemically characterised plant MTs, the observation of a protein corresponding to double the theoretical mass, as determined by gel filtration or native gel electrophoresis support the occurrence of dimerization (Gauthier et al., 1996; Murfitt et al., 2000; Preisig et al., 1989). The N-terminal region involved in dimer formation, designated as a dimerization domain, has now been used as a characteristic identification for plant MTs. Although the involvement of the dimerization domain remains true for most plant OMTs and NMTs, the N-terminal dimerization domain is absent in CCoAOMT. Even though CCoAOMT forms a dimer in solution, each substrate only interacts with one monomer (Ferrer et al., 2005). Therefore, it is regarded as a novel reaction mechanism involving a hydrophobic interaction through residues 26-40, 66-78, and 212-226. Metal independency of the rest of the MTs might have influenced the new feature of catalysis involving the dimerization domain. Additionally, there is some evidence of native plant MTs showing trimer formation. *C. japonica* scoulerine 9-OMT extracted from cultured cells was shown to be a homotrimer based on gel filtration size analysis (Sato et al., 1993). Although the recombinant scoulerine 9-OMTs studied since then show dimer formation in crystals as wells as in solution (Cabry et al., 2019; Valentic et al., 2020), the chance of these MTs behaving differently *in vivo* cannot be eliminated.

The degree of dimerization varies considerably between different classes of MTs, most likely due to varying hydrophobicity of substrates. Dimerization encloses most exposed residues, preventing exposure to the solvent. In most phenolic OMT proteins, nearly 30% of the available surface area was enclosed during dimerization, while in alkaloid NMTs the enclosed area is considerably less (Morris and Facchini, 2019). The extent of the dimerization interface is only 6.5% and 6.9% in TfNMT and CjNMT, respectively (Torres et al., 2016), and 7-8% of the total surface area for each monomer in coffee NMTs is also enclosed. It is also interesting to note the presence of salt bridges predicted to be involved in dimerization in GfNMT (Lang et al., 2019). This, along with the difference in the buried area of alkaloid MTs from other classes, may be indicative of dimerization having a varying impact on different alkaloid NMTs

Recent site-directed mutation studies have shown that even subtle changes in the protein N-terminal region can impact substrate preference. Sequence analysis of OMTs catalysing the same class of compounds showed changes in the N-terminal region that may be determinant of substrate specificity (Valentic et al., 2020). For example, two enzymes of *T. flavum*, TfS9OMT and Tf6OMT, accept distinct substrates and catalyse different steps in BIA biosynthesis. The two enzymes have no large structural differences except for two distinct structural perturbations in the N-terminal end which conferred a compressed cavity for the TfS9OMT active site (Valentic et al., 2020). Mutation of a single N-terminal residue (L13W) in the first α-helix of the N-terminal region resulted in promiscuity (less prominent compared to N terminal M111A mutant) in otherwise specific TfS9OMT, which allowed a second conformation favourable for producing both 9-O methylated and 2-O methylated versions of substrate (S)-scoulerine. Furthermore, the structure superimposition of two homologous (S)-scoulerine 9-OMTs, *T. flavum* (TfS9OMT) and *P. somniferum* (PSMT1), show well-conserved active site residues even across different species (Valentic et al., 2020).

In the case of cation dependant MTs, even with the absence of an interacting dimerization domain, specificity for CoA ester derivatives is due to the proper positioning of the N-terminus and insertion loops in the vicinity of the active site (Kopycki et al., 2008). Through biochemical characterisation of a novel Mg^{2+} dependant methyltransferase from *Mesembryanthemum crystallinum* accepting flavonoids and phenylpropanoids, Vogt (2004) demonstrated that the discrepancies in tolerance of recombinant PFOMT towards acceptance of a broad spectrum of substrates, compared to the *M. crystallinum* native PFOMT (N-terminally truncated), was due to variations in the N-terminal domain (Vogt, 2004). Removal of the first 11 amino acids of recombinant full-length PFOMT reverted the enzyme's regiospecificity and kinetic properties to that of its native counterpart (Vogt, 2004). Structural elucidation of the PFOMT from *M. crystallinum* further supported this observation by showing the involvement of the N-terminal region in broad substrate specificity (Kopycki et al., 2008). Based on the PFOMT crystal structure, Kopycki and colleagues (2008) developed N-terminal hybrids (exchange of PFOMT N-terminal residues with MsCCoAOMT N-terminal residues), which had a lower affinity towards flavonoid substrates and greater selectivity for phenylpropanoids.

Another important aspect of the N-terminal domain in catalysis is the cooperative effect via hinge movement. The movement is transferred via the dimerization domain to the other protomer and has been suggested to affect selectivity and catalysis (Valentic et al., 2020). Displacement of both domains upon substrate binding was observed in the case of bacterial small molecule MTs (Cooke et al., 2009). In the case of NMTs, although the role of the dimerization domain in catalysis was well established, any direct interaction with the substrate is still unclear.

1.6.3 Active site pocket

The size and shape of the active site, as well as the composition of active site lining residues, plays a very important role in substrate selectivity. Methyl acceptor binding pockets in MTs are generally lined with hydrophobic residues and the hydrophobicity of the substrate-binding pocket influences substrate selectivity. The broader substrate acceptance range of phenylpropanoid OMTs, with a preference for compounds with neutral aldehyde and alcohol propanoid tails over those with negatively charged acids, is suggested to be due to the structure and relative hydrophobicity of the tail binding pocket (Zubieta et al., 2002). In MsCOMT, most active site constituting residues are hydrophobic, namely Leu 136, Phe 172, Phe 176, and Ala 162. (Zubieta et al., 2002). However, the propanoid-tail binding region of the active site harbours hydrophilic residues that take part in the water-mediated binding of the hydrophilic tail. The residues in the tail binding pocket are thought to be involved in the electrostatic stabilization of the anionic intermediate (Green et al., 2014). Additionally, substrate distinction between scoulerine and norcoclaurine, respectively, by TfS9OMT and Tf6OMT of *T. flavum* is also attributed to the bulkier and more hydrophobic nature of residues that line the active site of TfS9OMT (Valentic et al., 2020).

Plant phenolic MT's substrates are primarily hydrophobic. Aromatic residues lining binding sites, and a closed conformation excluding solvent create an ideal space for catalysis of these lipophilic compounds. An exception is the case of anthocyanins which are more polar. Many OMTs have been shown to take part in *O*-methylation of anthocyanins (Gomez Roldan et al., 2014; Hugueney et al., 2009; Lücker et al., 2010), but none have yet been structurally characterised. An anthocyanin OMT crystal structure is warranted to better understand the chemistry behind catalysis and active site composition.

The varying size and shape of the binding pocket of the plant MTs are used to discern between related substrates. In the ChOMT active site, the substrate rotates 180° around the carbonyl carbon and adopts two different conformations which in turn gives the B-ring of chalcone two different binding modes, but it conserves the position of the A-ring where methylation occurs (Zubieta et al., 2001). The active site of MsCOMT is relatively spacious, which is suggested to be important for *in vivo* biological activity by allowing some level of substrate promiscuity. A larger substrate-binding pocket demonstrated by alkaloid Tf6OMT was ascribed to the binding of the possible inhibitor bis-BIA (Robin et al., 2016). A compressed cavity formed by bulky pocket lining residues restricts non-specific binding (Valentic et al., 2020).

The binding of a substrate to the active site of an enzyme necessitates a shape that complements that of the active site pocket. Therefore, substrate stereochemistry is also influential in substrate selectivity. Some enzymes can distinguish between substrates with the same stereochemistry. In such cases, small changes like the lack of hydroxyl help in distinguishing between structurally related substrates. (Liu et al., 2006; Zubieta et al., 2003). In contrast to this, chemically distinct compounds with nearly identical 3D-shape are known to be allowed into the active site of MTs. For example, 2,7,4′-trihydroxyisoflavanone and 6a-hydroxymaackiain are chemically very different. Nonetheless, the 2S, 3R-stereoisomer of 2,7,4′-trihydroxyisoflavanone forms a three-dimensional conformation with the A- and C-ring perpendicular to the B-ring and is identical in 3D-architecture to the 6aR, 11aR-stereoisomer of 6a-hydroxymaackiain. This enables both to bind to the Y-shaped active site of HI4'OMT (**Figure 1.11**; Liu et al., 2006).

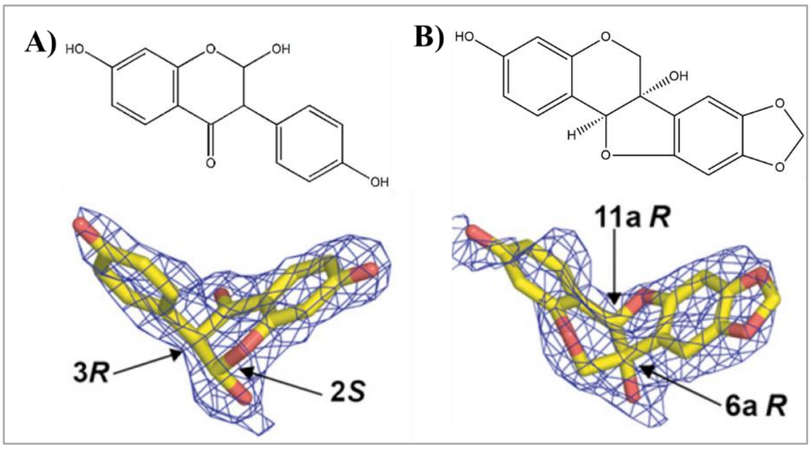


Figure 1.11: Schematic showing the chemical backbone of the 2S, 3R-stereoisomer of 2,7,4′-trihydroxyisoflavanone (A) and the 6aR, 11aR-stereoisomer of pterocarpan compound 6a-hydroxymaackiain (B) devising identical 3D shape. Some components of the figure are adapted from Liu et al., (2006) with publisher's permission.

Across the same species, active site examination of TfS9OMT and Tf6OMT revealed substrate distinction by inverse orientation of their respective substrates (S)-scoulerine and (S)-norlaudanosoline. All hydrophobic and H bond interactions were similar in both enzymes except for the M111 thioether group in TfS9OMT (A110 in Tf6OMT equivalent position) that would sterically clash with the C-ring of the (S)-norlaudanosoline making the reaction unfavourable (Cabry et al., 2019; Robin et al., 2016). This once again shows the impact a single residue can have on substrate distinction.

Plant MTs use different strategies to recognise the aromatic ring of a substrate and facilitate its positioning inside the active site pocket. To achieve this, OMTs mainly use a methionine thioether clamp formed by Met residues on both sides of the aromatic ring of the substrates (Robin et al., 2016). In carboxyl MTs and phenylpropanoid OMTs, the clamp holds the ring that presents the hydroxyl to be methylated in position. (Zubieta et al., 2003). Similarly, in flavonoid OMTs, the B-ring to be methylated is held in position by a Met thioester clamp, and the A- and C-ring, which are in the same plane, are held in place by a widely spaced clamp involving Met325 and Met-374 (Liu et al., 2006). In contrast, the NMTs of caffeine biosynthesis and a salicylic acid OMT were shown to use a unique π-π-stacking interaction for substrates aromatic ring recognition (McCarthy and McCarthy, 2007; Zhang et al., 2020; Zubieta et al., 2003).

Various functional groups present on a substrate can also influence selectivity via hydrogen bonding. Correct positioning of the substrate salicylic acid is achieved in SAMT using hydrogen bonding to the carboxylate moiety via Gln-25 and Trp151, but without interactions around the *2-OH*. Experimentally this meant the enzyme could not distinguish between benzoic acid, which only lacked the *2-OH*, and salicylic acid (Zubieta et al., 2003). Similarly, in HI4'OMT, Tyr-25 binds to the *2-OH* on the A-ring but another similar substrate (macckain) that lacks the *2-OH* could not bind to the active site (Liu et al., 2006).

Specific residues have been shown to be essential for catalysis by MTs, and experimental manipulation of these can eliminate activity due to an inability to bind to substrates. In particular, site-directed mutation studies have proved the effect even a single amino acid residue can have on the catalytic efficiency of MTs. For example, basil eugenol and chavicol OMT, which show remarkable similarity in sequence, were shown to swap their respective substrate preferences following site-directed mutagenesis of a single residue [F260S in ChavOMT and S261F in EOMT]. In all the structural studies so far (**Table 1.1**), a highly conserved histidine residue is designated as the key active site catalytic residue, but additional residues are also shown to play an associative role. These adjacent residues have been interpreted as forming a channel for methyl passage (Cabry et al., 2019). The sidechain of the designated catalytic dyad member Glu has been proposed to be the factor affecting substrate specificity. In *Cj*NMT, the Glu204 side chain was in close proximity to an alkaloid N atom, and its substitution to Ala had a detrimental effect on catalysis; however, Glu204Ala substitution in *GfNMT* retained some activity. Thus Glu appears to be less critical for catalysis. The equivalent residue in TfNMT was an alanine itself (Torres et al., 2016) and reticuline NMT has glycine at this position, further suggesting this substitution is not essential.

Substrate binding in NMTs show differences compared to OMTs. The binding site of the alkaloid substrate near the SAM binding site was found to involve residues overlapping the SAM-binding motifs (Morris and Facchini, 2019). NMTs examined to date have a typical NMT domain and an N-terminal three-helix extension, wrapped around the substrate-binding part of the dimer interface, which contributes to the NMT active site.

1.6.4 Active site gate/ 70s loop

In their work on BIA OMT, Robin and colleagues (2016) suggested the presence of gatekeeper residues controlling the active site pocket. These residues showed an orderly arrangement upon substrate binding and were coined as an active site ‘gate’. To be precise, shifting of an active site lid (formed by strands β3−β5 and helix α13) by 6.0−9.5 Å toward helix α11 in reference to the apoenzyme was observed during the SAM guided conformational change, which allowed the closure of the active site (Valentic et al., 2020). In NMTs, however, closing of the active site after substrate binding was dependent more on a gate-like loop (Morris and Facchini, 2019), which showed an orderly arrangement after SAH binding. The absence of this loop in the OMTs characterised so far suggests that this might be an NMT-specific structural feature.

Gatekeeper residues have been shown to determine substrate specificity in surface entropy reduction variant crystal structure studies of *P. somniferum* PSMT1 (Cabry et al., 2019; Lang et al., 2019; Torres et al., 2016; Walker et al., 2016; Zhao et al., 2020). In all BIA NMTs, upon binding, the loop region transforms from a disordered to an ordered helix structure, but no direct interaction with the substrate has been identified to date. The activity of BIA NMTs was shown to decrease when Glu82 of GfNMT, a residue involved in the positioning of the loop, was mutated (Lang et al., 2019). In contrast, when the adjacent residue to the loop in PavNMT was mutated, it surprisingly increased activity especially with a non-endogenous substrate (Torres et al., 2016). An alignment of various characterised (not structurally) NMT subtypes with PavNMT showed sequence conservation among them, strongly suggesting the contribution of gate/loop in functional diversity (Torres et al., 2016). Further mutagenesis studies could explore this idea.

In summary, plant MTs make use of various strategies to accommodate the vast chemical repository of secondary metabolite substrates. Subtle topological changes are enough to bring about difference in specificity. Functional evolution results from structural appendages in the Rossmann fold, as well as changes in N-terminal domains, and composition of the active site gate. It is noteworthy that altered specificity or activity is not always caused by structural changes as changes in active site lining residues and residues involving hydrogen bonding can also result in altered substrate preference.

1.7 Reaction mechanism

In general, SAM-dependent methylation occurs via an S_N2 like mechanism, where there is a direct transfer of the methyl group to a substrate via activation of a catalytic nucleophile on the substrate molecule by deprotonation (Woodard et al., 1980). The nucleophile activation mechanism differs depending on the chemistry of the target atom (Klimasauskas and Lukinavicius, 2007). Among the various phenolic OMTs reviewed here, there exists three main transmethylation mechanisms. First, the CoA ester OMTs and PFOMT follow deprotonation mediated by the divalent cation. The metal ion influences the pK_a of the substrate's phenolic

OH and promotes the loss of a proton and generates a nucleophilic phenolate anion.

The second type of transmethylation mechanism is catalysis driven by proximity and extrusion of water (desolvation). This mechanism was first suggested by Zubieta et al. (2003) after structural characterisation of SAMT. The enzyme's structural shape was such that the acceptor and donor were in proximity orienting themselves for nucleophilic substitution and desolvation from the donor-acceptor interface. The positioning of the negatively charged carboxylate anion within the proximity of the positively charged methyl of SAM is sufficient for the transmethylation of salicylic acid (Zubieta et al., 2003).

The third mechanism involves a general acid/base mediated mechanism. The majority of the metal independent OMTs and NMTs discussed in this review follow this mechanism of methyl transfer in which there exists a key active site residue that makes independent contributions to catalysis. Histidine is generally accepted as the active residue which deprotonates the substrate nitrogen or hydroxyl (Bennett et al., 2018; Zubieta et al., 2002). A lysine residue has been shown to perform the same role in the rat COMT (Schubert et al., 2003). The substitution of His residues showed decreased activity in NMTs, but almost abolished activity in OMTs (Zubieta et al., 2001; Zubieta et al., 2002). Therefore, it appears that the deprotonation mechanism is not the only thing majorly contributing to the reaction mechanism in NMTs, and the activity of NMT cannot yet be pinned to specific residues.

Adjacent residues to the key histidine have also been shown to be involved in catalysis, and unlike in OMTs, these residues are compulsory for catalysis in alkaloid NMTs. Structural studies show that in alkaloid OMTs, His-Asp acts as the catalytic dyad. In *P. somniferum* (S)-scoulerine 9-OMT, Asp255 was shown to be essential for catalysis, while His254 was deemed important, but nonessential (Cabry et al., 2019). Activity of TfS9OMT, however, is retained when either one was mutated (Valentic et al., 2020). In contrast, in alkaloid NMTs, His-Glu acts as the catalytic dyad where the carboxyl group of glutamic acid hydrogen-bonded with the imidazole ring of His to promote deprotonation. Mutation to any of the dyad residues in NMTs severely affected activity (Lang et al., 2019; Torres et al., 2016) except for the Glu207Ala mutant in CjCNMT (Bennett et al., 2018). In CjCNMT, mutation to Glu204, which is also in close proximity to the His208 catalytic residue, severely affected activity, suggesting a stronger role of Glu204 than Glu207 in catalysis (Bennett et al., 2018). An Asp residue is present adjacent to His in a majority of the isolated flavonoid and phenylpropanoid OMTs. Crystal structures analysis have shown the involvement of this Asp residue in hydrogen bonding network in enzyme-substrate complex (Green et al., 2014) but structural impacts were not confirmed. Further site-directed mutagenesis studies are required to prove the essentiality or involvement of Asp in catalytic dyad in flavonoid or phenylpropanoid OMTs.

The role of the His-Glu catalytic dyad in NMT activity is highly debated. It has been suggested that NMTs may not need residue deprotonation for catalysis (Morris and Facchini, 2019). In GfNMT, for example, the tertiary amine of the substrate stylopine had a pKa value of 5.3 which under physiological pH, would readily deprotonate and make the presence of the proposed His residue's role irrelevant. Interestingly, when His208 and Glu207 in GfNMT were mutated, the effects on catalysis were comparable to others MTs that require deprotonation indicating the importance of these residues in catalysis even in the absence of a general base mechanism (Lang et al., 2019). The involvement of the Glu residue in the NMT catalytic dyad is supported by further studies (Bennett et al., 2018; Torres et al., 2016). For example, glutamic acid mediation of substrate positioning (via direct or water-mediated H-bonding) has been suggested to electrostatically stabilize the intermediates employing a negatively charged side chain (Torres et al., 2016). In *Cj*NMT, however, Thr261 was suggested to play a role in the catalytic dyad instead of Glu residue. Also, in tea CkTcs NMT, instead of His, N9-methylation is suggested to happen through Thr31 facilitated attack of SAM (Zhang et al., 2020), indicating a presence of a possible NMT-specific catalytic mechanism that does not include a His mediated reaction. Structural studies of more purine alkaloid MTs are needed to confirm their methyltransferase mechanism.

1.8 Conclusion and future perspective

Plant secondary metabolites are extensively decorated and diversified by MTs. Methylation of various shikimate-derived compounds reviewed here, including information of the structural genes involved and biochemical characterisation of associated MTs, contributes to our understanding of plant *O*-, *N*- and *C*-MTs and provides a basis for the biosynthesis of complex plant metabolites. It is difficult to draw a clear distinction between the MTs that act on different substrate classes, like phenylpropanoids, alkaloids and flavonoids, as there are several variant MTs accepting compounds based on different chemical scaffolds. As many recombinant MTs act promiscuously *in vitro*, more research aimed at *in vitro*-*in vivo* integration, like cellular localization and expression analysis, is essential to establish an accurate picture of the varied biological roles of the MTs characterised to date.

Current classification systems cannot accommodate most of the recently isolated variant MTs.

More tertiary structure studies, together with evolutionary analyses and accurate cellular location studies of these MTs are necessary to establish a comprehensive and reliable classification system for plant small molecule OMTs. This will be aided by the increasing deposition of fully characterised plant methyltransferase sequences in databases.

Structure-function relationship studies of shikimate derived compounds and associated MTs show how structural relatedness or variations manifest in catalysis. The homodimeric MTs confine the methyl-acceptor in the inner pocket formed by one of the monomers and the N-terminal helix region of the neighbouring monomer. The functional divergence of SAM-dependent plant MTs is attributed to the capability of their flexible substrate-binding core in accepting a vast number of structurally dissimilar compounds Various strategies used by plant MTs for functional divergence include substrate promiscuity via subtle topological changes, structural appendages in the Rossmann fold, changes in N-terminal domains, and composition of the active site gate. Functional evolution is not always caused by structural changes. Subtle changes in active site lining residues and residues involving hydrogen bonding can also result in an altered substrate preference without modifying the structure. Structure comparison of MTs acting on different shikimate derived compounds show that substrate interactions in the 3D-architecture and active site residues in purine alkaloid NMTs were in consensus with the phenylpropene OMTs, whereas flavonoid OMTs generally show structural similarity with caffeic acid OMTs and alkaloid OMTs.

A large number of alkaloid MTs, especially BIA MTs, in the 3D-structure repository provide insights into the structural relatedness of proteins catalysing alkaloid compounds. This detailed representation has enabled accurate molecular modelling of BIA MTs from different plants as well as similar alkaloid MTs. A greater focus on the tertiary structure elucidation of more flavonoid and isoflavonoid MTs of different plant genera could produce information to produce more accurate and reliable homology modelling of a greater diversity of plant MTs.

1.9 Thesis significance

The research reviewed in the preceding sections depict how plant MTs contribute to the extensive chemical diversity observed in plant secondary metabolites. Compared to phenylpropanoid and alkaloid MTs, flavonoid MTs are under studied at biochemical, molecular as well s structural level.

Detailed elucidation of entire biosynthetic networks is essential for the proper utilization of plant metabolites. The groundwork for this has been laid by the biochemical investigations aimed at characterizing biosynthetic enzymes and their corresponding genes. For example, full molecular level characterisation as well as structure elucidation of BIA biosynthetic enzymes in *P. somniferum* has enabled the isolation as well as characterization of multiple enzymes in related alkaloid accumulating species which subsequently enabled *ex planta* reconstruction of entire alkaloid biosynthetic pathways and their production through bio-engineered microorganisms (Hagel and Facchini, 2013; Morris and Facchini, 2019). Recent reviews conducted by Panche et al. (2016) and Chen et al. (2018) exemplifies medicinal, cosmetic, and nutraceutical applications of modified flavonoids. Despite the therapeutic potential of methylated flavonoids, molecular level characterization of flavonoid OMTs has been done in limited genera including *Citrus*, *Solanum*, *Oriza*, *Zea*, *Triticum*, *Hordeum*, *Glycyrrhiza*, *Medicago*, *Lotus*, *Ocimum*, *Catharanthus*, *Chrysosplenium*, and *Perilla*. There is a significant lack of work on *in planta* methylation in tree species. Apart from a *Populus deltoids* OMT and two recently characterised multi-specific flavone OMTs from *Citrus species* (Itoh et al., 2016; Liu et al., 2020), no other characterised OMTs from tree species is available. No MTs have been biochemically characterised in *Eucalyptus*. Structural information available on flavonoid MTs is limited to isoflavonoid OMTs and a bifunctional Mg^{2+} dependent OMT that accepts flavonoids and phenylpropanoids.

Current understanding of *C*-methylation in plants is also very limited. *C*-methylation is often seen in flavonoids but research to date has not yet determined any CMT genes associated with it. Whilst some research has been carried out on sterol and γ-tocopherol CMTs, no sequence information is available on flavonoid CMTs in plants. The aims of this thesis seek to fill these gaps in current knowledge.

1.10 Thesis outline

This research project focuses on flavanone methylation in *Eucalyptus* species that are relatively rich in methylated flavanones and aims to isolate and characterise the enzymes responsible for flavanone methylation.

Chapter two focuses on the isolation and characterisation of the flavanone *O*-methyltransferase EnOMT1 from *Eucalyptus nitida*. Transcriptome analysis of *Eucalyptus nitida* (rich in *O*-methylated flavanones) was employed to obtain putative sequences encoding the *O*-methyltransferase responsible for the production of the modified flavanone backbone of pinocembrin. The isolated recombinant enzyme of *EnOMT1* catalysed the conversion of pinocembrin to a mono-methylated flavanone, pinostrobin.

Chapter three describes a detailed characterisation of the isolated EnOMT1. Biochemical characterisation of the enzyme showed that it can act on other classes of flavonoids but is 7-OH-specific. Two homologous genes from different *Eucalyptus* species were also isolated and characterised to compare with EnOMT1. Tertiary structure prediction by homology modelling along with *in vitro* validation by site-directed mutagenesis shed light on the active site chemistry of the flavanone 7-*O*-methyltransferase.

Chapter four focuses on the strategies for the identification of flavanone *C*-methylation in *Eucalyptus*. A species-wide screening was performed to identify the presence and composition of various methylated compounds in the foliar tissue particularly looking for a species that is rich in *C*-methylated flavanones. Using a comprehensive bioinformatics analysis of expressed transcripts in various *Eucalyptus sp*. the genetic basis and an evolutionary relationship were analysed and possible candidates for *C*-methyltransferase in *Eucalyptus diversifolia* were narrowed down. The chapter also discusses the attempts to uncover the ecological aspects of flavanone methylation. Limited laboratory access due to COVID-19 precluded some of the planned experiments (like native *C*-methyltransferase analysis and elicitation of callus culture) and constrained the chapter to a theoretical approach.

Chapter five summarizes the findings of the entire research project and discusses possible future directions for flavonoid methylation research in *Eucalyptus* as well as in other plant species.

CHAPTER-2

MOLECULAR CLONING AND ISOLATION OF FLAVANONE 7-*O*-METHYLTRANSFERASE FROM *EUCALYPTUS*

2.1 Introduction

The plant kingdom abounds with a bewildering array of natural products called secondary metabolites. The term 'secondary' is a misnomer dating back to when such compounds were widely considered metabolic waste products. They have subsequently been shown to be indispensable components in the survival strategy of plants (Harborne, 2001). Tens of thousands of plant secondary metabolites have been structurally elucidated and classified into groups based on shared chemical features. The largest groups are the terpenes, alkaloids and phenolics and these are further sub-divided into classes based on specific structural arrangements. For example, within the phenolic group, one of the major classes is the flavonoids – compounds that are ubiquitous in plants and fulfill many important physiological roles (Panche et al., 2016).

Flavonoid is the generalized term given to compounds whose basic structure includes a 15-carbon skeleton derived from the precursor compounds malonyl CoA and *p*-coumaroyl CoA. The general structure of flavonoids involves two benzene rings bridged by a heterocyclic pyran ring. Conventionally the rings are named A, B and C (**Figure 2.1**). A large diversity of flavonoids structures are found naturally in plants and they are divided into nine major classes: flavones, flavanones, flavanols, anthocyanins, dihydroflavonols, flavonols, isoflavonoids, neoflavonoids, aurones and chalcones, based on differences in bond saturation, hydroxylation, degree of oxidation and ring position on the basic backbone (Kumar and Pandey, 2013). The general structures of the major classes of flavonoids are shown in **Figure 2.1**. These classes can be further divided into sub-classes by the presence of additional chemical moieties such as methyl, hydroxyl and glucosyl groups. The occasional addition of aromatic or aliphatic acids, sulphate, isoprenyl or phenyl groups to the compounds escalates the number of naturally occurring flavonoids to more than 9000 (lwashina, 2000).

The role of flavonoids in plants includes protection from various biotic (pathogens, herbivores) and abiotic (UV-B, drought, salt, frost) stresses (Tucker et al., 2010), maintenance of redox state in cells, auxin transport regulation (Mathesius et al., 1998), pollen tube growth and pollen functionality control (Mo et al., 1992; Taylor and Jorgensen, 1992) and root nodule development (Hirsch et al., 1989; Mathesius et al., 1998; Zhang et al., 2009). Flavonoids also act as antioxidants, colour and taste pigments, attractants, growth stimulators of certain insects (Thoison et al., 2004) and precursors of inducible compounds (Dixon et al., 1983).

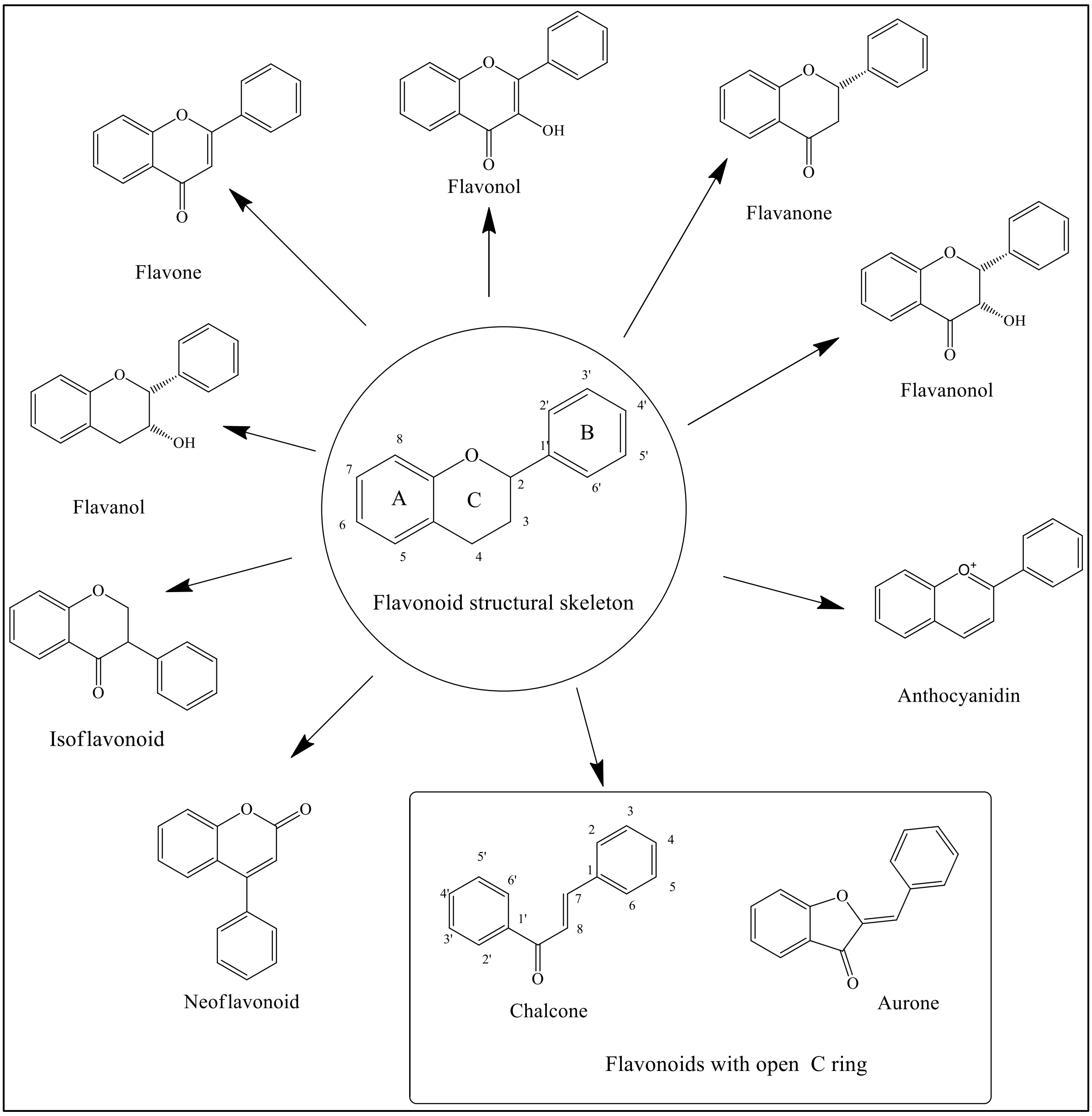


Figure 2.1: Flavonoid subclasses based on the presence of a bond between carbon 3 and 4 of the C-ring, C-ring substitution pattern and degree of oxidation. The numbering of carbons is different for flavonoids with an open C ring structure as indicated.

Flavonoid compounds are well recognized for their diverse health-promoting properties in humans. They are very effective antioxidants due to their excellent free radical scavenging activity. This property is often associated with the protective effect of flavonoids in coronary heart disease, diabetes, and stroke (D'Andrea, 2015). The pharmacological activity of compounds from major flavonoid subclasses, including flavonols, flavones and anthocyanidins are currently being investigated. One minor group of flavonoids – the flavanones – have recently attracted a great deal of attention due to their medicinal efficacy as hepatoprotective, antiproliferative (Gao et al., 2015) antioxidant (Bedane et al., 2016; Miler et al., 2016), anti-inflammatory (Sudsai et al., 2014), antimutagenic (Atun et al., 2013) and neuroprotective agents (Kapoor, 2013).

Natural product diversity is attributed to various tailoring mechanisms such us glycosylation, methylation, and acylation which contribute to a diverse set of possible structures. For example, the flavonoid core has eleven different positions where a hydroxyl group can attach. With this arises numerous possible combinations of methyl or glycosyl group additions. In addition to this, various modification mechanisms, such as acylation, and prenylation, result in a wide variety of flavonoid compounds with distinct chemical signatures. In plants, flavonoids occur in three main forms: aglycones, glycosides or methylated variants of the first two forms. Enzymatic glycosylation of flavonoids is performed by glycosyltransferases (GTs) and glycoside hydrolases (GHs) (Lombard et al., 2013), collectively designated as Carbohydrate-Active enZymes (CAZymes). Enzymatic methylation of flavonoids is performed by MTs [EC 2.1.1.6.x], which involves the transfer of a methyl group from *S*-adenosyl-methionine (SAM) to oxygen or carbon to form *O*-methylated or *C*-methylated forms, respectively. These modifications reduce the reactivity of their hydroxyl groups and increase their antimicrobial and/ or antiviral activity (Ibrahim et al., 1998; Koirala et al., 2016; Tuchinda et al., 2002).

Flavonoids have various promising applications as dietary supplements, functional food, and cosmetic ingredients. Most of the current literature on flavonoids pays much attention to pharmacological aspects. However, many plant compounds that show activity *in vitro* fail to perform the same *in vivo* due to low bioavailability (Otake et al., 2002; Walle et al., 2007) and lack of membrane transport (Cao et al., 2013). Modifications such as methylation and glycosylation help to ameliorate these problems by increasing metabolic stability, bioavailability, and bioactivity (Walle et al., 2007), and enhancing membrane transport (Cao et al., 2013). This is exemplified in the work undertaken by Ta and Walle (2007) where methylated forms of flavone chrysin showed high transport capacity, intestinal absorption, and metabolite resistance in a human intestinal transport model. Likewise, 5,7-dimethoxyflavone and 5,7,4′-trimethoxyflavone inhibited DNA synthesis in human oral (tongue) squamous cell carcinoma-9 cells 10 times more effectively than their unmethylated analogues chrysin and apigenin, respectively (Walle et al., 2007). In human umbilical vein endothelial cells, even though both methylated and unmethylated flavones showed similar antioxidant activity, 7-hydroxy,8-

methoxy flavone showed a longer-term effect than 7,8- dihydroxyflavone without affecting cell viability (Koirala et al., 2014). Furthermore, methylated flavonoid compounds of rhamnetin and cirsiliol were demonstrated to enhance the radiosensitivity of human lung cancer cell lines (Kang et al., 2013) and lower triacylglycerols in rats fed on a cholesterol-enriched diet (Igarashi and Ohmuma, 1995). Collectively, these studies highlight the importance of methylation as a modification technique for drug delivery as well as the commercial application of flavonoid compounds.

Despite the discovery of *O*- and *C*-methylated flavonoid compounds from various plant species, very little is known about the underlying mechanism of *in planta* methylation. A detailed insight into this aspect is essential. For this, we need to understand both *C*- and *O*-methyltransferase enzymatic machinery. A considerable amount of literature has been published on isolation and characterization of *O*-methyltransferases (OMTs) from different plant sources. To date, more than hundred plant OMT sequences have been elucidated and can be retrieved from sequence databases. Nonetheless, even for a similar class of flavonoids, different MTs may be responsible for the production of a range of methylated flavonoids, and these enzymes differ substantially in their primary sequence except for five highly conserved motifs regions. As a result, using the sequence information across distantly related species remains an imprecise approach for identification of MTs, limiting further progress in the research field. Furthermore, very little information can be found in the literature on the question of regiospecificity and substrate specificity of flavonoid MTs. Many OMTs have been shown to possess narrow substrate specificity (Ibrahim et al., 1998) which is dependent on their stereospecificity as well as the pattern of hydroxyl group substitution on substrates. Studies have shown that even a single amino acid change can result in modified substrate specificity (Gang et al., 2002).

Eucalyptus is an excellent candidate species for the study of enzymes involved in flavonoid methylation. *Eucalyptus* is a genus of evergreen hardwood flowering trees belonging to the family Myrtaceae in which a disproportionately high number of flavanones have been detected (Goodger et al., 2019; Goodger et al., 2016; Marsh et al., 2019). The trees are the iconic and abundant native flora of Australia and are ecologically as well as economically significant. More than 800 species occur throughout the continent except in the central arid zone and the rainforest region (Ghisalberti, 1996). These trees are commercially exploited primarily for their essential oil content and as a source of timber, wood pulp, gum, resins, and honey. The research on foliar secretory cavities (glands) showed that a variety of different non-volatile components constitute up to 50% of the secretory gland volume (Goodger et al., 2016). In some species, the non-volatiles are comprised of several unsubstituted B-ring flavanones, including non-methylated pinocembrin, mono-methylated pinostrobin and alpinetin and di-methylated dimethylpinocembrin, in high abundance. Therefore, *Eucalyptus* has been proposed as a potential commercial source of methylated flavonoids (Goodger et al., 2019; Goodger et al., 2016). Some *Eucalyptus* species were also reported to contain high levels of *C*-methylated flavonoids (Goodger et al., 2016; Marsh et al., 2019). The presence of both *C*- and *O*-methylated compounds in abundance makes the genus *Eucalyptus* an excellent candidate for the exploration of *C*- and *O*-methylation mechanisms. To date, few studies have investigated general methylation processes in these trees, and none have specifically addressed flavanone methylation. Only the structural gene of lignin OMT in *Eucalyptus gunnii* (Poeydomenge et al., 1994) and several other OMTs in *E. grandis* (Gion et al., 2000; Myburg et al., 2014) have been identified by sequence homology as well as gene prediction methods, but none have yet been characterized or their substrate specificity assessed. Some major enzymes of flavonoid biosynthesis have been studied in *Eucalyptus* (**Figure 2.2**) (Kanehisa and Goto, 2000). Nonetheless, further research is required to investigate the order in which the methylation steps occur on flavanones, and whether these compounds are modified at an earlier or later stage of biosynthesis.

The research described in this chapter aims to isolate the *O*-methyltransferase enzyme(s) from *Eucalyptus* subgenus *Eucalyptus* that are responsible for the differential accumulation of methylated flavanones. This will be accomplished using transcriptomic sequencing and analysis techniques to enable isolation of poorly understood genes in *Eucalyptus*. Specifically, the chapter reports on the functional expression and biochemical characterization of the recombinant EnOMT1 responsible for the production of 7-*O*-methylated flavanone, pinostrobin in *Eucalyptus nitida*.

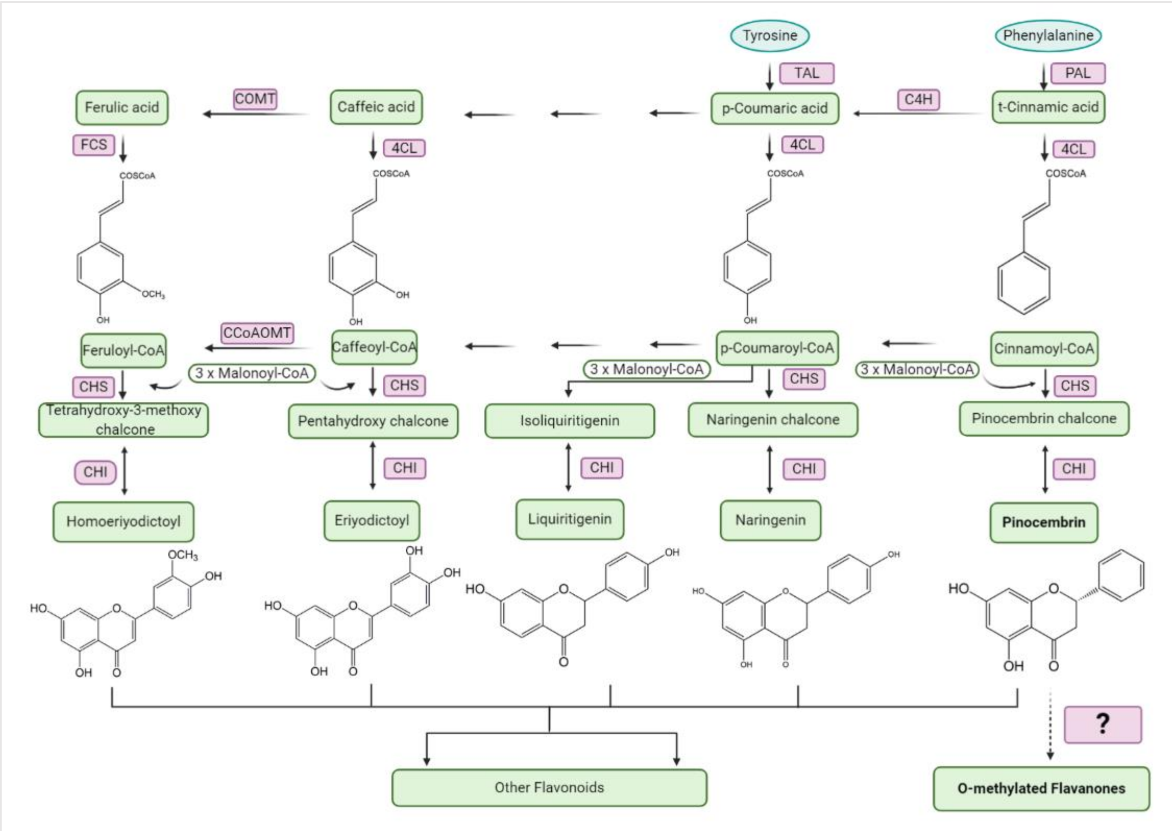


Figure 2.2: Schematic showing biosynthesis of different flavanone compounds across various Eucalyptus species. PAL- phenylalanine ammonia-lyase; TAL- tyrosine ammonia-lyase; C4H- trans-cinnamate 4 hydroxylase; COMT- caffeic acid *O*-methyltransferase; 4CL- 4-coumarate CoA ligase; FCS-feruloyl CoA synthase; CHS- chalcone synthase; CHI- chalcone isomerase (Kanehisa and Goto, 2000).

2.2 Materials and methods

2.2.1 Materials

Plant materials: All plant materials used in this study were collected from various locations in Victoria, Australia. This study used two *Eucalyptus* species belonging to subgenus *Eucalyptus*, namely, *E. nitida* known to contain 7-*O*-methylated pinostrobin (Goodger et al., 2016) (hereafter termed *O-methylated species 1*) and a species known to contain only pinocembrin (unpublished observations; the name of this species is withheld due to commercial considerations and is hereafter referred to as *non-methylated species*). *Non-methylated species* leaf tips were collected from one-year-old plants grown in the System Garden, The University of Melbourne Parkville campus, whereas *O-methylated species 1* leaf samples were collected from plants grown at the University of Melbourne Dookie campus. Seeds of *O-methylated species 1* were purchased from Nindethana seeds (Albany, Western Australia) and were propagated in the University of

Melbourne glasshouse and the seedlings planted at the Dookie campus in November 2016. *Non-methylated species* seeds were collected from a natural population (collection details are withheld due to commercial considerations) and sown in October 2016 and planted at System Garden in June 2017.

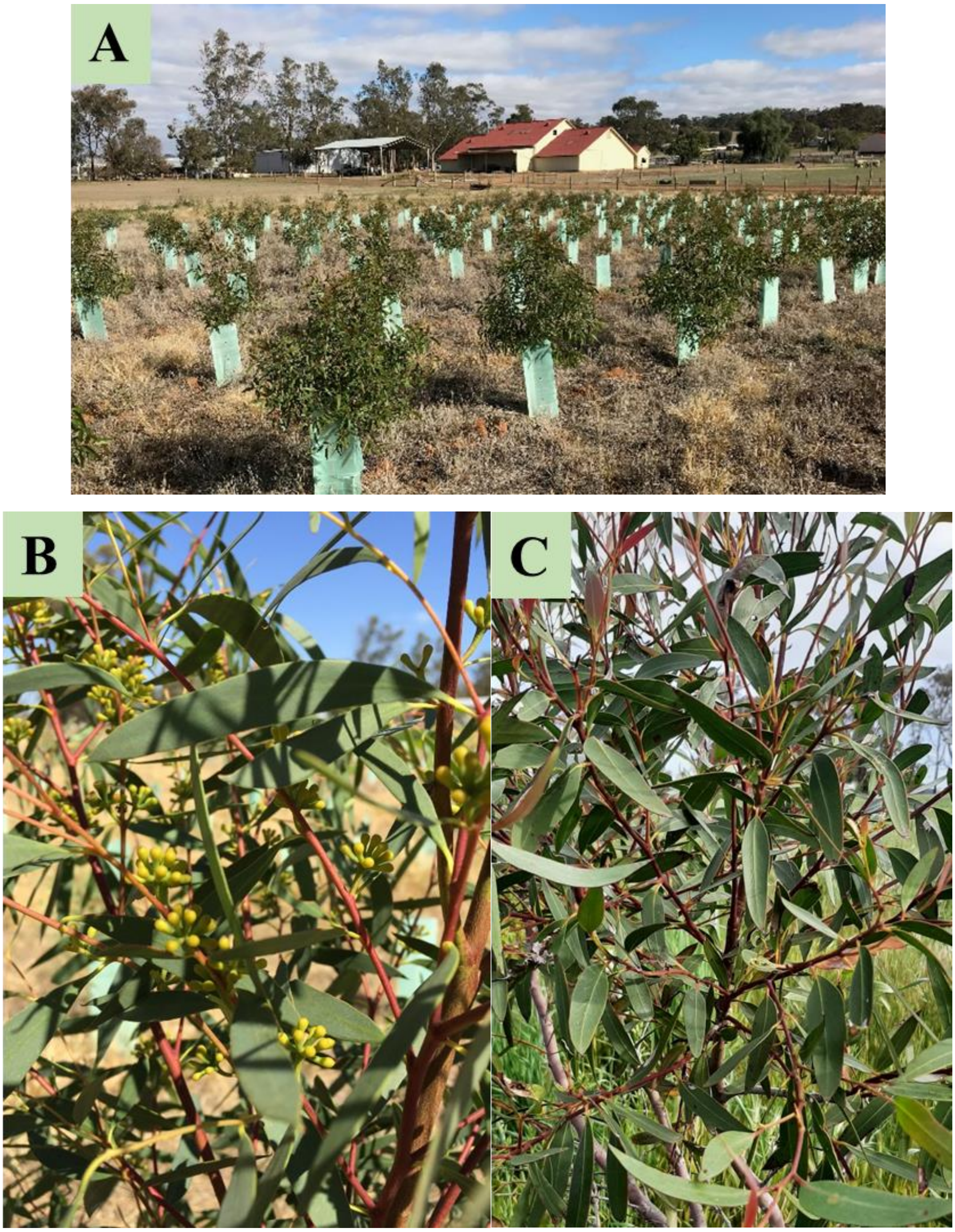


Figure 2.3: Plant material used for the study. A) The University of Melbourne Dookie campus *Eucalyptus* plot; B) *O*-methylated species 1; C) Non-methylated species.

Chemicals*:* All chemicals used were of analytical grade unless specified otherwise. Milli-Q purified water was used for HPLC analyses. All other reagents and materials used in molecular biology experiments were purchased from Thermo Fisher Scientific Australia Pty Ltd, if not specifically stated otherwise.

Bacterial culture media: Luria-Bertani (LB) broth [1% (w/v) tryptone, 0.5% (w/v) yeast extract, 171 mM NaCl, pH 7.2] was employed as a bacterial culture medium. When required, antibiotics ampicillin (Amp) or spectinomycin (Spec) were added at a final concentration of 50 µg/mL. Solid media was prepared by adding 1.5% (w/v) bacteriological agar (Scientifix, VIC, Australia). All media were sterilized by autoclaving at 121 °C, 15 psi for 20 min.

2.2.2 Flavanone profiling of *Eucalyptus* species

Flavanones in *Eucalyptus* leaves were chromatographed and quantified on a reversed-phase HPLC system. A fresh whole leaf (mature/ young) was collected from respective *Eucalyptus* trees and dried for 1-2 days in a hot air oven at 55 °C to remove most of the volatiles. Once dried, the leaf samples were finely ground to powder in a pestle and mortar using liquid nitrogen. The powder was weighed (approx. 0.1 g) and extracted with 4 mL acetonitrile (100%, HPLC grade) in a 4 mL glass vial. Weight of each empty vials were noted down to calculate leaf dry weight later. Samples were stirred on an Intelli mixer (99 rpm, room temperature) for 24 h. The extract was then centrifuged (5000 × *g*, 5 min) to remove the leaf residue and the supernatant was filtered through a syringe filter (PTFE membrane, 0.45 µm pore size) before HPLC analysis. The vial was left in the fume hood to evaporate the residual solvent and weighed accurately to calculate the leaf dry weight .

Chromatographic separation of the reaction mixture was achieved on an Agilent 1220 Infinity High- Performance Liquid Chromatography (HPLC) system with a Gemini C18 analytical column (5 µm, 150 × 4.6 mm; Phenomenex), Degasser G1311 A, pumpL7100, autosampler L-2700 and a photodiode array with UV detection (283 nm). A gradient elution method with water (solvent A) and acetonitrile (solvent B) at a flow rate of 1 mL min^{-1} were employed. The column temperature was maintained at 28 ± 5 °C and the injected volume varied between 0.1 µL and 100 µL according to the sample′s composition. The elution method used was as follows: linear increase from 30% to 50% solvent B in 7 min, thereafter to 95% in the next 7 min, isocratic elution with 95% solvent B for 6 min, then a linear gradient of 95% to 100% solvent B over 15 sec followed by 100% solvent B for 3.75 min. The column was returned to initial equilibrium with a linear decrease to 30% B over the next 30 seconds and remained the same for the last 4 min making a total run time of 29 min. Further analysis of the data was done through the associated EZ Chrom Elite software (Agilent technologies).

Different flavanones in various *Eucalyptus* species were identified by comparing their retention time to that of the known standards (pinocembrin, pinostrobin, alpinetin, dimethyl pinocembrin) and peak integrals quantified by comparison with calibration series of standard compounds.

2.2.3 RNA extraction

In preparation for RNA isolation, freshly collected *non-methylated species* leaf tips were immediately frozen in liquid nitrogen, while *O-methylated species 1* leaf tips were collected and stored in RNAlater solution until extraction. RNA extraction methods described by Untergasser (2008) was utilised with some modifications. All the materials used for RNA extraction were wiped with RNaseZAP (Sigma) before use, to inactivate RNases. Fresh *Eucalyptus* leaf tips (approx. 300 mg each), powdered in liquid nitrogen, were immediately suspended in 3 mL extraction buffer [2% (w/v) CTAB, 1.4M NaCl, 20mM EDTA, 100mM TRIS-HCL pH 8, 2% (w/v) Polyvinylpyrrolidone (PVP-40)]. The total cell extract was then mixed with an equal volume of chloroform and centrifuged at 16000 × *g* for 2 min to separate the phases. Proteins, lipids, and other cell debris from the aqueous extract were selectively partitioned to the organic phase. The supernatant was collected and further mixed with an equal volume of chloroform and centrifuged at full speed for 2 min to separate the residual contaminants. To precipitate the nucleic acids, the supernatant was collected without disturbing the white interphase and mixed with an equal amount of 100% isopropanol. Once precipitated the genomic material was washed with 500 µL of 70% ethanol and then resuspended in 95 µL of RNase free water. Further RNA purification was achieved using RNeasy mini spin column from RNeasy Plant mini kit (QIAGEN) following the manufacturer′s instruction. Briefly, the resuspended genomic materials along with 345 µL of buffer RLC were transferred to the mini spin pink column, 200 µL 100% ethanol was added to provide appropriate binding conditions, centrifuged at 8000 × *g* for 30 seconds, and the flow-through was discarded. The genomic material, now bound to the column, was then washed with 350 µL of wash buffer RW1, and DNA contamination in the extract was removed by an on-column treatment with DNase I (Bioline, Australia). 10 µL of DNase1 was mixed with 90 µL DNase reaction buffer and was applied on the mini spin column, incubated for 10 min at room temperature for the reaction to proceed. The column was then washed three times and total RNA was eluted with 30 µL RNase free water. In the occasional absence of an on-column DNase treatment, the contaminating genomic DNA in the eluted RNA preparation was removed by incubating with an amplification grade DNase 1 (Invitrogen, Australia) as follows: 30 µL purified RNA, 3 µL of DNase 1 (1U/ µL) and 3 µL of 10X buffer was mixed and incubated at room temperature for 15 min; the reaction was then stopped by incubation with 2.5 mM EDTA at 65 °C for 10 min.

The quality and quantity of purified RNA were analysed using Agilent 2200 TapeStation system (Agilent Technologies, US). The sample was prepared by mixing 5 µL RNA ScreenTape sample buffer with 1 µL RNA. The mixture was vortexed for 1 min, centrifuged and denatured at 72 °C for 3 min. The heat-treated sample was then cooled on ice for 2 min and was again centrifuged to collect solution at the bottom of the tube. The samples were then subject to electrophoresis following the 2200 TapeStation Controller Software Eukaryotic RNA protocol, and the results were analysed using TapeStation Analysis Software, providing an assessment of RNA quantity and integrity (RNA integrity number - RIN). RNA samples prepared for RNA-Seq analysis were applied to GenTegra™ RNA tubes and dried using a rotational vacuum concentrator (RVC 2-33, CHRiST) before sending for transcriptome sequencing.

2.2.4 Transcriptome Sequencing

Purified total RNA was sent to Novogene (https://en.novogene.com) for transcriptome sequencing using RAN-Seq technology (Wang et al., 2009). The workflow of the sequencing process followed by Novogene is depicted in **Figure 2.4**. The NEBNext® Ultra™ II RNA Library Prep Kit for Illumina® (New England Biolab Inc.) was employed to convert RNA into high-quality non-directional libraries for next-generation sequencing on the Illumina® platform. Quality control (QC) was performed at each major step of the procedure to ensure the reliability of data. In short, samples that passed initial quality control were subjected to mRNA enrichment using oligo(dT) beads before cDNA synthesis and library construction. Qubit 2.0 fluorometer (Life Technologies Australia Pty Ltd. Australia) aided quantification of the cDNA library, following which, the library was fed into the Illumina HiSeq 4000 platform for sequencing. The original data from Illumina platform were converted to Sequenced Reads by base calling method, producing millions of 150 bp paired-end reads. Clean reads, after quality control, were *de novo* assembled for transcriptome reconstruction with the use of the bioinformatic platform Trinity (Grabherr et al., 2011).

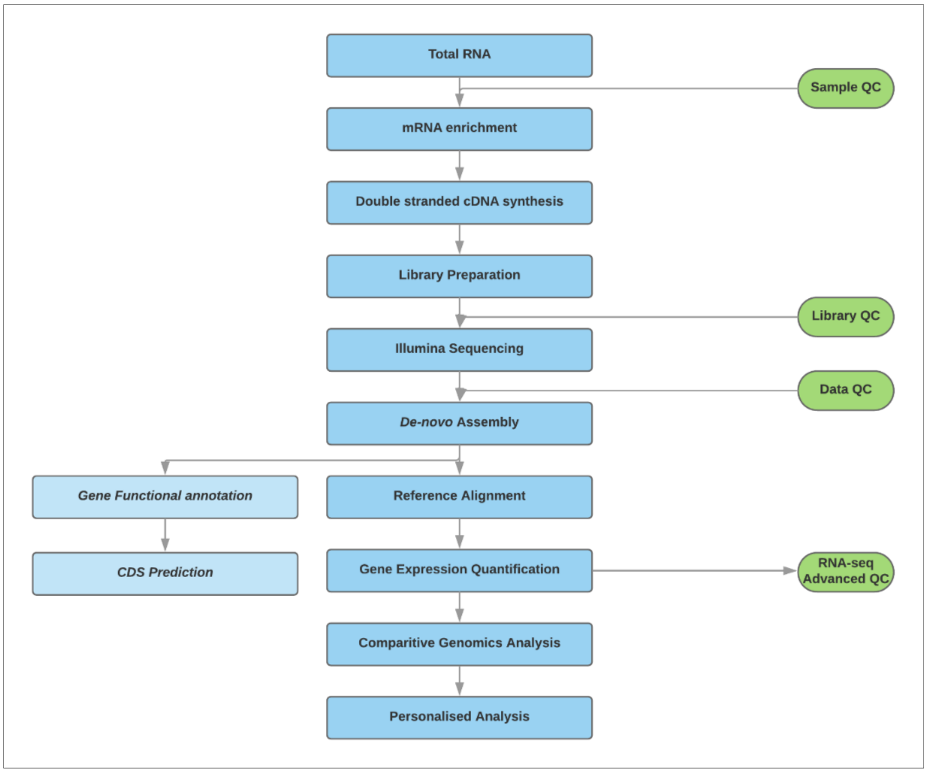


Figure 2.4: General work-flow diagram of transcriptome sequencing.

The reads, having been multi-mapped to the transcriptome, were filtered, and the resulting contigs were subjected to Hierarchical Clustering by Corset, version 1.05 (Davidson and Oshlack, 2014). The longest transcripts of each cluster were then selected as unigene clusters, and their functional annotation was performed through reference to several databases: namely Nr (NCBI non-redundant protein sequences), Nt (NCBI nucleotide sequences), Swiss-Prot, KOG (euKaryotic Orthologous Groups), Pfam (Protein family), KEGG (Kyoto Encyclopedia of Genes and Genome) and GO (Gene Ontology). Furthermore, the gene expression level was analysed and estimated as FPKM value (expected number of Fragments Per Kilobase of transcript sequence per Million base pairs sequenced) using RSEM (Li and Dewey, 2011). The annotated unigene clusters were later analysed using Geneious Prime software for identification of putative flavanone *O*-methyltransferase genes

2.2.5 Oligonucleotides

Synthetic oligonucleotides used in sequencing and polymerase chain reactions were designed and analysed using an online OligoAnalyzer® tool (Integrated DNA Technologies, Inc., US) and were synthesised by Sigma-Aldrich (New South Wales, Australia). All the oligonucleotide sequences used in this study are listed in **Table 2.1.**

Table 2.1: List of oligonucleotide primers used in the study

Oligonucleotides				Annealing Temp (°C)
Oligonucleotides used for cloning pDEST17				
EnOMT1 F	5'-ATGGATCCTCTCGATGAGACGTTAAG-3'	*EnOMT1 R*	5'-TCACAGCGGGAAAGCTTCAA-3'	64.9
EnOMT2 F	5'-ATGAGCGAGAGCGAGATCTTC-3'	*EnOMT2 R*	5'-TCAAGGATAAACCTCAATGAGAG-3'	60.9
EnOMT3 F	5'-ATGGATCCTAGCGAGACCC-3'	*EnOMT3 R*	5'-TCAAGGATAAACCTCAATGAGAG-3'	58.3
EnOMT4 F	5'-ATGGAGCGAGGCTGGGACAAG-3'	*EnOMT4 R*	5'-CTAATAGAGGCGCCTGCAAAG-3'	64
EnOMT5 F	5'-ATGGAGCGAGGCTGGGACAAG-3'	*EnOMT5 R*	5'-CTAAAAGAGGCGCCTACAAAGC-3'	64.6
EnOMT6 F	5'-ATGGCCGACAACCAAGAACG-3'	*EnOMT6 R*	5'-GGAAAGACTTAATGACCTGCC-3'	61.7
EnOMT7 F	5'-ATGGAATCGCTTGATGAAACTG-3'	*EnOMT7 R*	5'-TCACAGTGGATAGGCTTCAATG-3'	62
Oligonucleotides used for pHUE cloning				
EnOMT1 F Kpn	5'-GCGGTACCATGGATCCTCTCGATGAGAC-3'	*EnOMT1 R NotI*	5'-TAGCGGCCGCTCACAGCGGGAAAGCTTCAA-3'	71.9
EnOMT2 F Kpn	5'-AGCCGGTACCATGAGCGAGAGCGAGATCTTC-3'	*EnOMT2 R NotI*	5'-ACCTGCGGCCGCTCAAGGATAACCTCAATG-3'	72
EnOMT3 F Eco	5'-TTGAGAATTCATGGATCCTAGCGAGACCC-3'	*EnOMT3 R Hind*	5'-AGCGAAGCTTTCACTTCTTTTTGTAAACTC-3'	69.1
EnOMT4 F Eco	5'-GTGAGAATTCATGGAGCGAGGCTGGGACAAGG-3'	*EnOMT4 R Hind*	5'-AGGTAAGCTTCTAATAGAGGCGCCTGCAA-3'	71.8
EnOMT7 F Eco	5'-GCGGTACCATGGAATCGCTTGATGAA-3'	*EnOMT7 R Hind*	5'-TAGCGGCCGCTCACAGTGGATAGGCTTC-3'	72
Oligonucleotides used for qRT-PCR				
qRef4F†	5'-TACAAGCGCTGTTGATATGTGGGC-3'	*qRef4R*	5'-TTGCCAATGAGGCGGATTCACAAG-3'	68.3
qOMT1AF	5'-GCATGGCGTGCTTGAATAAG-3'	*qOMT1AR*	5'-GTGTGGGTACAACCTGACTATC-3'	63
qOMT1BF	5'-CAGACGGTGATGGGCTATTC-3'	*qOMT1BR*	5'-TTCTTCCATTCCGGTTCACTC-3'	63.1
Oligonucleotides used for CMT				
EdCMT1 F Eco	5'-GCGAATTCATGGATCCTAGCGAGACCCAATC-3'	*EdCMT1 R Hind*	5'-TGAAGCTTTCACTTCTTTTTGTGAAGCTCCG-3'	70.3
EdCMT2 F Eco	GAGAGAATTCATGGATTCARCAGGATCCRAG-3'	*EdCMT2R Hind*	5'-CACCAAGCTTTCACTTCTTTTTGTGAAACTC-3'	69.3
EdCMT3 F Eco	5'-GCGAATTCAATAGAAGCATGAGTGCAC-3'	*EdCMT3 R Hind*	5'-TGAAGCTTTCACATCCTTTTGTGTAACTCG-3'	67.9
EdCMT4 F Eco	TAGCGAATTCATGGATCCTAGCGAGACCC-3'	*EdCMT4 R Hind*	5'-CACCAAGCTTTCACTTCTTTTTGTGAAACTC-3'	68.2
EdCMT5 F Eco	ATGAGAATTCATGGACTCGGTGGTCTGTGG-3'	*EdCMT5 RHind*	5'-ACGTAAGCTTCTAAGAATCTGCCAGGGGTTG-3'	72
EdCMT6 F Eco	GTGAGAATTCATGTCGAAAGCAGGAGCGATGG-3'	*EdCMT6 R Hind*	5'-AGTTAAGCTTTCACTGATTCTCAGATAGGG-3'	68.1

†Ref4- putative cyclin-dependent protein kinases.

2.2.6 cDNA synthesis and Polymerase Chain Reaction

RNA harvested from *Eucalyptus* leaf tips were used as the template for first-strand complementary DNA synthesis using the SuperScript™ III Reverse Transcriptase (Invitrogen, Australia, Australia), in a total volume of 20 µL. In brief, the following components were added to a nuclease-free microcentrifuge tube: 1 µL of oligo(dT)$_{20}$ (50 µM); 11µL total RNA (10pg-5µg); 1 µL 10 mM dNTP mix. The mixture was heated to 65 °C for 5 minutes then incubated on ice for at least 1 minute. The contents of the tube, collected by brief centrifugation (5000 × *g*), were added to 4 µL 5X First-Strand Buffer [250 mM Tris-HCl (pH 8.3), 375 mM KCl, 15 mM $MgCl_2$], 1 µL 0.1 M DTT, 1 µL RNaseOUT™ Recombinant RNase Inhibitor (40 units/µL), 1 µL of SuperScript™ III RT (200 units/µL). The solution was mixed by pipetting gently up and down and was subsequently incubated at 25 °C for 10 minutes, 37 °C for 45 minutes, 42 °C for 45 minutes and finally, the reaction was inactivated by heating at 70 °C for 15 min and cooled to 4°C.

PCR amplifications were performed using Phusion® High-Fidelity PCR kit in a MyCycler™ Thermal Cycler System (Bio-Rad, California, US). For PCR amplification, 1 µL of the template (50 ng cDNA or 10 ng of plasmid) was combined with 19 µL of PCR master mix [11.8 µL of water, 4 µL of 5X Phusion HF Buffer (containing 7.5 mM MgCl2), 0.4 µL of 10 mM dNTPs, 1 µL each of 0.5 µM forward and reverse primer, 0.6 µL DMSO, 0.2 µL Phusion DNA polymerase (2U/ µL)]. Initial denaturation was achieved by preincubation of samples at 98 °C for 30 seconds followed by template amplification through 30 cycles of varying parameters including, denaturation at 98 °C for 10 s; annealing of primers at 45 - 72°C (temperature dependent on the melting temperature Tm of the primers) for 30 s; and an extension at 72 °C for 40 s. A final extension was performed at 72 °C for 10 min.

2.2.7 Quantitative PCR

To validate the differential expression pattern between *Eucalyptus* species determined by FPKM values of the transcriptome sequencing data, quantitative real-time PCR was performed on potential *Eucalyptus* OMT candidate genes. A gene in *Eucalyptus*, encoding protein highly similar to Eucons04 (putatively encodes a protein highly similar to cyclin-dependent protein kinases, qRef4) was selected as a reference gene based on de Oliveira et al. (2011). Corresponding primers for EnOMT1 gene (**Table 2.1**) were designed using the online PrimerQuest tool from Integrated DNA Technologies (https://sg.idtdna.com/Primerquest/Home/Index). To monitor chromosomal DNA contamination, mock cDNA synthesis reactions having no reverse transcriptase were employed as negative controls in subsequent amplification and a set of primers were designed flanking the intron-exon junction in the target gene transcript wherever possible. qPCR reactions were carried out using QuantiNova™ SYBR® Green PCR kit following manufactures instructions. Briefly, a reaction mix was set up containing 5 µL dye, 0.5 µL each of forward and reverse primer, 2.5 µL of 1/10 diluted cDNA and made up to 10 µL with Milli-Q water. Samples were loaded into individual wells in a 384 well qPCR plate along with negative controls containing no cDNA and analysed using CFX384 Touch Real-Time PCR detection system (Bio-Rad). Initial heat activation was set to 95 °C for 2 min followed by thirty cycles of denaturation (95 °C for 5 s) and combined annealing or extension (60 °C for 10 s). The real-time cycler built-in programme was used for melting curve analysis and specificity of the PCR products were further checked by agarose gel electrophoresis. Further analysis and interpretation of experimental results were done utilising the accompanying CFX Maestro Software (Bio-Rad).

2.2.8 Resolution and recovery of DNA fragments by gel electrophoresis

Size separation of DNA fragments (i.e., amplified PCR products or digested plasmid DNA) was performed via gel electrophoresis on a 1% (w/v) agarose gel containing ethidium bromide, at 120 V in 1X TAE buffer [40 mM Tris, 0.114% (v/v) glacial acetic acid, 1 mM EDTA, pH 8.3]. DNA was mixed with loading buffer containing bromophenol blue and loaded in the wells. The size of amplified or digested products was determined by comparison with HyperLadder, 1 kb (Bioline). To visualise the DNA, the gel was placed on a Gel Doc™ EZ System (Bio-Rad, California, US) and documentation and analysis were performed using the Image Lab™ software associated with the Gel Doc™ EZ System.

2.2.9 DNA purification

PCR products which required purification for subsequent cloning or sequence analysis were generally purified directly from the PCR reaction solution using QIAquick® PCR Purification Kit (QIAGEN) following the manufacturer′s instructions. In cases where the PCR reaction produced multiple bands, the desired band was purified using the QIAquick® Gel Extraction Kit, (QIAGEN) per the manufacturer′s instructions. Succinctly, desired DNA bands were precisely excised from the agarose gel with a sterile, sharp scalpel blade. Three volumes of Buffer QG was added to 1 volume of gel (100 mg ~ 100 µL) and incubated at 50°C until the gel slice was completely dissolved. The solution was added to one gel volume of isopropanol, mixed well to precipitate DNA, and applied to the QIAquick column for further purification. When purifying a PCR product from solution, Buffer PB (containing guanidine hydrochloride and isopropanol to enable efficient binding to column) with a buffer to sample ratio of 4:1 was utilised. Following gentle mixing with buffer, the sample was applied to a QIAquick column and centrifuged at 11000 × *g* for 1 min for further purification. Impurities and salts were removed from the column by washing with 0.5 mL of Buffer PE, centrifuged at 11000 × *g* for 1 min and the flow-through was discarded. The QIAquick column was once again centrifuged for an additional 1 min at ≥10,000 ×*g* to dry off the membrane. To elute DNA, 50 µL milli-Q water was added to the centre of the QIAquick membrane and centrifuged for 1 min at maximum speed. The elute was collected in a sterile microcentrifuge tube and stored at $^{-}$20 °C until further use.

2.2.10 Plasmid DNA extraction

Plasmid DNA was isolated using the QIAprep® Spin Miniprep kit (QIAGEN, California, USA) as per the manufacturer′s instructions. Bacterial colonies were used to inoculate cultures grown overnight at 37 °C in ~3 mL LB broth (supplemented with the appropriate antibiotic: 50 µg/mL ampicillin or spectinomycin). Overnight cultures were pelleted by centrifugation at 6800 × *g* for 3 min at room temperature, resuspended in 250 µL Buffer P1 followed by alkaline lysis in 250 µL Buffer P2. Upon addition of Buffer P2 colour of the solution changed to blue due to the presence of LyseBlue reagent in Buffer P1. The solution was mixed by gently inverting the microcentrifuge tube until a homogeneously coloured blue suspension was achieved. Care was taken not to allow the lysis reaction to proceed for more than 5 min. In the next step 350 µL of neutralisation buffer, Buffer N3 was added to provide high salt conditions essential for binding of DNA to the spin column silica membrane and it was mixed thoroughly by inverting the tube 4–6 times until the solution became colourless. The mixture was centrifuged for 10 min at 17,900 × *g* in a table-top microcentrifuge and 800 µL supernatant containing DNA was applied to the QIAprep spin column. Washing and elution were performed as described in section **2.2.9**.

2.2.11 Restriction enzyme digestion

Restriction digestion of purified plasmid DNA was performed using various enzymes namely: EcoR1 (10U/ µL, Promega, US), EcoRV (10U/ µL, Promega, US), HindIII (New England BioLabs, US) and BamHI (New England BioLabs, US) according to the manufacturer′s instructions. Briefly, DNA digestion reactions routinely consisted of 4 µL of plasmid DNA, 0.5 µL of restriction enzyme, 1.5 µL of CutSmart buffer [20 mM Tris-acetate, 10 mM Magnesium acetate, 50 mM potassium acetate, 100 µg/mL BSA (pH 7.9 at 25 °C)], made up to a final reaction volume of 15 µL with distilled water. Reactions were incubated at 37 °C for 60 min and the products were analysed by gel electrophoresis on a 1% agarose gel.

2.2.12 DNA sequencing and analysis

Once purified, approximately 800 ng of DNA sample (PCR product or plasmid DNA) was mixed with 1 µL forward or reverse primer, made up to 12 µL with Milli-Q water and sent to Australian Genome Research Facility (AGRF, Melbourne, Australia) for Sanger sequencing. On obtaining the results, sequence analysis was carried out using Geneious prime software (version 11.1.5, Biomatters Ltd., Auckland, New Zealand).

2.2.13 Molecular cloning and expression of recombinant flavanone MTs in *E. coli*

Cloning of PCR products was achieved by utilizing the pCR™8/GW/TOPO® TA Cloning Kit

(Invitrogen). Ligation reactions were performed at an insert: vector molar ratio of 9:1. Ligation into the pCR™8/GW/TOPO® entry vector (**Figure 2.3A**) requires a 3′ A overhanging on the end of the PCR product. Therefore, before cloning, 1 µL of 10 mM ATP with 1µL non-proof-reading DNA polymerase (MyTaq HS, Bioline, Australia) was added to approximately 10 µL of PCR product and incubated at 72 °C for 10 min for the addition of a 3′A overhang. Subsequently, 4.5 µL of the above mixture was added to 1 µL salt solution, 0.5 µL pCR™8/GW/TOPO® entry vector and incubated for 1 h at room temperature for ligation. This reaction mixture was then mixed with 30 µL of One Shot™ TOP10 Chemically Competent *E. coli* cells (Life Technologies Australia Pty Ltd. Australia) by gentle tapping and left on ice for 30 min. The mixture was then heat-shocked at 42 °C for 45 s and returned to the ice for 5 min. For cell recovery, 250 µL of LB media was added to the mixture and agitated at 37 °C for 1 h at 225 rpm in a shaking incubator. Subsequently, the cells were spun down at 5000 × *g*, 150 µL supernatant was discarded, the pellet was resuspended in the remaining 100 µL media, plated onto LB agar plates containing 50 µg/mL spectinomycin and incubated overnight at 37 °C. Multiple transformed colonies (2 - 6) were randomly picked and separately cultured overnight in LB medium containing 50 µg/mL spectinomycin before plasmid isolation following the method described in section **2.2.10**. The isolated plasmids were analysed by restriction analysis or PCR (using GW1 and GW2 primers, which bind either side of the insert entry point) to confirm the presence and correct orientation of the insert.

The verified entry clones, with the gene of interest in the correct orientation, were then subjected to LR Gateway cloning reaction, as described by the manufacturer (Invitrogen), to generate an optimised construct for recombinant protein expression. The LR reaction is a recombination reaction between the attL sites (flanking the gene of interest) in the purified entry clone and the attR sites in the Gateway® destination vector (pDEST17), which enables the effective transfer of gene of interest from entry clone to destination vector. A schematic diagram of the entry and destination vectors used in this study are shown in **Figure 2.5.**

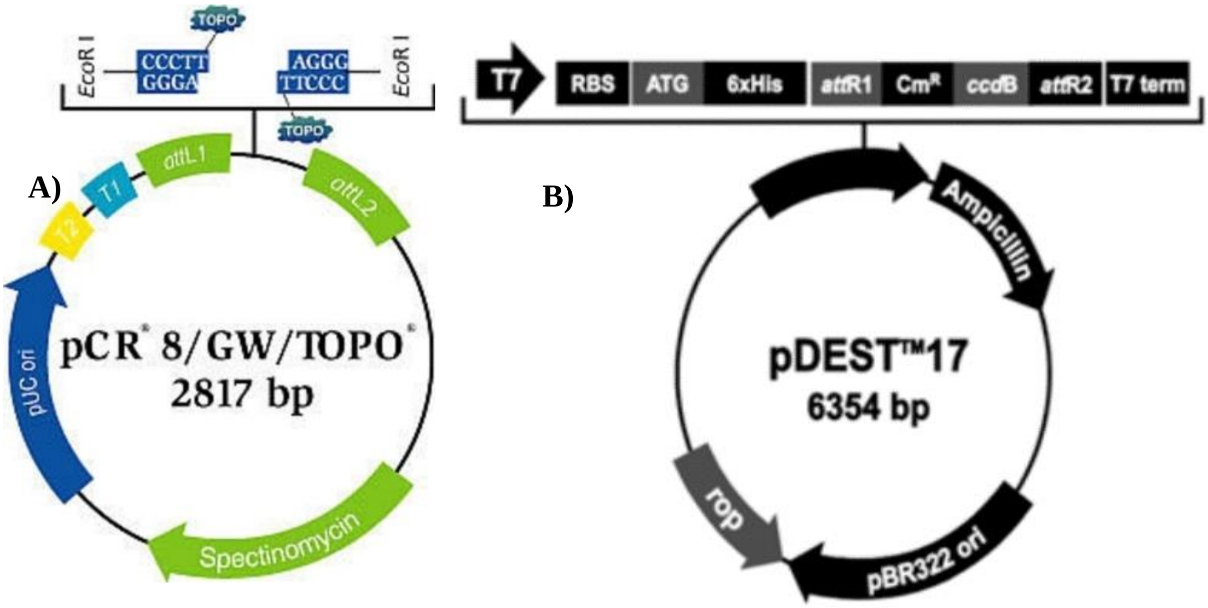


Figure 2.5: A schematic diagram of the entry and destination vectors used in this study. (A) Schematic diagram of a pCR8/GW/TOPO entry vector attL1 and attL2 sites for recombination-based transfer into destination vector (B) Schematic diagram of pDEST17 gateway destination vector.

Detailed LR reaction steps are as follows: LR reaction mixture [3 µL entry vector, 1µL destination vector, 1 µL LR Clonase enzyme mix, 5µL Milli-Q water] was incubated at 25 °C for 1 h, to enable the gene transfer, followed by addition of 1 µL proteinase K and further incubation at 37 °C for 10 min to terminate the reaction. The LR reaction mixture was then transformed into One Shot™ TOP10 Chemically Competent *E. coli* cells. Following plasmid isolation and confirmation of successful gene transfer into the pDEST17 vector, the plasmid was then introduced into the *E. coli* BL21 strain following the heat shock transformation procedure described above, and plated onto LB agar plates containing 50 µg/mL ampicillin. Transformed colonies were checked for the presence of the plasmid and sub-cultured for expression when required.

As an alternative protein expression system, the coding sequences of MTs were cloned into the Histidine-tagged Ubiquitin Expression vector (pHUE) (Catanzariti et al., 2004). pHUE contains the inducible T7 RNA polymerase promoter, a histidine tag at the 5′ end of a ubiquitin open reading frame and an extended poly-linker (**Figure 2.6**) (Catanzariti et al., 2004).

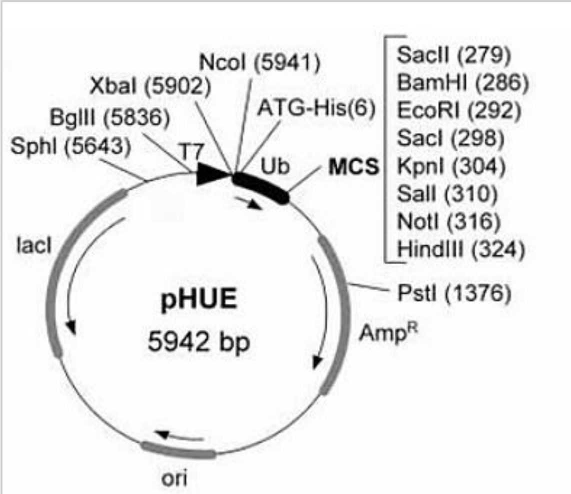


Figure 2.6: Schematic representation of pHUE cloning vector showing the ubiquitin coding region (Ub); the T7 polymerase promoter (black triangle); polyhistidine tag(His 6); β-lactamase gene;colE1 origin of replication (ori) and the lacI repressor gene (lacI). Arrows indicate the direction of transcription (Catanzariti et al., 2004).

To clone into the pHUE vector, genes were amplified via PCR using oligonucleotide primers with suitable restriction enzyme sites incorporated into forward and reverse primers. After amplification, the products were purified and then digested with NotI and KpnI. For 20 µL of the purified PCR product, the reaction mixture contained 2.5 µL of CutSmart buffer and 1 µL each of restriction enzymes. Similarly, approximately 1 µg (~4 µL) of pHUE vector was also digested in a total volume of 20 µL with the same restriction enzymes. Digested PCR product and vector were purified separately, before ligation, following the steps described in section **2.2.9** and **2.2.10**. The ligation reaction comprised of 3 µL digested vector, 9 µL digested PCR product, 1.5 µL 10X buffer and 1.5 µL T4 ligase in a total volume of 15 µL and it was left at room temperature for 1-2 h or at 4 °C overnight. The ligated vector was then transformed into TOP10/JM109 competent *E. coli* cells and once verified was subsequently transformed into *E. coli* BL21 strain following the same procedure used for pDEST17 cloning described above.

For *in vitro* protein expression, 2 mL starter culture of the selected colony was added to 200 mL of LB medium (supplemented with 50 µg/mL ampicillin). The culture was grown to an OD_{600} of 0.6, on a 180 rpm shaking incubator, at 37 °C, followed by the addition of isopropyl-*β*-D-thiogalactopyranoside (IPTG) to a final concentration of 0.1 mM. The cultures were then incubated at 16 °C overnight and harvested by centrifugation (6,000 × *g*, 10 min). The resulting pellet was resuspended in lysis buffer [50 mM NaH_2PO_4, 10 mM imidazole, 300 mM NaCl one protease inhibitor cocktail tablet (Roche, NSW, Australia), 1mg/mL Lysozyme] by pipetting up and down, homogenised and intracellular contents were released by sonicated (30 s, 1 min cooling on ice, 4 times). The resulting solution was then centrifuged at 10,000 × *g* for 5 min, and the supernatant was used as a crude enzyme solution for the enzyme assays.

2.2.14 Recombinant protein purification and Dot-Blot / Western Blot assay

Crude enzyme extracts were purified by passing them through a His-Tag protein purification column (Ni-NTA Agarose; QIAGEN) following the manufacturer′s instructions. Efficient binding of the protein to the beads was promoted by mixing 2 mL of clear lysate with 3 mL of Ni-NTA agarose beads (pre-equilibrated with lysis buffer) for 1 h at 4 °C. The mixture was then packed into a standard glass Econo-Column (15 cm × 0.5 cm, Bio-Rad) by gravity flow and washed two times with a wash buffer (0.5 M Tris-HCl, pH 8.0, 5 M NaCl, 1 M imidazole, 0.5 M EDTA) and subsequently eluted with an elution buffer containing an increasing concentration of imidazole in wash buffer with a gradient of (100-600 µM) of imidazole in wash buffer. The fraction containing purified protein was determined by dot blot assay utilizing a His-tag Monoclonal Antibody (Life technologies) following manufacturer′s instruction. Briefly, 2µL crude lysate/eluate was immobilized on to the nitrocellulose membrane, allowed to dry, and incubated by rotating overnight at 4 °C with anti-His antibody (MA1-21351 from Thermo Fisher, 1 in 1000 dilution in TBST buffer [0.2 M Trizma base, 1.5 M NaCl, pH 7.2- 7.4, 0.05% Tween 20 containing 3% milk powder] in a sealed plastic bag. The membrane was then washed thrice, for 5 min each, in TBST buffer before applying secondary antibody [horseradish peroxidase (HRP) conjugated anti-mouse IgG, 1 in 10,000 dilutions in TBST buffer containing 3% milk powder] and further incubated for 1 h at room temperature following which, the membrane was washed three times (5 min each) in TBST buffer and the signal was developed using a Thermo Scientific™ SuperSignal™ West Femto Chemiluminescent Substrate Kit, by incubating the membrane in a 1:1 mixture of both Femto developing solutions for 3 min. Chemiluminescent detection was done using Bio-Rad ChemiDoc MP Imaging System

2.2.15 *O*-Methyltransferase assay

Activity of purified ZmOMT1 was assayed with quercetin as substrate, following the protocol described by Itoh et al. (2016) with an incubation time of 30 min. The activity of crude OMT was assayed with various substrates, in 500 µL reactions containing 50 mM potassium phosphate buffer (pH 7), 200 µL crude OMT enzyme (approx. 0.6 mg protein), 500 µM SAM and 10-100 µM substrate (dissolved in DMSO). Various negative controls, as described in **Table 2.2**, were run along with the test to account for the accuracy of the analysis.

Table 2.2: List of negative controls used in recombinant methyltransferase assay.

Name	Buffer	Substrate	Enzyme	SAM
Control A	☑[†]	☑	X[‡]	X
Control B	☑	X	☑	X
Control C	☑	X	X	☑
Control D	☑	☑	☑	X
Control E	☑	X	☑	☑
Control F	☑	☑	X	☑

[†]☑- component present [‡]X- component absent

All reactions were incubated at 30 °C with gentle shaking for 20 min and terminated by the addition of 500 µL acetonitrile. The soluble fraction of the assay mixture, containing the methylated products, was separated from the precipitate by centrifugation and analysed using RP-HPLC as described in section **2.2.2**. To ensure the feasibility of the proposed method for expression, purification and activity assay of MTs, an *O*-methyltransferase with a known activity (ZmOMT1) was also used aa a positive control, with quercetin as the substrate (Zhou et al., 2008). All assays were performed in triplicate. The identity of the methylated derivatives was confirmed by ESI-LCMS/MS.

2.2.16 Assay product analysis

Chromatographic separation of the reaction mixture was achieved on an Agilent 1220 Infinity High- Performance Liquid Chromatography (HPLC) system with a Gemini C18 analytical column (5 µm, 150 × 4.6 mm; Phenomenex) and a diode array UV detection (283 nm). A gradient elution method, as described in section **2.2.2**, with acetonitrile and water (both acidified with 2% acetic acid) was employed. Flavanones in the assay product were quantified by comparing the absorbance to that of pure compound. Where required, identification of reaction products was done using ESI-LC-MS/MS.

Electro-Spray Ionisation Liquid Chromatography-Mass Spectrometry (ESI-LCMS) of non-volatile compounds was conducted on an Agilent 6520 quadrupole time of flight mass spectrometer system (Agilent Technologies, USA) with a dual spray ESI attached to an Agilent 1200 series HPLC with a diode array detector using a Gemini C18 (5 µm, 150 × 4.6 mm;

Phenomenex) analytical column eluted at a flow rate of 0.8 mL min^{-1}. A gradient elution method with water (solvent A) and acetonitrile (solvent B), both acidified with 0.1% acetic acid, was employed. The eluent system was a gradient of solvent B from 30 to 50% over 7 min, followed by 50–95% over 7 min and then isocratic at 95% for a further 12 min, and finally a gradient of 95–100% over 0.25 min. Electrospray ionization was achieved in positive ion mode using the following conditions: nebulizer pressure 37 psi, gas flowrate 12 L min^{-1}, gas temperature 350°C, capillary voltage 4000V, fragmenter 150 and skimmer 65 V. The conditions were modified depending on the compound to be analysed whenever needed and other source parameters were set to standard conditions recommended by Agilent. TOF mass range was set from *m/z* 70–1700. Accurate mass data calculation of the molecular ions, as well as the processing of chromatograms, were performed using Mass Hunter Workstation (Agilent Technology).

2.3 Results

2.3.1 Transcriptomic analysis of *O-methylated species* 1 versus *non-methylated* species

To isolate OMTs from *Eucalyptus*, transcriptome sequencing of two *Eucalyptus* species with different complements of flavanone compounds was performed. Specifically, a species with a significantly high concentration of *O*-methylated flavanones (*O-methylated species 1*) and a second species with no detectable *O*-methylated flavanones (*non-methylated species*) were chosen. The flavanone profiles of the two selected *Eucalyptus* species are shown in **Figure 2.7**.

Figure 2.7: Flavanone profile of *O-methylated species 1* and *non-methylated species*. The data are average values of twelve individual trees.

Preliminary attempts at RNA extraction from *Eucalyptus* leaves using the commercially available spin-column matrices were unsuccessful. Perhaps it was due to the presence of high amounts of extractable polyphenols, and other secondary metabolites in *Eucalyptus* leaves. The oxidized forms of these compounds can covalently link to nucleic acid and lead to production of low quality total RNA (Salzman et al., 1999). Also, the contaminated DNA removal utilizing DNase I with heat inactivation of the enzyme also resulted in RNA damage. Therefore, RNA extraction using CTAB buffer with 2% PVP-40 (inhibit polyphenol oxidases) and an on-column DNA digestion on a RNeasy mini spin column was employed. Total RNA was successfully extracted from leaf tips of the two species. All RNA samples passed the quality check for library construction and sequencing, with RNA integrity number (RIN) values ranging between 8.3 and 9.2 (**Figure 2.8**). The concentration of RNA samples ranged from 140-515 ng/µL and the A260/A280 ratio was greater than 2.1 for all the samples.

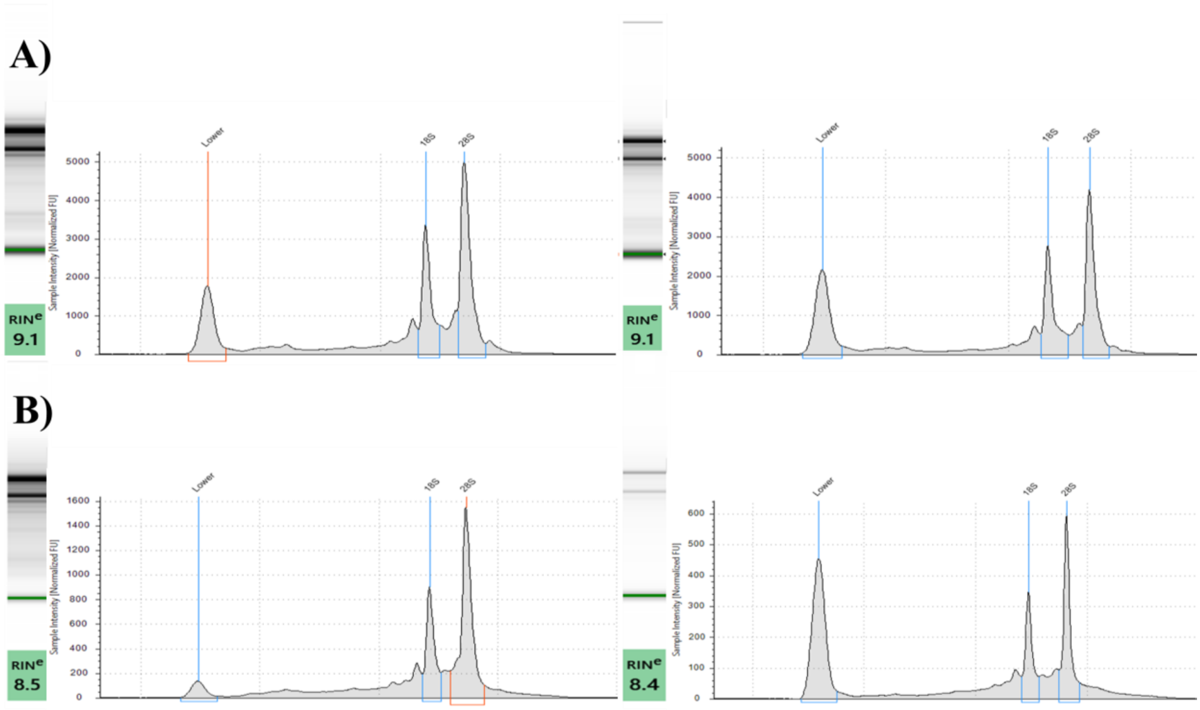


Figure 2.8: RNA quality analysis using TapeStation. RNA electrophoresis gel image with RNA integrity number (RIN) values (left side) and corresponding electropherograms with fluorescence unit (FU) on the Y-axis and fragment size (bp) on the X-axis. Peaks of 18S and 28S rRNA label for are shown for *O-methylated species 1* (A) and *Non-methylated species* (B).

iv. <u>Library preparation and quality control</u>

A cDNA library was successfully constructed and the raw data from Illumina sequencing was first passed through quality control. The *O-methylated species 1* raw data had 59,662,210 raw read counts whereas *non-methylated species* had 49,048,364 raw read counts. The graph of error rate distribution along the reads (**Figure 2.9**) shows an average sequencing error rate of 0.01% (*O-methylated species 1*) and 0.02% (*non-methylated species*). Error rate increases as the sequencing reads are extended and sequencing reagents become more and more scarce. Generally, single base error distribution rates below 1% are considered as acceptable quality. Therefore, in our samples, the average sequencing error rate of 0.02% indicates high sequencing quality.

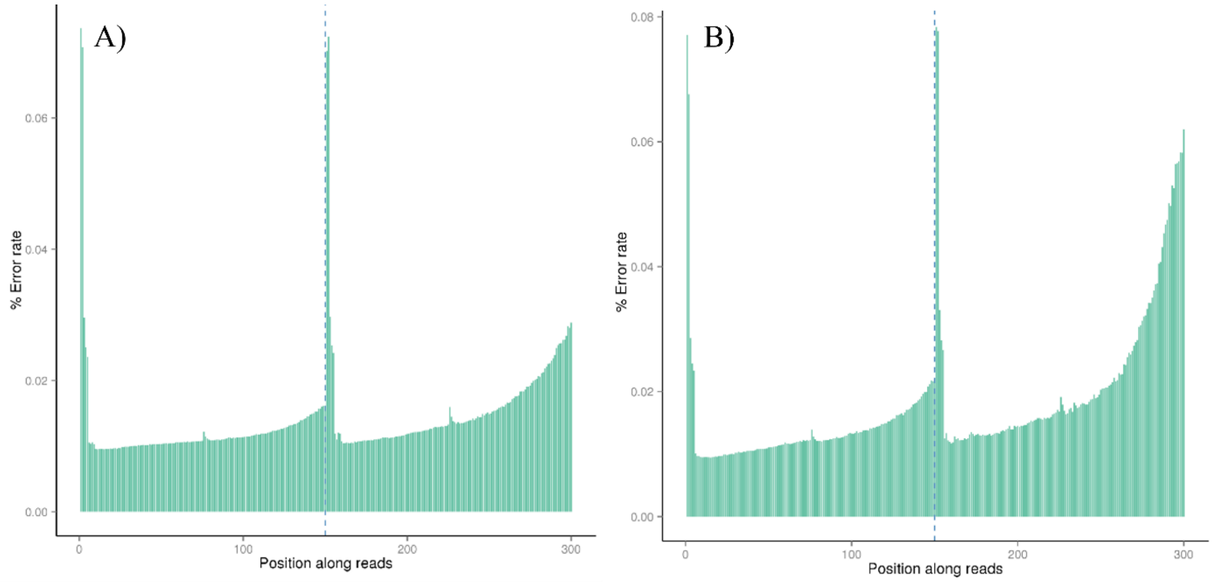


Figure 2.9: Error rate distribution along reads for A) *O-methylated species 1* and B) *non-methylated species.*

Uncertain bases in the sequenced reads were designated as N. To improve the quality, the sequenced data were filtered by removing reads according to three criteria: low quality, reads with more than 10% N, and reads with adaptor contamination. After stringent quality assessment and data filtering, 94.79% of *O-methylated species 1* raw-reads and 94.48% of *non-methylated species* raw-reads (**Figure 2.10**) were selected as high-quality reads for further analysis. A summary of the raw data quality check is shown in **Table 2.3**.

Table 2.3: Raw data quality control summaryshowing read counts after filtering low quality reads

Sample	Raw Reads *	Clean Reads †	Clean Bases ‡	Error (%)	Q20 (%) ¶	Q30 (%)††	GC (%)
O-methylated species 1	59662210	56555618	8.5G	0.01	97.38	93.66	51.09
Non-methylated species	49048364	46342630	7G	0.02	95.92	90.56	51.79

*Raw Reads: the original sequencing reads counts. †Clean Reads: number of reads after filtering. ‡ Clean Bases: clean reads number multiplied by read length, saved in G unit. ¶ Q20: percentages of bases whose correct base recognition rates are greater than 99% in total bases. †† Q30: percentages of bases whose correct base recognition rates are greater than 99.9% in total bases.

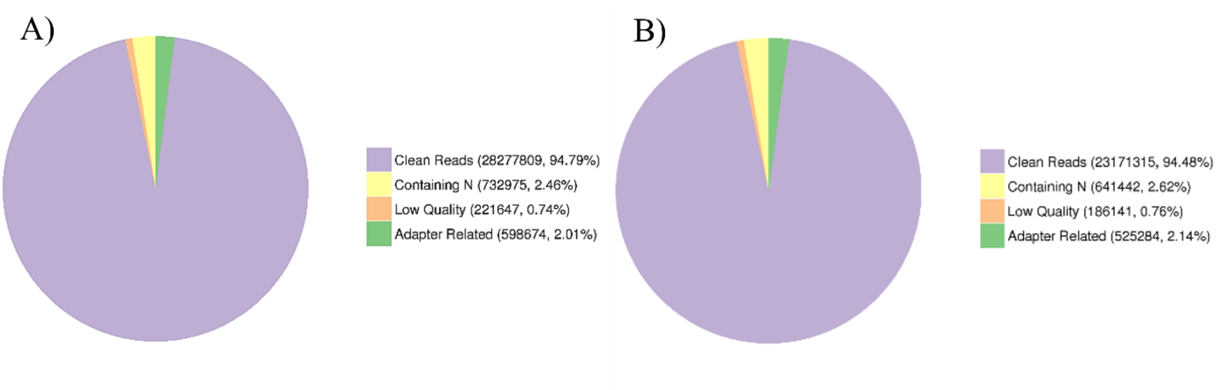


Figure 2.10: Classification of raw reads for *O-methylated species 1* (A) and non-methylated species (B).

v. <u>Transcriptome reconstruction</u>

Clean reads were *de novo* assembled by Trinity (Grabherr et al., 2011) to get an assembly transcriptome followed by Hierarchical Clustering by Corset (Davidson and Oshlack, 2014) to remove redundancies. The longest transcripts of each cluster were then selected as unigenes. Length distribution information of transcripts and unigenes are listed in **Table 2.4 and Figure 2.11**. *Non-methylated species* had a higher number of short transcripts compared with *O-methylated species 1*.

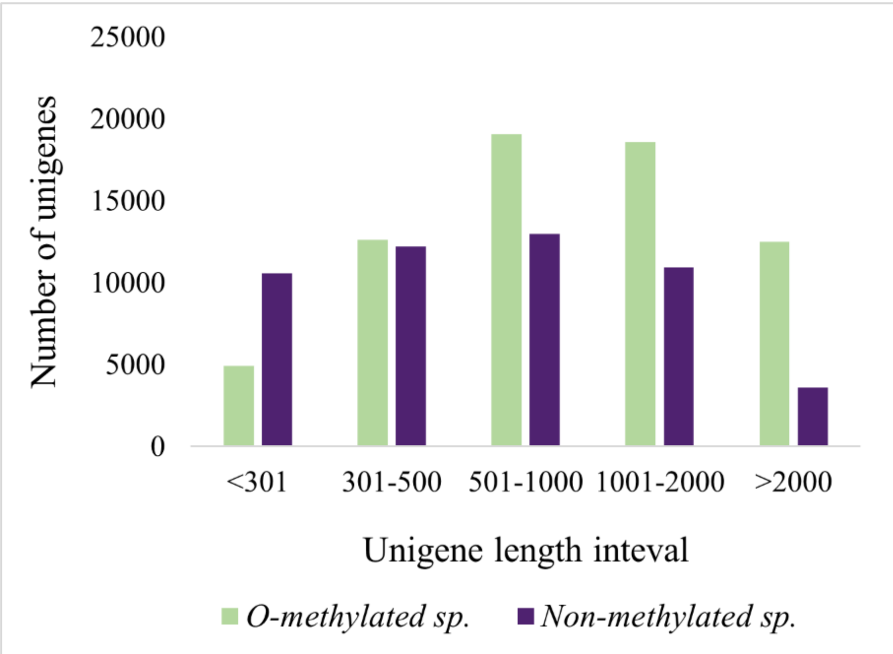


Figure 2.11: Unigene length distribution in *O-methylated species* 1 and *non-methylated species*.

Table 2.4: Transcript gene grouped according to length in both *O-* and *non-methylated* species transcriptome.

Transcript length interval		200-500bp	500-1kbp	1k-2kbp	>2kbp	Total
O-methylated species 1	Number of transcripts	17561	19071	18589	12522	67743
	Number of Unigenes	17546	19071	18589	12522	67728
non-methylated species	Number of transcripts	22805	13003	10922	3588	50318
	Number of Unigenes	22788	13003	10922	3588	50301

vi. Gene function annotation

The function of each unigene cluster was annotated with the help of seven mainstream databases: namely, Nr (NCBI non-redundant protein sequences), Nt (NCBI nucleotide sequences), Pfam (Protein family), KOG (euKaryotic Orthologous Groups)/COG (Cluster of Orthologous Groups of proteins), Swiss-Prot, KEGG (Kyoto Encyclopedia of Genes and Genome), and GO (Gene Ontology). The percentage of unigenes annotated in at least one database in *O-methylated species 1* and *non-methylated species* was 80.45% and 86.41%, respectively. The majority of the unigenes were annotated based on their sequence similarity with the NCBI nucleotide sequences and Swiss-Prot entries. 10% of *O-methylated species 1* unigenes and 13.6% of *non-methylated species* unigenes were annotated in all seven databases. **Figure 2.12** shows gene function annotation results using five selected databases out of the seven. Both species had a very low E-value distribution which emphasises a high annotation accuracy.

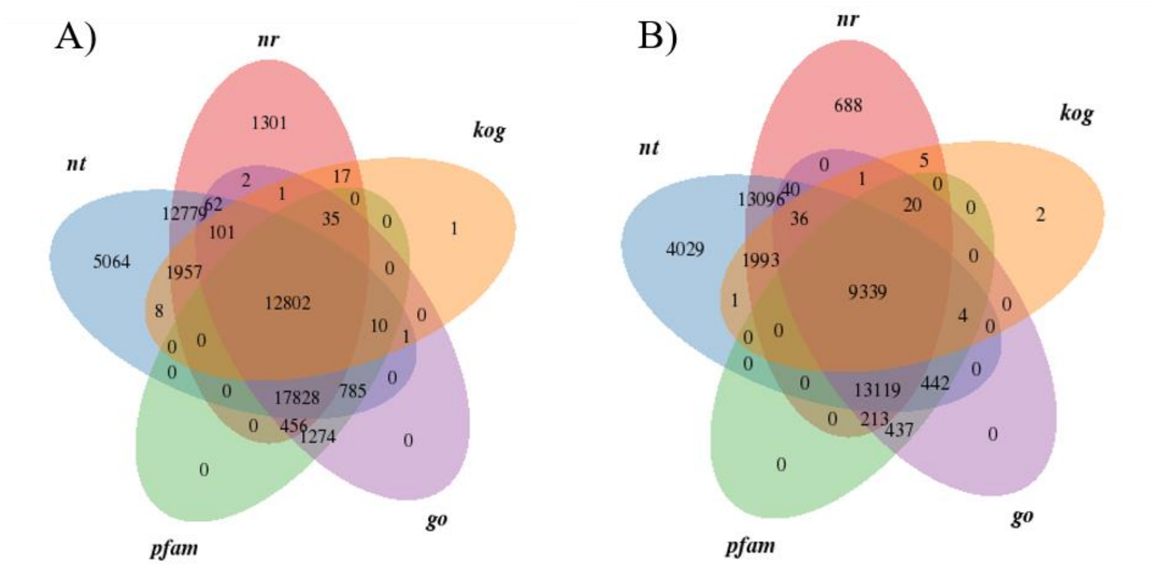


Figure 2.12: Venn diagrams showing gene function annotation results using five selected databases, including Nt (blue), Nr (red), KOG (orange), GO (purple), and Pfam (green) showing number of annotated genes in A) *O-methylated species 1* and B) *Non-methylated species*. The numbers in each intersection of Venn diagram represent the number of transcripts with significant hits in respective databases. Abbreviations are as follows: Nr- NCBI non redundant protein database, Nt- NCBI nucleotide database, KOG- eukaryotic orthologous groups, GO- gene ontology database, Pfam- Protein family database.

vii. Gene expression analysis

Using the *de novo* transcriptome as a reference, reads were mapped back to transcriptome and the expression level was quantified employing RSEM (Li and Dewey, 2011). Out of the total reads, 77.37% and 72.71% were mapped to the reference genome in *O-methylated species 1* and *non-methylated species*, respectively (**Table 2.5**). For calculating the expression level, read counts of each sample were calculated and converted to a corresponding FPKM (expected number of Fragments Per Kilobase of transcript sequence per Millions base pairs sequenced) values.

Table 2.5: Summary of read mapping to reference genome. Values are total number of sequenced reads after quality control (total reads) and number/percentage of reads mapped to the reference genome in the selected species of *Eucalypts*.

Sample name	Total reads	Total mapped
O-methylated species 1	56555618	43755712 (77.37%)
non-methylated species	46342630	33694156 (72.71%)

FPKM values are reliable for estimating gene expression levels as they consider the effects of both sequencing depth and gene length on the counting of fragments. Density distributions of FPKM displaying the overall gene expression levels are shown in **Figure 2.13**.

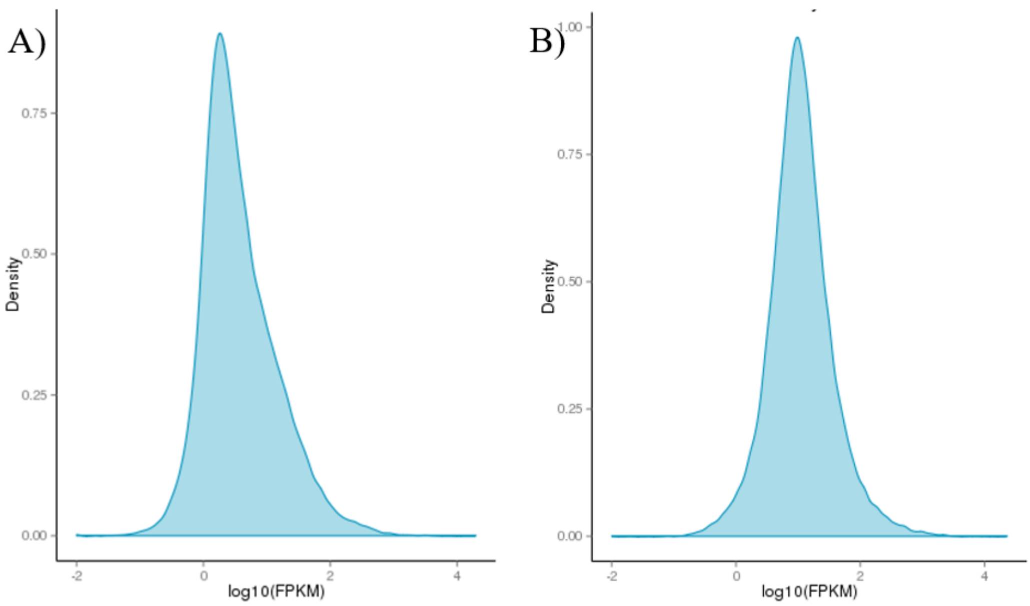


Figure 2.13: FPKM density distribution. A) *O-methylated species 1*. B) *Non-methylated species*.

2.3.2 Putative MTs identification from transcriptomic data

The sequenced transcriptomes of both species were screened for putative MTs. Many MTs exhibit a highly conserved C-terminal sequence (Joshi and Chiang, 1998; Lam et al., 2007), which was utilized for the selection of putative flavanone OMTs. In *non-methylated species*, 55 candidate (0.11%) unigene clusters were annotated as MTs whereas, in *O-methylated species 1*, 63 candidates were annotated (0.099%).

Unigene clusters that were annotated as MTs (based on the sequence similarities with the existing characterized MTs) in the respective *Eucalyptus* species were compared and screened based on their FPKM value, which correlates directly with their expression levels (**Figure 2.14**). *O-methylated species 1* is rich in 7-*O*-methylated flavanone pinostrobin and non-methylated aglycone pinocembrin, whereas *non-methylated species* lacks any of the methylated counterparts and has very high amounts of pinocembrin. Therefore, transcripts that are highly expressed in *O-methylated species 1* were "blasted" against the *non-methylated species* transcript gene database, and among the top "hits", those which showed high expression in *O-methylated species 1* but low expression in *non-methylated species*, were selected as the gene clusters for further analysis (**Table 2.6)**. The difference in gene expression levels of EnOMT1 and ExOMT1 between the two respective species, *O-methylated species 1* and *non-methylated species* was further confirmed using qPCR (**Appendix E**).

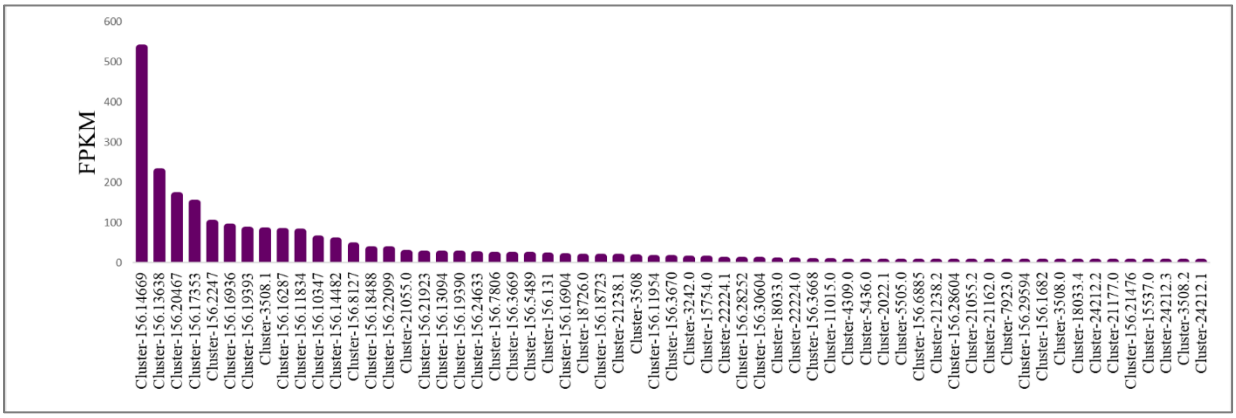


Figure 2.14: Unigene clusters annotated as MTs in *O-methylated species 1* (X-axis) with corresponding FPKM values (Y-axis).

Table 2.6: List of candidates tested for *O*-methyltransferase activity.

O-methylated species 1			Similar *Non-methylated species* gene cluster	Protein pairwise identity (%)	Query coverage	*Non-methylated species* FPKM
gene name	gene cluster name	FPKM				
EnOMT 1	Cluster-156.14669	534.6	Cluster-13293.15022	94	100	157.4
EnOMT 2	Cluster-156.13638	227.8	Cluster-8398.0	64	100	18.2
EnOMT 3	Cluster-156.20467	168.1	Cluster-13293.12940	91	100	35.7
EnOMT 4	Cluster-156.16936	89.1	Cluster-13293.14312	98	100	283.5
EnOMT 5	Cluster-156.28252	6.2	Cluster-13293.23918	97.6	100	54.6
EnOMT 6	Cluster-156.2247	99.3	Cluster-13293.14817	98.5	100	14.1
EnOMT 7	Cluster-3508.1	80.1	Cluster-7646.0	83.2	88	10.1

long with the relative expression levels, screening based on characteristic plant OMT motifs (suggested by Ibrahim et al., 1998; Joshi and Chiang, 1998) aided narrowing down putative OMTs to seven unigene clusters, named EnOMT 1 through 7, as candidates for initial analysis (**Figure 2.15**). Three out of seven candidates namely, EnOMT4, EnOMT5, EnOMT6 were selected primarily based on high sequence identity to that of the flavonoid OMTs (Lam et al., 2007) previously characterized from other species, as flavanones fall in the same class of secondary metabolites. Both EnOMT4 and EnOMT5 showed a significance difference in motifs B and C. Although EnOMT4 and EnOMT5 showed higher expression in *non-methylated species* (therefore, did not follow the relative expression criteria), I was interested to see whether they have the methyltransferase activity.

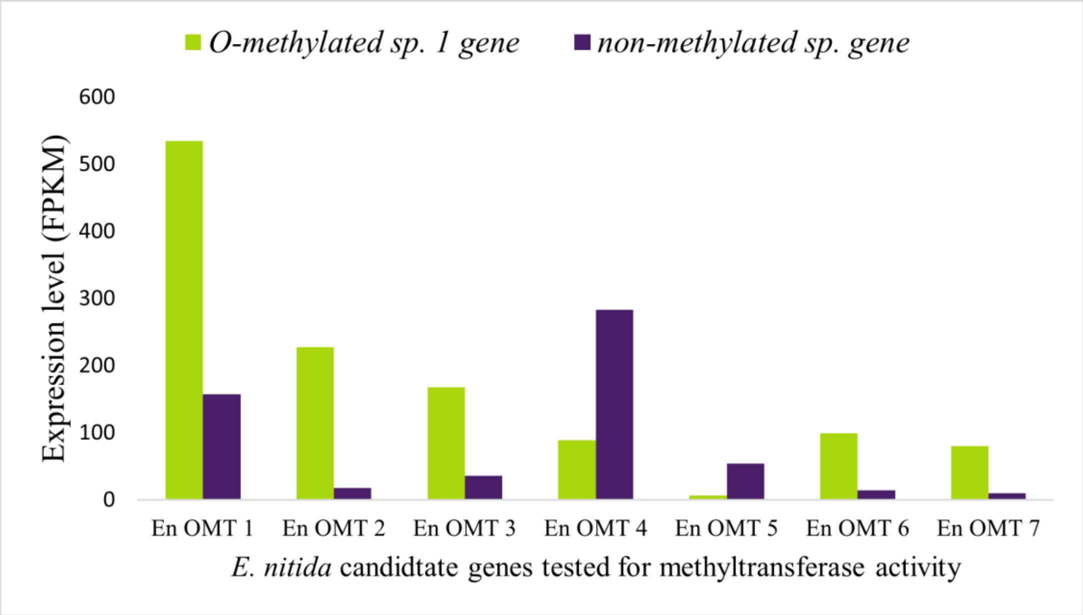


Figure 2.15: Expression level comparison of *O-methylated species 1* gene candidates tested for *O*-methyltransferase activity with corresponding similar genes from non-methylated species.

Predicted amino acid sequences of *O-methylated species 1* genes selected for further analysis showed high sequence identity and conservation of motifs predicted to be involved in methyltransferase activity (**Figure 2.16**).

Figure 2.16: Multiple sequence alignment of putative proteins tested for in this study with conserved motif locations suggested by Joshi and Chiang (1998). Motif A (blue); Motif B (green); Motif C (red).

2.3.3 Cloning, expression, and purification of *Zea mays (*ZmOMT1*)* in *E. coli*:

An OMT with a known activity was used as a positive control to ensure the feasibility of the proposed method for expression, purification, and activity assay of MTs. A fully characterized flavone-specific *O*-methyltransferase gene from *Zea mays*, *ZmOMT1*, that preferably acts on the flavone tricetin was selected for this purpose (Zhou et al., 2008). The *ZmOMT1* was successfully cloned, expressed in *E. coli*, and subsequently purified. The activity was assayed with quercetin, a flavone substrate of less preference (only 77% activity with respect to tricetin). The HPLC chromatogram of enzyme assay products shows that the purified extract catalysed quercetin conversion into a product with a retention time of 7.86 min (**Figure 2.17**). About 19.7% of the quercetin was converted into the new product under the assay conditions. Incubating the reaction mixture for a longer time did not seem to affect the conversion rate possibly because the product formed inhibited the enzyme activity.

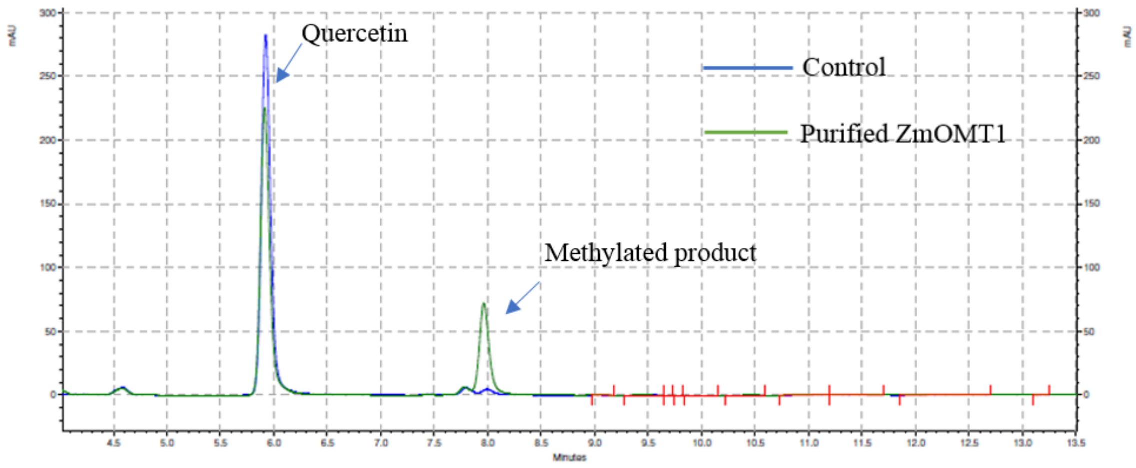


Figure 2.17: HPLC Chromatogram of ZmOMT1 assay products overlayed with substrate control. Additional peak with a later retention time compared to the substrate quercetin corresponds to the mono-methylated quercetin derivative.

LC-MS/MS was performed to determine the identity of the newly formed compound. Mass spectra of the enzyme assay products shows that quercetin eluted at 6.7 min with *m/z* 301 [M-H]$^-$. The product eluted at 8.6min with *m/z* of 315 [M-H]$^-$ (**Figure 2.18**-A). This increase of 15 units in the *m/z* value of the product, compared to that of quercetin, corresponds to the molecular mass of a methyl group, suggesting that the product formed is monomethylated quercetin. Subsequent fragmentation of quercetin on MS2 showed the parent ion *m*/*z* 301 [M-H]$^-$and a key MS2 fragment ion of *m*/*z* 151 [M-H]$^-$ due to cleavage across C-ring. The *m*/*z* 178.9988 is a contaminant ion (**Figure 2.18**-B). Further characteristic fragmentation analysis of the OMT product on MS2 resulted in peaks corresponding to a methylated parent ion with an *m*/*z* of 315 [M-H]$^-$, a fragment with a loss of methyl group with an *m*/*z* of 300 [M-H]$^-$ and a key MS2 fragment ion with an *m*/*z* of 151 [M-H]$^-$due to cleavage across the C-ring.

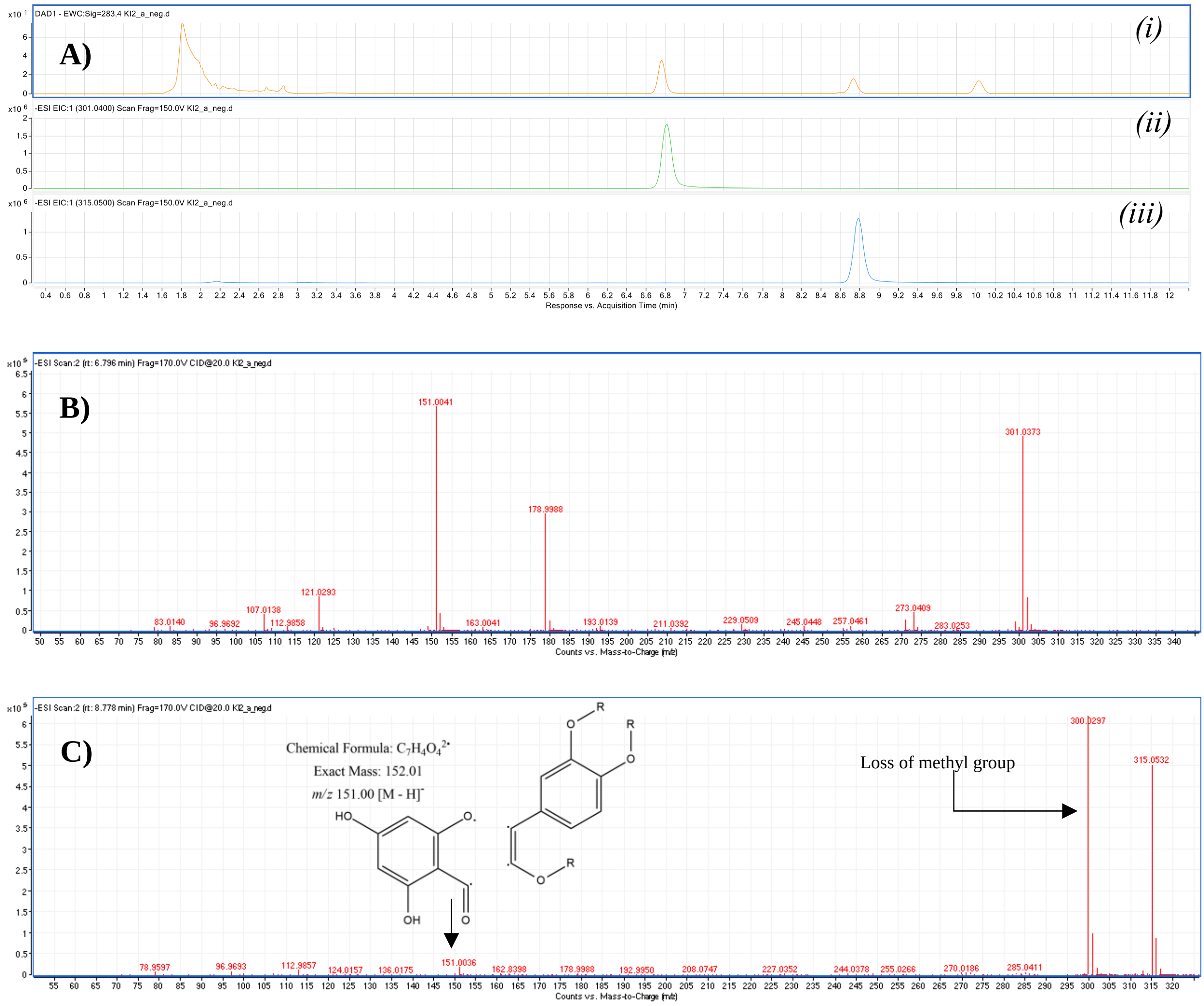


Figure 2.18. Representative mass spectra of quercetin and ZmOMT1 assay product. A) Quercetin and methylated quercetin UV data at 283 nm) (*i*) and Extracted Ion Current for quercetin (*ii*) and mono-methylated quercetin (*iii*). B) Quercetin MS2, showing parent ion m/z 301 $[M\text{-}H]^-$ and key MS2 fragment ion of m/z 151 $[M\text{-}H]^-$ due to cleavage across C-ring. C) OMT product MS2, showing methylated parent ion m/z 315 $[M\text{-}H]^-$, MS2 loss of methyl group m/z 300 $[M\text{-}H]^-$ and key MS2 fragment ion of m/z 151 $[M\text{-}H]^-$ due to cleavage across the C-ring.

Quercetin can form six possible mono *O*-methylated compounds and the presence of a characteristic MS2 fragment ion of *m/z* 151 $[M\text{-}H]^-$ indicates that the methylation is not on the A-ring (**Figure 2.19**). *Zea mays* OMT must have transferred a methyl group preferentially to the C- or B- ring hydroxyl groups, which is consistent with the results obtained by Zhou et al., (2008), where the OMT catalysed B-ring *O*-methylation. The successful enzyme assay using ZmOMT1 proved that our proposed method and techniques are adequate to isolate and characterize *Eucalyptus* OMTs identified through the transcriptomic analysis.

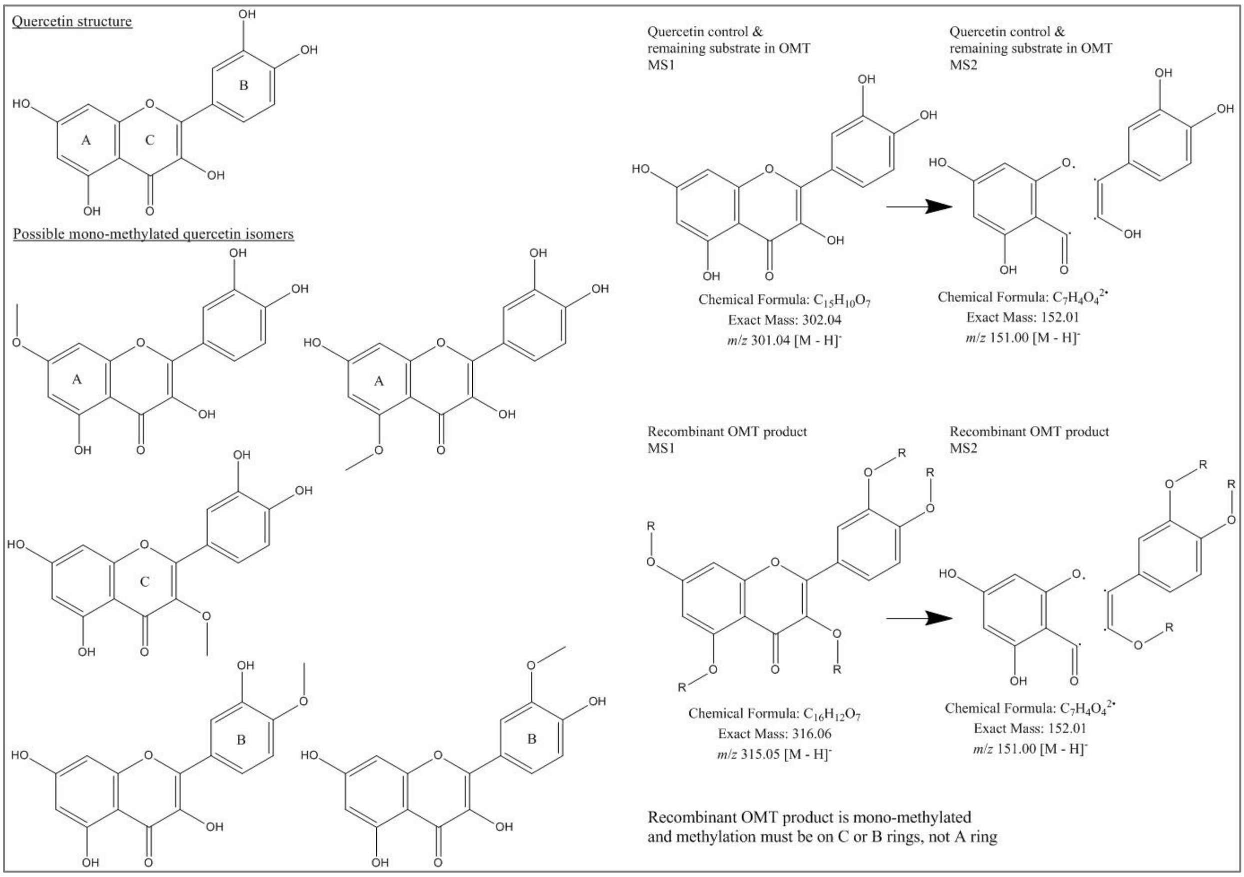


Figure 2.19. Characteristic fragmentation patterns of quercetin observed in ESI-LCMS/MS

2.3.4 Cloning and analysis of *O-methylated species 1* OMT candidates

Full-length coding sequences of putative OMTs from transcriptomic data were amplified with specific primers flanking the coding region of selected unigene clusters using *O-methylated species 1* leaf cDNA as a template. All primers used to clone these genes are listed in **Table 2.1**. All seven genes were successfully cloned into the pDEST17 vector system and expressed in *E. coli* BL21 cells. The recombinant proteins were obtained as a fusion protein with a 6X Histidine tag at the N-terminus to facilitate detection and purification when required. To determine enzyme activity, crude cell lysate was used directly in assays containing substrate pinocembrin and the methyl donor SAM. Control assays (with no substrate or with cell lysate of an expressed empty vector) were performed and showed no activity. HPLC chromatograms of the assay products of various candidates tested are given in **Figure 2.20.** Among the seven-candidate OMT gene clusters that were successfully cloned, expressed and analysed, EnOMT1 successfully metabolized pinocembrin into a product with a later retention time to that of pinocembrin (**Figure 2.20)**.

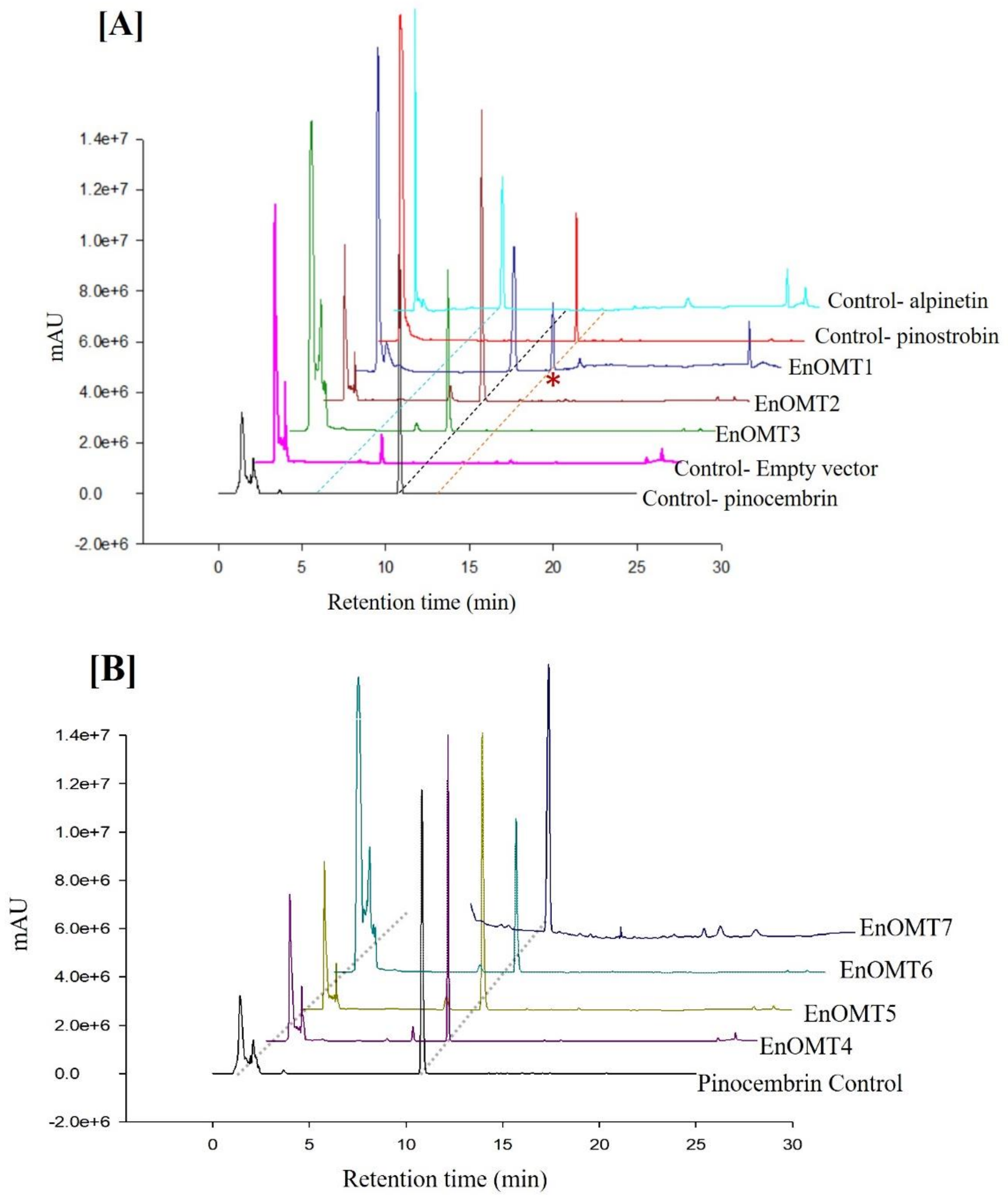


Figure 2.20: HPLC chromatograms showing assay products of EnOMT clones tested. A) EnOMT1, EnOMT2, and EnOMT4 aligned with empty vector control and flavanone standards. The * denotes peak corresponding to the methylated product. The initial peaks correspond to buffer and other polar constituents. B) EnOMT3, EnOMT5, EnOMT6, EnOMT7 aligned with pinocembrin control showing peaks corresponding to pinocembrin but no methylated products.

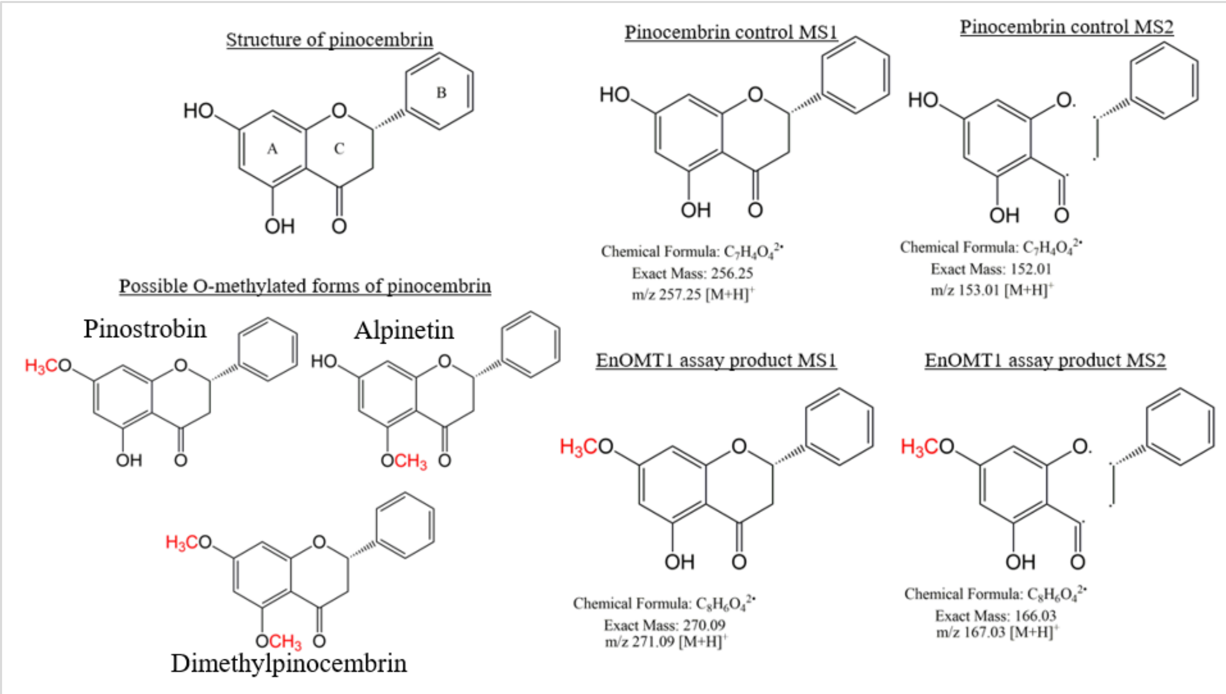


Figure 2.21. Characteristic fragmentation patterns of pinocembrin and EnOMT1 assay product observed in ESI-LCMS/MS.

There are three possible *O*-methylated forms of pinocembrin: two monomethylated (pinostrobin, alpinetin) and one di-methylated derivative (dimethylpinocembrin; **Figure 2.21**). Definitive identification of the assay product catalysed by EnOMT1 was made by positive-ion electrospray ionization (ESI)-LC-MS/MS analysis of the reaction mixture. Mass spectra of the control assay showed pinocembrin eluting at 11.8 min with *m/z* 255.05 $[M – H]^-$ whereas the EnOMT1 assay product eluted at 14.328 min with *m/z* 271 $[M+H]^+$ (**Figure 2.22 A, B**).

Figure 2.22. Representative mass spectra of EnOMT1 assay product with pinocembrin as a substrate. (A) (i)UV data at 283 nm and Extracted Ion Current for (ii) mono methylated pinocembrin m/z 271 $[M + H]^+$ in MS1and (iii) mono methylated pinocembrin m/z 271 $[M + H]^+$ in MS2. (B) EnOMT1 assay product MS1 showing mono methylated pinocembrin m/z 271 $[M + H]^+$ in MS1. (C) EnOMT1 assay product MS2 showing methylated parent ion m/z 271 $[M + H]^+$ and key MS2 fragment ion of m/z 167 $[M + H]^+$ due to cleavage across C-ring.

An increase in the m/z value of the product to that of pinocembrin of 15 units was detected, which corresponds to the molecular mass of a methyl group, suggesting that the product formed is mono-methylated pinocembrin. Subsequent fragmentation of pinocembrin in control assay on MS2 showed the parent ion *m/z* 257.25 $[M+H]^+$and a key MS2 fragment ion of *m/z* 153.01 $[M+H]^+$ due to cleavage across C-ring. Further characteristic fragmentation analysis of the EnOMT1 product on MS2 resulted in peaks corresponding to a methylated parent ion m/z 271 $[M+H]^+$, a key MS2 fragment ion of m/z 167 $[M+H]^+$due to cleavage across the C-ring- the most common point of weakness in flavanones (**Figure 2.21**, **Figure 2.22C**) and another fragment produced by cleavage across the C-ring, combined with the loss of H atoms (i.e. double bond formation in the remnant C-ring) with m/z 131 $[M+H]^+$ (**Figure 2.23**; Sun et al. 2016)

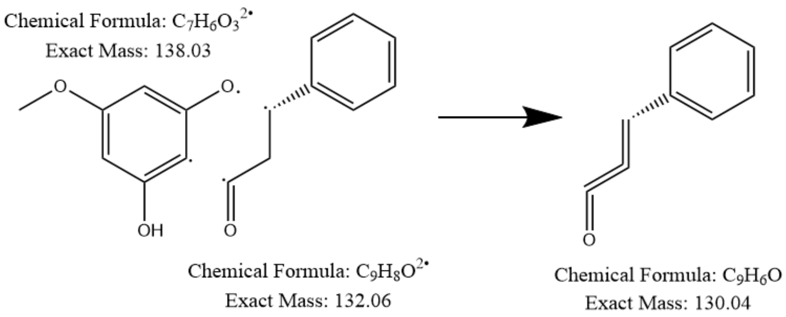


Figure 2.23: Characteristic fragmentation pattern that results in the formation of MS2 fragment ion of *m/z* 131.04 $[M+H]^+$.

The MS2 fragmentation pattern makes it clear that the mono-methylation is on the A-ring of pinocembrin. Therefore, the product can either be alpinetin (5-*O*-methylated) or pinostrobin (7-*O*-methylated). Since fragmentation across the A-ring is rare, MS2 data alone is insufficient to determine the exact position of the methyl group on A-ring. Therefore, RP-HPLC retention time comparison of standard solutions of pure alpinetin and pinostrobin, to that of the unknown assay product, was carried out. Alpinetin showed a retention time of 7.5 min whereas pinostrobin eluted at 13.4 min, which matched that of the methylated product (**Figure 2.20**A). Hence, it can be concluded that EnOMT1 catalysed the formation of pinostrobin, a 7-*O*-methylated derivative of pinocembrin.

Although we successfully isolated and expressed a flavanone OMT, the expression of these *Eucalyptus* OMTs in *E. coli* was inconsistent. Re-cloning of the genes into the pHUE vector significantly improved the expression and solubility of the recombinant proteins. To confirm that the negative results of other OMT candidates tested were not due to inconsistent expression, all candidates were cloned into the pHUE vector and re-assayed. Dot blot analysis demonstrated the effectiveness of expression (**Figure 2.24**).

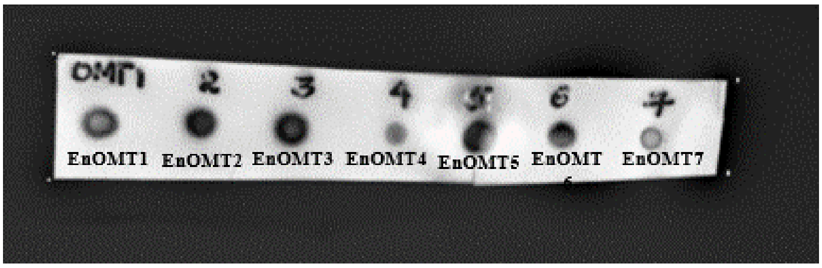


Figure 2.24: Dot Blot analysis of all EnOMT clones showing successful expression in the soluble

fractions.

Despite the improved expression using the pHUE vector, the other clones failed to produce any methylated products. Enzyme activity of the soluble EnOMT1 protein was not maintained at −80 ˚C (**Figure 2.25**). Therefore, the enzyme was extracted freshly before each assay from stored *E. coli* pellets.

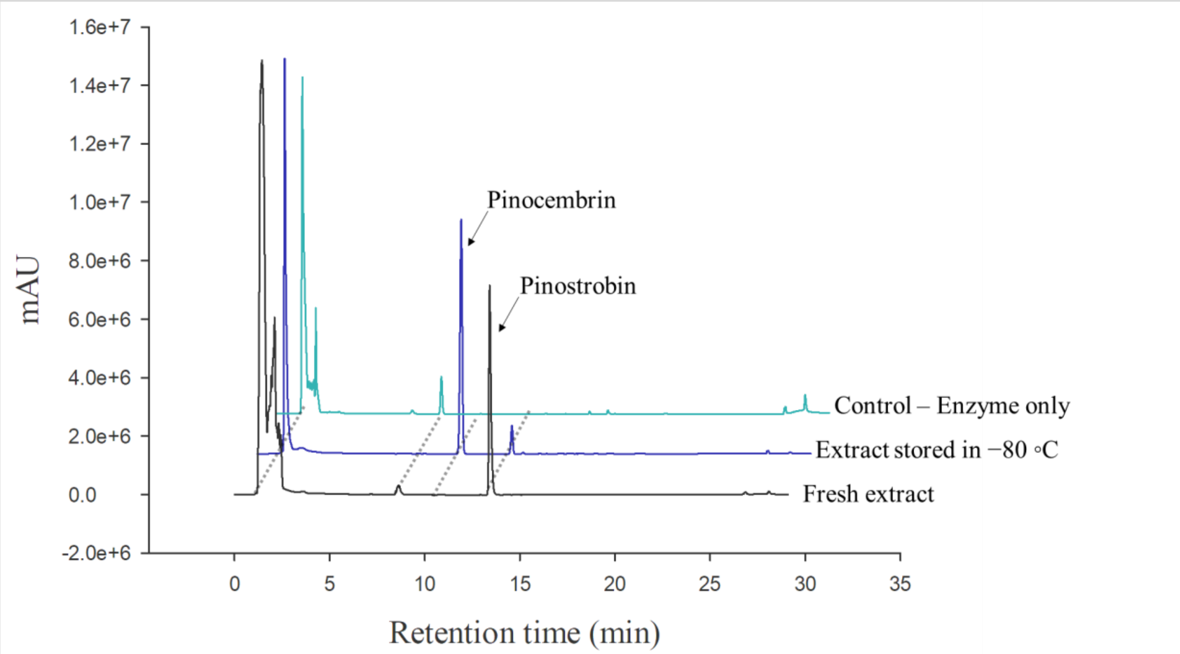


Figure 2.25: Decrease in activity of EnOMT1 after storage in −80 ◦C. HPLC chromatogram of EnOMT1 assay products using a freshly extracted enzyme (black), extract stored overnight in −80 ◦C (blue) and enzyme extract without substrates (cyan) showing decreased activity after storage All assays were carried with 100 µM pinocembrin for 1hr.

2.3.5 Sequence analysis using bioinformatics tools

The open reading frame of the isolated EnOMT1 gene encodes a 349 amino acid protein with a predicted molecular weight of 38.15 kDa (**Figure 2.26**). EnOMT1 contains conserved motifs which are likely to be involved in SAM/ substrate binding. Motifs 1 through 4 (Ibrahim et al., 1998) and motif A, and B (Joshi and Chiang, 1998) are conserved with at most two mismatches whereas motif 5, predicted to be the divalent cation binding region, and motif C have more mismatches.

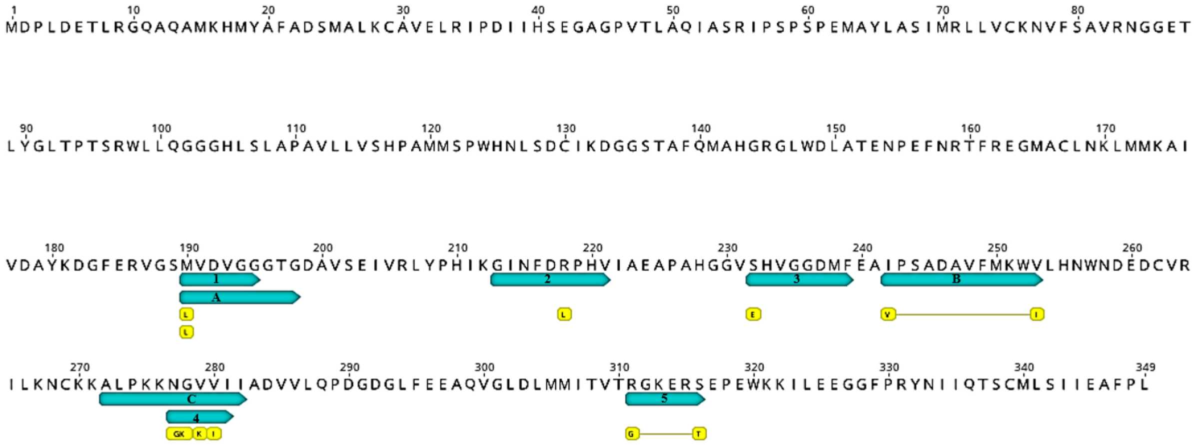


Figure 2.26: EnOMT1 full coding region with labelled conserved motifs A, B, C (Joshi and Chiang, 1998) and motif 1-5 (Ibrahim et al., 1998). Amino acid changes to that of the suggested motifs are highlighted in yellow.

The sequence of EnOMT1 is largely hydrophobic (**Figure 2.27**A) with 9% each of glycine, alanine and valine (**Figure 2.27**B). The enzyme EnOMT1 exhibited similarity to other plant OMTs that do not utilize divalent cations (Joshi and Chiang, 1998). The EnOMT1 protein sequence has a 53% amino acid identity with xanthohumol 4-OMT (Uniprot ID- B0ZB56.1) from *Humulus lupulus* (Nagel et al., 2008) and a 50% sequence identity with (R, S)-reticuline 7-OMT (Q6WUC2.1) from *Papaver somniferum* (Ounaroon et al., 2003). Alignment of EnOMT1 with all characterized (whose catalytic activity has been measured) flavonoid OMTs from other plant species is given in **Figure 2.28**. The maximum sequence identity to a flavonoid OMT is 40% for an isoflavone 7-OMT (AAC49927) from *M. sativa*. Additionally, the sequence identities of EnOMT1 to the other flavonoid 7-OMTs are 39% and 31%, respectively, for a flavonol 7-OMT (AY337457) from *M. piperita* and a flavanone naringenin 7-OMT (AB692949) from *O. sativa*.

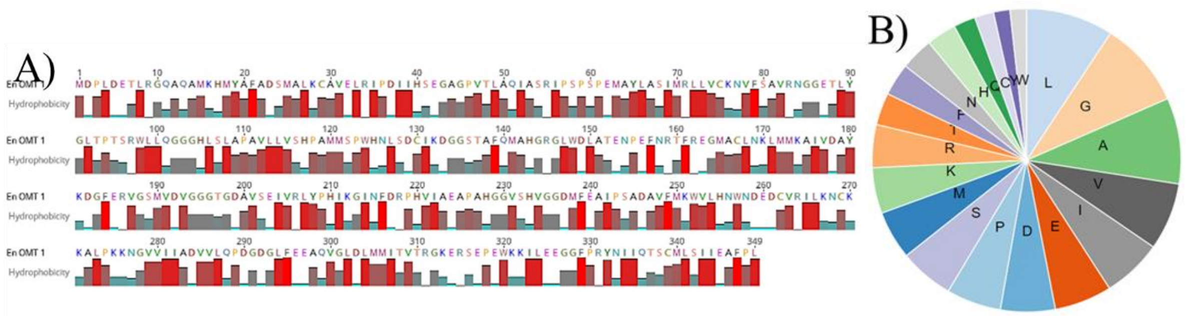


Figure 2.27: EnOMT1 coding sequence analysis. A) Amino acids colour coded based on the degree of hydrophobicity [red hydrophobic, blue polar]. B) Amino acid composition of EnOMT1.

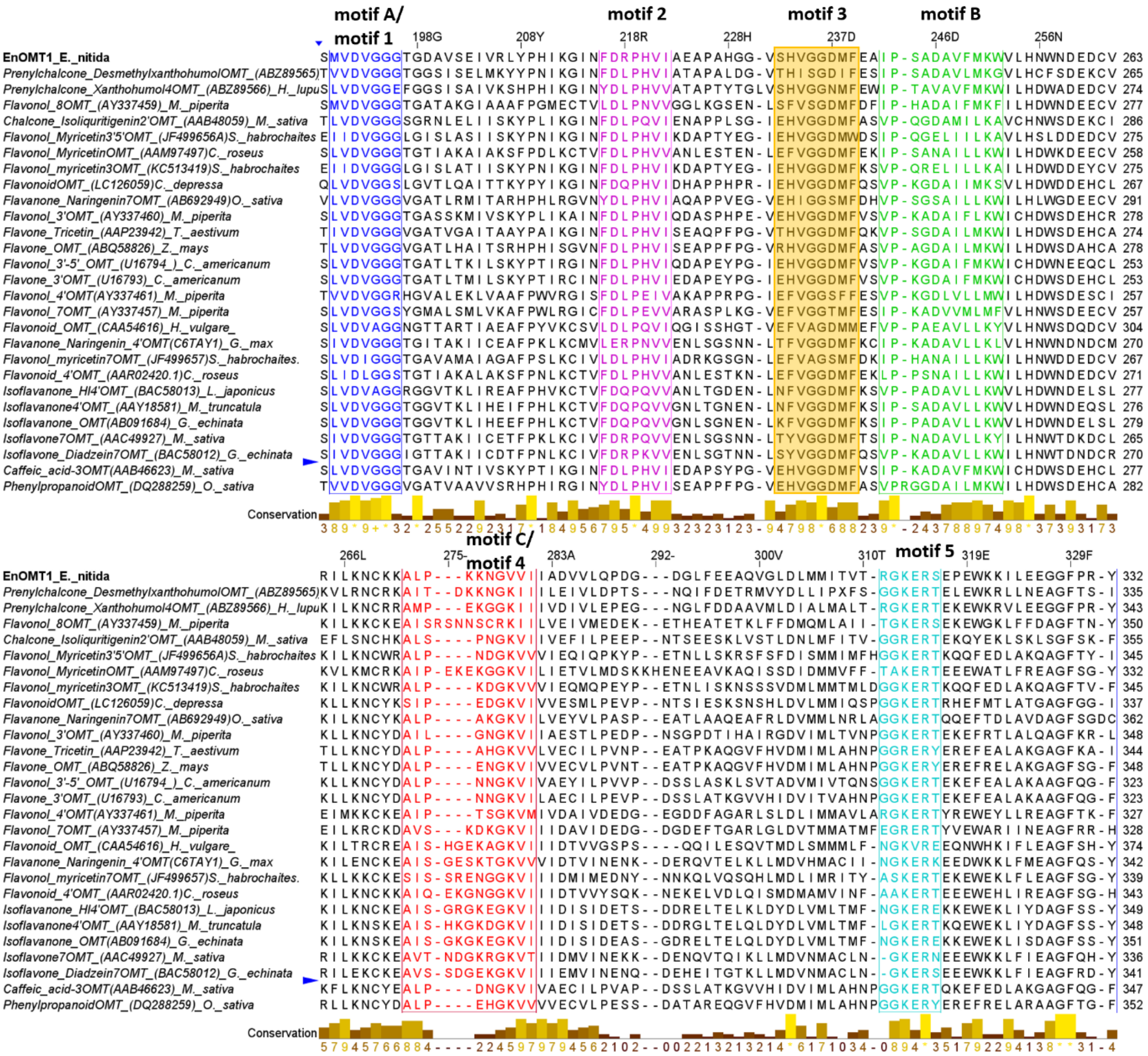


Figure 2.28: Alignment of EnOMT1 with functionally characterized plant flavonoid OMTs showing *C*-terminal amino acid identity and conserved motifs. Motif A, B, C (Joshi and Chiang, 1998) and motif 1-5 (Ibrahim et al., 1998).The last two sequences at the bottom belong to phenylpropanoid OMTs. Conservation of residues across the sequences is denoted by scores ranging from 1- 10 where score of 10 represents high conservation.

2.4 Discussion

This chapter documents the successful isolation of a gene encoding flavanone *O*-methyltransferase EnOMT1, which catalyses the biosynthesis of pinostrobin from the substrate pinocembrin. An initial objective of the work was to identify two species that exhibit a contrasting profile of unsubstituted B-ring flavanones, as the strategy for gene isolation was transcriptome sequencing and correlation between flavanone content in the leaves and transcript level. Based on the RP-HPLC analysis, the flavanone profile of *O-methylated species 1* matched with those observed in Goodger et al. (2016) with the presence of mono- and di-methylated flavanones. Similar to the *non-methylated species* several species of *Eucalyptu*s that only accumulate high amount of pinocembrin have also been reported previously (Goodger et al., 2019). Whether the difference in the phenolic profile is due to the lack of efficient enzymatic machinery responsible for methylation or to the differential expression of these genes between species, will be investigated in chapter 3.

2.4.1 Comparative transcriptomic analysis is an effective tool for gene isolation from *Eucalyptus* species.

In this study, seven candidate OMTs from *O-methylated species 1* were identified through RNA seq analysis, and one (EnOMT1) was shown to act on the unsubstituted B-ring flavanone, pinocembrin. The homology-based screening approach for isolation of similar genes using gene sequence information of previously isolated sequences is a popular technique for gene isolation. Using this technique, OMT clones that act on various flavonoid compounds have been isolated and characterised in different plant species. (Ibrahim and Muzac, 2000). For example, a multi-specific flavonoid OMT has recently been isolated from citrus by a >73% sequency identity to *A. thaliana* flavonoid 3'-OMT (Liu et al., 2020). Three out of seven candidates tested in this study, namely, EnOMT4, EnOMT5, and EnOMT6, were selected primarily based on high sequence identity to the flavonoid OMTs previously characterized from other genera (Lam et al., 2007). Nevertheless, none of them showed activity with flavanone compounds tested. Whether these clones are active on any other flavonoid and related compounds still needs to be investigated. Moreover, the isolated EnOMT1 coding sequence, shown in this chapter to act on flavanone pinocembrin, is more similar to alkaloid MTs than to flavonoid OMTs. Together, these results indicate that a homology-based approach alone may not be the most effective strategy in *Eucalyptus*.

This chapter focused on flavanone profiling of the two *Eucalyptus* species candidates and comparative analysis of their corresponding transcriptome using RNA-seq, rather than solely relying on homology-based gene isolation. Similar research by other groups using differential gene expression analysis by transcriptome sequencing of multiple plant parts or differentially treated individuals has enabled successful identification of genes controlling frost tolerance (Fernández et al., 2015), wood traits (Thavamanikumar et al., 2014), disease resistance (Santos et al., 2020), and pest resistance (Oates et al., 2015) in various *Eucalyptus* species. Several pieces of research have successfully utilized the approach of comparison of transcriptome \ data of multiple plant parts with metabolome analysis for identification and isolation of plant biosynthetic enzymes (Chen et al., 2017; He et al., 2018; Zhong et al., 2020). In the case of *Eucalyptus*, there exists a differential accumulation of flavanones in different plant parts with a major accumulation in foliar sub-dermal glands (Goodger et al., 2019). Considering the difficulty in extracting high-quality RNA from woody plant parts (Suzuki et al., 2003), RNA-sequencing of two *Eucalyptus* species with a contrasting metabolite profile is a better approach for future gene isolation research than a differential gene expression analysis of the transcriptome of different tissues.

The reads generated by Illumina sequencing were *de novo* assembled with the Trinity software program (Grabherr et al., 2011) even though a full genome sequence of *E. grandis* (uncharacterised) is available. RNA-seq technology with *de novo* assembly is generally used for transcriptome assembly of non-model organisms or plants without a reference genome (Farrell et al., 2014; Mizrachi et al., 2010). *De novo* assembly was chosen for the transcriptome assembly as a direct alignment of short reads to a reference genome may result in mismatches due to genetic variation as well as mapping to multiple locations (Robertson et al., 2010). Sequencing quality control produced about 95% clean reads in both species and the assembled transcriptomes had 67,728 and 50,301 unigenes in *O-methylated species* 1 and *non-methylated species*, respectively. Consistent with previous studies on *Eucalyptus* transcriptome sequencing and other non-model organisms, this report suggests that *de novo* assembly can be successfully utilised for gene isolation in non-model organisms (Li et al., 2014; Mizrachi et al., 2010). The assembled transcriptome of *O-methylated* and *non-methylated species* showed a higher mean unigene length (1240 bp and 829 bp respectively) and N50 scores (1816 and 1229 respectively) compared to previous *Eucalyptus* transcriptome assemblies (Annadurai et al., 2013; Fernández et al., 2015) which suggest better quality. In addition, several parameters reported in this work, including successful CDS and Pfam prediction of assembled unigenes, and high confidence BlastX similarity scores of unigenes to protein sequences databases stand as various evidence that substantiates the robustness of the assembled transcriptome.

2.4.2 Isolation of a flavonoid methyltransferase in *Eucalyptus*

We have successfully isolated the first flavanone MTs from *Eucalyptus* species. Several sequences within the *E. grandis* and *E. camaldulensis* genome were annotated as putative MTs based on the homology with characterised caffeic acid OMTs (COMT) and caffeoyl-CoA OMTs

(CCoAOMT) from other genera, but none have yet been characterised (Harakava, 2005). A cDNA (X74814) isolated by screening a xylem cDNA library from *E. gunnii* with tobacco COMT (Poeydomenge et al., 1994) was localised to lignifying cells of xylem and mid-rib of *Eucalyptus* leaves and is thus purportedly associated with lignin biosynthesis as a COMT (Hawkins et al., 2003). Similarly, gene silencing and overexpression of several putative CCoAOMT ESTs and cDNA clones of genus *Eucalyptus* showed an effect on lignin content but there is no biochemical characterisation data available for the proteins encoded by the cDNA/EST (Paux et al., 2004; Xiao et al., 2020). In addition, EnOMT1 reported here is the first flavanone OMT shown to be acting on the unsubstituted B-ring flavanone pinocembrin. Several OMTs acting on flavonoids have been characterised in other genera, of which three specifically act on flavanones with B-ring substitutions. For example, a 4'-OMT from *C. roseus* (AAR02420.1) acts on homoeriodictyol (Schröder et al., 2004) and a 7-OMT from *O. sativa* (AB692949) (Shimizu et al., 2012) and a 4'-OMT from *G. max* (Uniprot ID-C6TAY1) (Kim et al., 2005) acts on naringenin, but none of these enzymes have been tested against unsubstituted B-ring flavanones. The specificity of EnOMT1 towards various flavonoid substrates will be analysed in chapter 3.

The isolated EnOMT1 sequence encodes a 349 amino acid protein and has the hallmark SAM-binding sequences moderately maintained. Conserved motif analysis showed that Motifs 1 through 3 (Ibrahim et al., 1998) and Motif A, and B (Joshi and Chiang, 1998) are moderately conserved with at most two mismatches, whereas Motif C (overlapped by Motif 4) has the most mismatches. Although EnOMT1 did not require any divalent cations for activity and exhibited similarity to other cation-independent plants OMTs (Joshi and Chiang, 1998), Motif 5, which is associated with divalent cation binding (Ibrahim et al., 1998), is present with only two mismatches. The presence of this metal-binding motif suggests close evolutionary relatedness to cation-dependent phenylpropanoid OMTs. The spatial separation between the motifs in EnOMT1 is consistent with a suggestion by Joshi and Chiang (1998) with motif A separated from motif B by 52 amino acid residues and 20 amino acid between motif B and C. This indicates a highly conserved tertiary structure of the C-terminal region and suggests that homology modelling for structure prediction should provide reliable results.

Notably, the flavanone methyltransferase EnOMT1 isolated here has higher sequence similarity to prenylchalcone OMTs and alkaloid OMTs than to other flavonoid OMTs. The highest identity is with xanthohumol 4-OMT (53%) from *Humulus lupulus* which acts on xanthohumol, a prenylchalcone that accumulates primarily in peltate glandular trichomes (Nagel et al., 2008), and with (R, S)-reticuline 7-OMT (50%) from *Papaver somniferum* (Ounaroon et al., 2003), which is involved in alkaloid biosynthesis. EnOMT1 has a maximum of 40% sequence identity with the structurally characterised isoflavone-7OMT (IOMT) from *Medicago sativa* (Zubieta et al., 2001). The sequence of the equivalent OMT1 sequence (Cluster-13293.15022) of *non-methylated species*, with which we compared for candidate selection, is slightly different from that of the EnOMT1 sequence (94% sequence identity). Further investigation into whether this sequence difference has any role in the catalytic activity is described in Chapter 3.

2.5 Conclusion

Trees in the genus *Eucalyptus* are an ecologically important component of the native flora of Australia and are also economically important in many regions of the world. The trees have long been commercially exploited primarily for their hardwood timber, and to a lesser degree, for their essential oils. In more recent times there has been growing interest in other natural products of eucalypts, including their methylated and non-methylated flavonoids. Comparative RNA-seq transcriptome data analysis of two *Eucalyptus* species with distinct flavanone profile has resulted in the successful isolation of the first flavonoid methyltransferase from *Eucalyptus* as well as the first plant OMT acting on the unsubstituted B-ring flavanone, pinocembrin. Further biochemical, as well as structural characterization of EnOMT1, is described in Chapter 3. It would be useful to apply this comparative transcriptomics approach to other *Eucalypts* with differences in their natural product profiles. For example, this approach might help elucidating the biosynthetic network of various terpenoids or other non-volatile compounds like monoterpene acid glucose ester in *Eucalyptus*. Particularly, such an approach could enable isolation of CMTs from *Eucalyptus* provided there is a species with a contrasting profile accumulating *C*-methylated flavanones. This idea will be explored further in Chapter 4.

CHAPTER-3

CHARACTERIZATION OF A FLAVANONE 7- *O*-METHYLTRANSFERASE FROM *EUCALYPTUS*

3.1 Introduction

In plants, *O*-methylation is an important modification reaction involved in the biosynthesis of various metabolites including some flavonoids. Several OMTs that accepts flavonoid substrates have been isolated and characterised (**Section 1.3.2**; Appendix A). Plant OMTs are divided into two main categories: Class I contains cation-dependent, small molecular weight OMTs whereas Class II hosts larger, cation-independent OMTs (Lam et al., 2007). Both classes accommodate representative OMTs that catalyse flavonoid methylation (reviewed in **Section 1.3.2**).

Flavonoid OMTs in particular are a versatile group of enzymes. Majority show broad substrate specificity by accepting flavonoids of various subclasses. Some are multifunctional whereas others are strictly position-specific (Cacace et al., 2003; Itoh et al., 2016; Liu et al., 2020; Zhou et al., 2006). For example, flavonoid OMTs from *Citrus depressa* (CdFOMT5) and *Citrus reticulata* (CrOMT2) are multifunctional where CdFOMT5 catalyse 3-, 5-, 6- and 7-*OH* on a range of flavonols while CrOMT2 methylate 3'-, 5'-, and 7-*OH* on flavones and flavonols (Itoh et al., 2016; Liu et al., 2020). Additionally, flavonoid OMTs from *Catharanthus roseus* (CrOMT2*) and *Triticum aestivum* (TaOMT2) are multifunctional but demonstrate B-ring specific sequential methylation. CrOMT2* methylation is proposed to proceed in the order 5'-followed by 3'-*OH* while TaOMT2 catalysis was predicted to proceed through 3′-mono- → 3′,5′-di- → 3′,4′,5′-trimethyl ether derivatives.(Cacace et al., 2003; Zhou et al., 2006). Unlike these multifunctional OMTs, multiple flavonoid OMTs isolated from *Chrysosplenium americanum* are shown to be strictly position-specific (Barron et al., 2020). For example, four distinct OMTs are shown to respectively catalyse 3-, 6-, 7-, and 4′-OH of flavonols and among them 3-OMT was specific towards flavonols with 3′,4′ (or 4′,5′) hydroxylation pattern (De Luca and Ibrahim, 1985; Khouri and Ibrahim, 1987). Limited information is available on OMTs involved with flavanone substrates from tree species. A 7-*OH* specific OMT from *Populus deltoids* showed an altered regiospecificity for 4'-*OH* when a flavone substrate with an existing 3'-*OH* methylation was given as the substrate (Kim et al., 2008). Flavonoid OMTs from *Citrus* species responsible for poly-methoxylated flavones have been demonstrated to catalyse methylation of various hydroxyls on all three rings of flavones (Itoh et al., 2016; Liu et al., 2020). More studies are needed to confirm whether this lack of strict regioselectivity is a characteristic of tree species OMTs. Functional characterisation of an EnOMT1 from *Eucalypt* can therefore contribute to the existing knowledge pool and help explain the substrate specificities.

Structurally, plant OMTs are distinct from their monomeric mammalian counterparts as they are shown to be homodimers (Vidgren et al., 1994; Zubieta et al., 2001). In silico analysis of various characterised plant OMTs has yielded information regarding residues that are involved in catalysis; however, insufficient candidate proteins in each of the substrate/position categories make it difficult to generalise the results (Liscombe et al., 2012; Morris and Facchini, 2019).

As described in chapter 2, a relatively higher concentration of methylated flavanones in subgenus *Eucalyptus* makes it a potential commercial source of methylated flavonoids (Goodger et al., 2019). The essential basic step for any kind of commercialisation of a product or compound is understanding their biosynthesis by functional characterization of the enzymes involved. Therefore, the work described in this chapter aims to characterise the OMT identified in the previous chapter. First, biochemical characterization will be achieved through enzyme kinetic and substrate specificity analysis. Secondly, structural characterization will be accomplished through *in silico* techniques including homology modelling and molecular docking. Additionally, through isolation and biochemical characterisation of orthologues genes from different *Eucalyptus* species, it is intended to identify the specific residues involved in catalysis.

3.1 Materials and methods

3.1.1 Plant material

In addition to the two species used in Chapter 2, this study used an additional *Eucalyptus* species, which also belongs to subgenus *Eucalyptus,* namely*, E. croajingolensis* (hereafter *O-methylated species 2*). *O-methylated species 2* leaf samples were collected from plants grown at the University of Melbourne Dookie campus. Seeds of the species were purchased from Seeds of Gippsland (Woodside and Stradbroke provinces, Australia) and propagated in the University of Melbourne glasshouse. The seedlings were then planted at the Dookie campus in November 2016.

3.1.2 Gel electrophoresis and Western Blot

To confirm the molecular weight of the recombinant OMTs, denaturing gel electrophoresis was performed using ready to use SDS-PAGE gel (Abcam Australia) following the manufacturer's instructions. Native gel electrophoresis was performed using Mini-PROTEAN® TGX™ Precast Gels (BioRad) under native conditions following the manufacturer′s protocol. After separation, the protein bands were transferred onto a nitrocellulose membrane for antibody staining and detection via western blotting following the same protocol described for Dot Blot analysis. When required, gels after electrophoretic separation were stained with Coomassie Brilliant Blue R-250 staining solution (Bio-Rad) overnight and washed multiple times with distilled water to visualize protein bands.

Protein concentration in the crude lysate was measured by Bradford assay method (Bradford, 1976) using bovine serum albumin as the standard protein and the Bio-Rad protein assay concentrate reagent (Bio-Rad) and absorbance were measured using a NanoDrop spectrophotometer (Thermo Fisher Scientific).

3.1.3 Kinetic constants determination and substrate specificity assay

A range of phenolic compounds were analysed to determine the substrate specificity of EnOMT1 (**Table 3.2**). Enzyme kinetic parameters of EnOMT1 were calculated using pinocembrin as the substrate. The initial velocity (v_0) for each pinocembrin concentration (2 µM -200 µM), determined from the slope of the curve at the beginning of a reaction, was plotted against different substrate concentrations, and the apparent kinetic constants V_{max} (maximum velocity) and K_m (Michaelis constant/substrate concentration at half the maximum velocity) of EnOMT1 were determined using the Enzyme Kinetics add-on programme in SigmaPlot. The specific activity of the enzyme (U mg protein^{-1}) and catalytic rate constant (k_{cat}) were calculated from the maximum catalytic velocity of purified EnOMT1. For determining substrate specificity of EnOMT1, duplicate assays with the crude enzyme and varying substrate concentrations (2 µM - 50 µM) at a saturating concentration of 500 µM SAM were carried out for time intervals ranging from 3 to 10 min, where the reaction velocity is in the linear range with respect to time for EnOMT1 with pinocembrin as a substrate. Control reactions without EnOMT1 were run to determine the initial concentration of substrates. After RP-HPLC analysis, the peak area of the remaining substrate in the assay mixture was compared with that of the control and the difference in peak size area was used as the amount of substrate utilized during the reaction. For calculation of relative activity, the reactions at saturating concentrations (10 µM substrate, 500 µM SAM) were used. Relative activity was calculated by setting the amount of pinocembrin utilized by EnOMT1 as 100%.

3.1.4 Bioinformatic analysis of protein sequence

The molecular weight of the EnOMT1 protein was calculated using ProtParam (Gasteiger et al., 2005). Secondary structure analysis of the isolated proteins was investigated using online tools PSIPRED, ProteinPredict and Geneious software. PSIPRED (http://www.psipred.net) incorporates two feed-forward neural networks which perform an analysis on the output obtained from PSI-BLAST (Position-Specific Iterated-BLAST).

Functional effects of single amino acid substitutions were predicted with SNAP2 (https://www.rostlab.org/services/SNAP/), which considers the evolutionary information taken from an automatically generated multiple sequence alignment, predicted secondary structure and solvent accessibility and functional annotation, if available, of the sequence.

Tertiary structure analysis was performed using two online tools, namely, I-TASSER (https://zhanglab.ccmb.med.umich.edu/I-TASSER/) and Phyre2 (http://www.sbg.bio.ic.ac.uk/servers/phyre2/html/page.cgi?id=index). In I-TASSER, 3D models are built based on multiple-threading alignments by LOMETS (Local Meta-Threading Server) and iterative template fragment assembly simulations. I-TASSER was ranked as the most effective/accurate server for protein structure prediction in recent years (Zheng et al., 2019). Crystal structure of an isoflavone OMT (PDB: IFP2) and chalcone OMT (PDB: 1FP1) from *Medicago sativa*, S-norcoclaurine 6-*O*-methyltransferase (PDB:5ICC) from *Thalictrum flavum*, caffeic acid O-methyltransferase from *Lolium perenne* (PDB:3P9C) and coniferyl alcohol 9-*O*-methyltransferase from *Linum nodiflorum* (PDB:4E70) were used as templates for homology modelling. Predicted structure molecular data was visualised and analysed using a molecular graphics visualization tool - RasMol and Geneious Prime software with protein analysis plugin. Models predicted by different tools were compared by the sequence-independent protein structure comparison tool by TM-align (Zhang and Skolnick, 2005).

The phylogenetic relationship was investigated using Geneious Prime software with the PHYML plugin. The sequences used in the phylogenetic comparisons with the genes reported here were obtained from public databases GenBank and Uniprot. Details of the sequences used are listed in Appendix B. Multiple sequence alignment for phylogeny tree development was done using inbuilt tools: MUSCLE (Edgar, 2004) with a user-defined maximum number of iterations of 10 or Clustal Omega (Sievers and Higgins, 2018). Subsequent tree construction was done in Geneious tree builder by Neighbour-Joining (NJ) and also maximum likelihood (ML) algorithm with bootstrap resampling (1000 iterations) to assess the level of statistical support for each tree node. The rat catechol OMT (accession number: BC081850) was set as the outgroup. Evolutionary distances were computed and are in the units of the number of amino acid substitutions per site.

The ligand-binding site prediction in the modelled protein structure was carried out using a combination of COFACTOR (a program that employs global structural alignment; Zhang et al., 2017) and TM-align (Zhang and Skolnick, 2005) to identify possible binding templates and then employ local 3D motif matches to obtain the binding-site residues. Using the 3D model predicted using I-TASSER, protein-ligand interaction was analysed by docking pinocembrin into EnOMT1 with the help of AutoDock Vina tool (Trott and Olson, 2010) in UCSF Chimera (Pettersen et al., 2004).

3.1.5 Site-directed mutagenesis

Site-directed mutagenesis was achieved by a primer extension method that involves incorporating mutagenic primers in independent, nested PCRs before combining them in the final product (Reikofski and Tao, 1992). Oligonucleotide primers containing the chosen mutations were designed for three different sites of EnOMT1, hereafter referred to as region I through III (**Table 3.1**). Amino acids in the EnOMT1coding sequence were mutated towards the ExOMT1, the EnOMT1 orthologue sequence from *non-methylated species*. The region I-III primers contained nucleotide mismatches resulting in three amino acid changes each, and the other three primer pairs (W253G, V254I, N257D) resulted in single amino acid changes.

Briefly, in the first round of PCR, the EnOMT1 forward primer (*EnOMT1 F*; **Table 2.1**) and OMT mutated reverse primer (e.g. *EnOMT1_255R*) were used on EnOMT1 template, as were EnOMT1 mutated forward primer (e.g. *EnOMT1_255F*) and EnOMT1 reverse primer (*EnOMT1 R*). After the first round of PCR, the products were purified and combined to form a new template (the products hybridise due to overlapping regions). In the second PCR, with OMT1 forward primer (*EnOMT1 F Kpn*) and EnOMT1 reverse primer (*EnOMT1 R Not1*), the full-length EnOMT1 with incorporated mutations from the mutated primers was amplified and subsequently cloned into pHUE via restriction digest using the protocol described in section **2.2.13**. PCR conditions were the same described in section **2.2.6**.

Table 3.1: Oligonucleotide primers used for site-directed mutagenesis.

Primer name	Nucleotide sequence	Annealing temperature
EnOMT1_255F	TCATGAAGGGGATCCTGCACGATTGGAAT	77.7
EnOMT1_255R	ATTCCAATCGTGCAGGATCCCCTTCATGA	77.7
EnOMT1_300F	GCTATTCCACGAGGTACAAGCAGGGCTC	74.7
EnOMT1_300R	GAGCCCTGCTTGTACCTCGTGGAATAGC	74.7
EnOMT1_310F	GATGATCACGCTGAGCGGAGGTAAAGAG	71.0
EnOMT1_310R	CTCTTTACCTCCGCTCAGCGTGATCATC	71.0
EnOMT1_W253G_F	GTTTTCATGAAGGGGGTCCTGCAC	72.1
EnOMT1_W253G_R	GTGCAGGACCCCCTTCATGAAAAC	72.1
EnOMT1_V254I_F	TTTCATGAAGTGGATCCTGCACAATTGG	96.4
EnOMT1_V254I_R	CCAATTGTGCAGGATCCACTTCATGAAA	69.4
EnOMT1_N257D_F	GGTCCTGCACGATTGGAATGATG	71.5
EnOMT1_N257D_R	CATCATTCCAATCGTGCAGGACC	71.5

3.2 Results

3.2.1 Kinetic analysis of EnOMT1

Routine kinetic assays of EnOMT1 were undertaken with crude *E. coli* extracts following confirmation that extracts from *E. coli* cells harbouring an empty vector as a negative control showed no product formation with pinocembrin as a substrate (detectable on RP-HPLC) and no other ability to diminish pinocembrin concentration. EnOMT1 catalytic velocity estimates at a series of substrate concentrations showed that the enzyme follows Michaelis-Menten kinetics (Michaelis and Menten, 2007). The inclusion of bivalent cations (Mg^{2+}, $Cu^{2+}$and Ca^{2+}) did not affect the activity of the enzyme. At each substrate concentration tested, the amount of product formed increased approximately linearly with time. The assay duration was adjusted to maintain linearity at low substrate concentrations (**Figure 3.1**). **Figure 3.2** shows the Michaelis-Menten curve of EnOMT1, with the initial catalytic velocity (v_0; the slope of the curve at the beginning of a reaction for each substrate concentration) plotted against the substrate concentration.

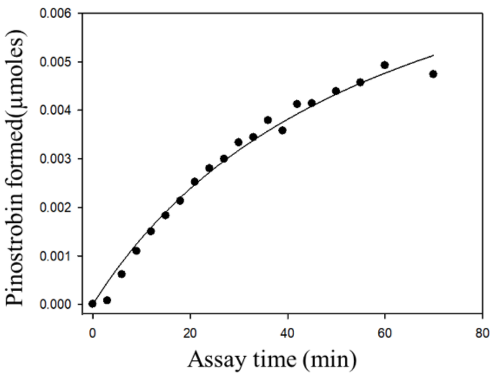


Figure 3.1: Rate of formation of pinostrobin catalysed by EnOMT1 at a substrate concentration of 300 µM pinocembrin with respect to assay time. Only the initial linear part of the curve was used to estimate catalytic velocity.

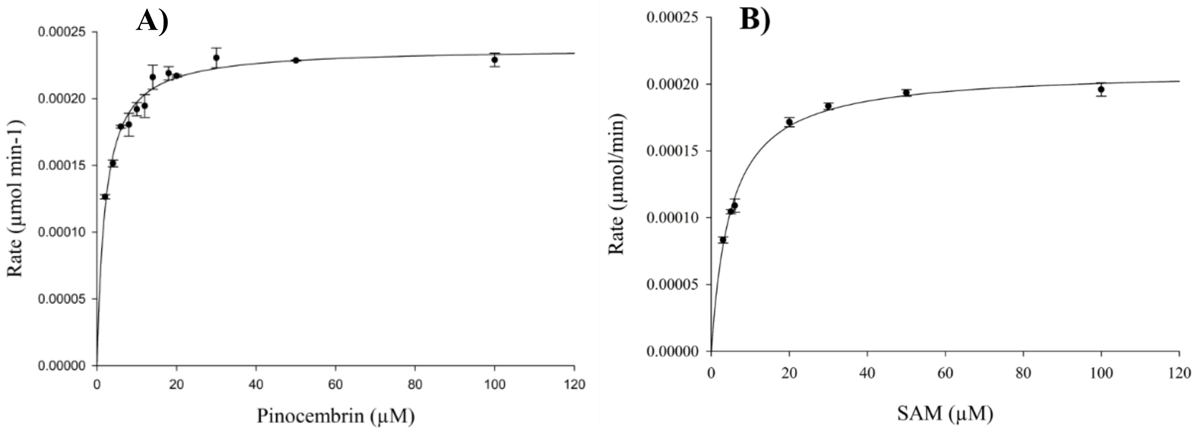


Figure 3.2: Michaelis-Menten curve of pinostrobin formed with EnOMT1 for pinocembrin (A) and SAM (B).

Apparent kinetic constants ($K_{m\ app}$ and $Vmax_{app}$) were calculated as the enzyme assays were conducted with crude/partially purified enzyme. With an excess SAM concentration (100 µM) and varying quantity of pinocembrin as a substrate, EnOMT1 showed an apparent $K_{m\ app}$ value of 2.05 ± 0.1581 µM. At a saturating concentration of pinocembrin (50 µM), EnOMT1 showed $K_{m\ app}$ of 4.99 ± 0.29 µmolar for SAM. The specific activity of the enzyme (U mg protein^{-1}) and rate constant (k_{cat}) were calculated by doing a short assay with substrate concentration at $K_{m\ app}$ with enzyme purified through His-tag affinity chromatography. The enzyme concentration after purification was estimated through Bradford assay. The enzyme's purity was estimated by comparing the staining density of the bands on SDS-PAGE with Coomassie blue staining (**Figure 3.3**B**)**. The size of the partially purified EnOMT1 protein estimated by SDS-PAGE **(Figure 3.3**A) is comparable to the calculated molecular weight of 38.15 kDa. The affinity-purified eluted fraction of EnOMT1 showed a specific activity 0.0033 U mg of protein^{-1} and has a catalytic rate constant (k_{cat}) of 0.1 s^{-1} and a catalytic efficiency (k_{cat}/K_m) of 0.05 $\mu M^{-1} s^{-1}$.

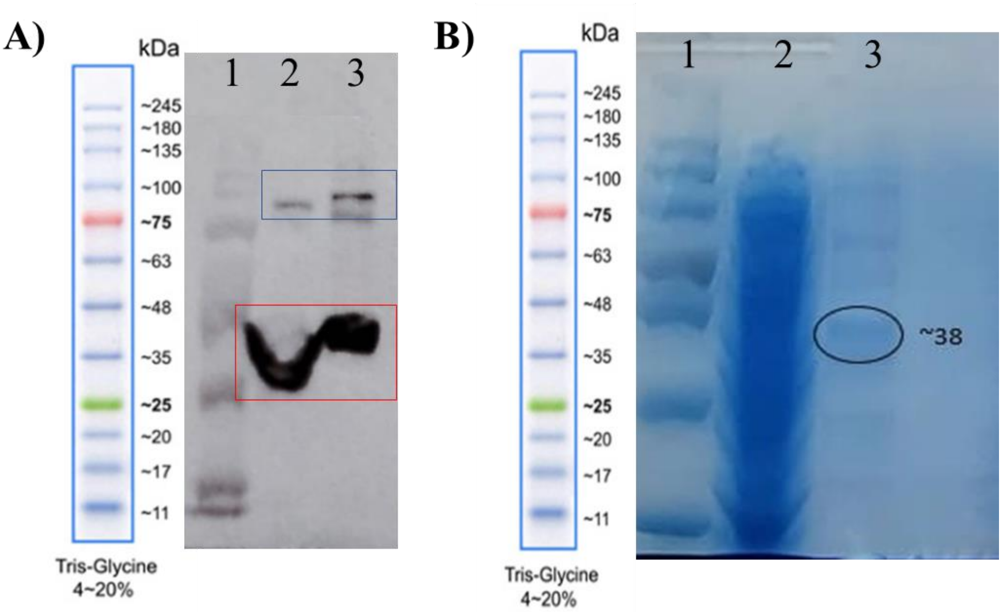


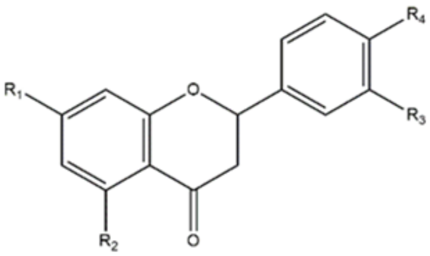
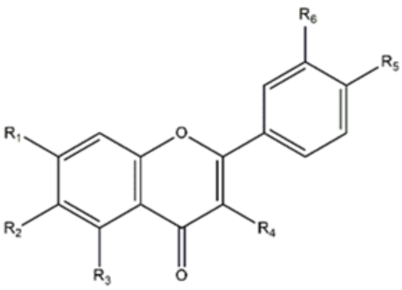
Figure 3.3: Western blot and SDS-PAGE analysis of EnOMT1. Lane 1-prestained protein ladder. Band separation in tris-glycine buffer is shown alongside for size reference; Lane 2- crude enzyme extract; Lane 3- purified His-tagged recombinant EnOMT1 in eluted fraction with 600µM Imidazole. A) Western blot analysis of fractions from affinity-purification of EnOMT1 under reducing conditions showing His-tagged recombinant EnOMT1 with anti-His antibody detection (red box). Bands with twice the molecular weight of the EnOMT1 are indicated by a blue box and may suggest homodimers. B)SDS-PAGE gel image of crude (lane 2) and purified EnOMT1 (lane 3) fractions showing protein bands after Coomassie-brilliant blue staining.

3.2.2 Substrate specificity

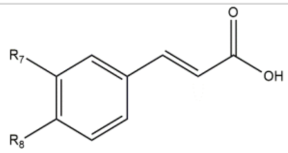
The catalytic specificity of EnOMT1 was analysed against a range of related phenolic compounds, to obtain substrate preference and apparent kinetic data. All flavonoids tested are known to be present in the genus *Eucalyptus* (Conde et al., 1992; Wollenweber and Kohorst, 1981). The structure of the compounds used in the substrate specificity analysis are summarised in **Table 3.2** and **Figure 3.4**. With SAM as a methyl donor, the flavanones alpinetin, naringenin, and hesperetin, and the flavones 7-hydroxy flavone, apigenin and chrysin showed an extra peak compared to that of the control without enzyme (**Figure 3.5**). Among the flavonoid compounds tested, no activity was observed with 5-hydroxyflavone, 6-hydroxyflavone, luteolin and quercetin. The range of chosen substrates for analysis was expanded out of the flavonoid category as there is evidence of flavonoid OMTs exhibiting a broad substrate specificity, accepting other phenylpropanoid compounds (Chiron et al., 2000; Gauthier et al., 1998).

Nevertheless, no activity was detected with caffeic acid, simple phenols, or other tested phenylpropanoids.

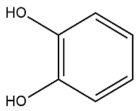
Table 3.2: Structure of flavonoid compounds tested for substrate specificity of EnOMT1.

Substrate	Structure	R_1	R_2	R_3	R_4	R_5	R_6
Alpinetin		OH	OCH_3	H	H	-	-
Naringenin		OH	OH	H	OH	-	-
Hesperetin		OH	OH	OH	OCH_3	-	-
5-HF†		H	H	OH	H	H	H
6-HF		H	OH	H	H	H	H
7-HF		OH	H	H	H	H	H
Apigenin		OH	H	OH	H	OH	H
Chrysin		OH	H	OH	H	H	H
Luteolin		OH	H	OH	H	OH	OH
Quercetin		OH	H	OH	OH	OH	OH

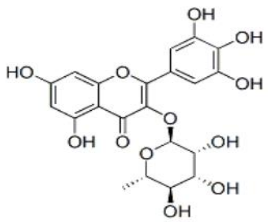
†HF - Hydroxyflavone.



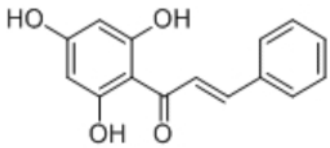
	R_7	R_8
Caffeic acid	OH	OH
p-Coumaric acid	H	OH
Trans- cinnamic acid	H	H



Catechol



Myricitrin



Pinocembrin chalcone

Figure 3.4: Structure of substrates outside the flavonoid aglycones category used for EnOMT1 substrate specificity assays.

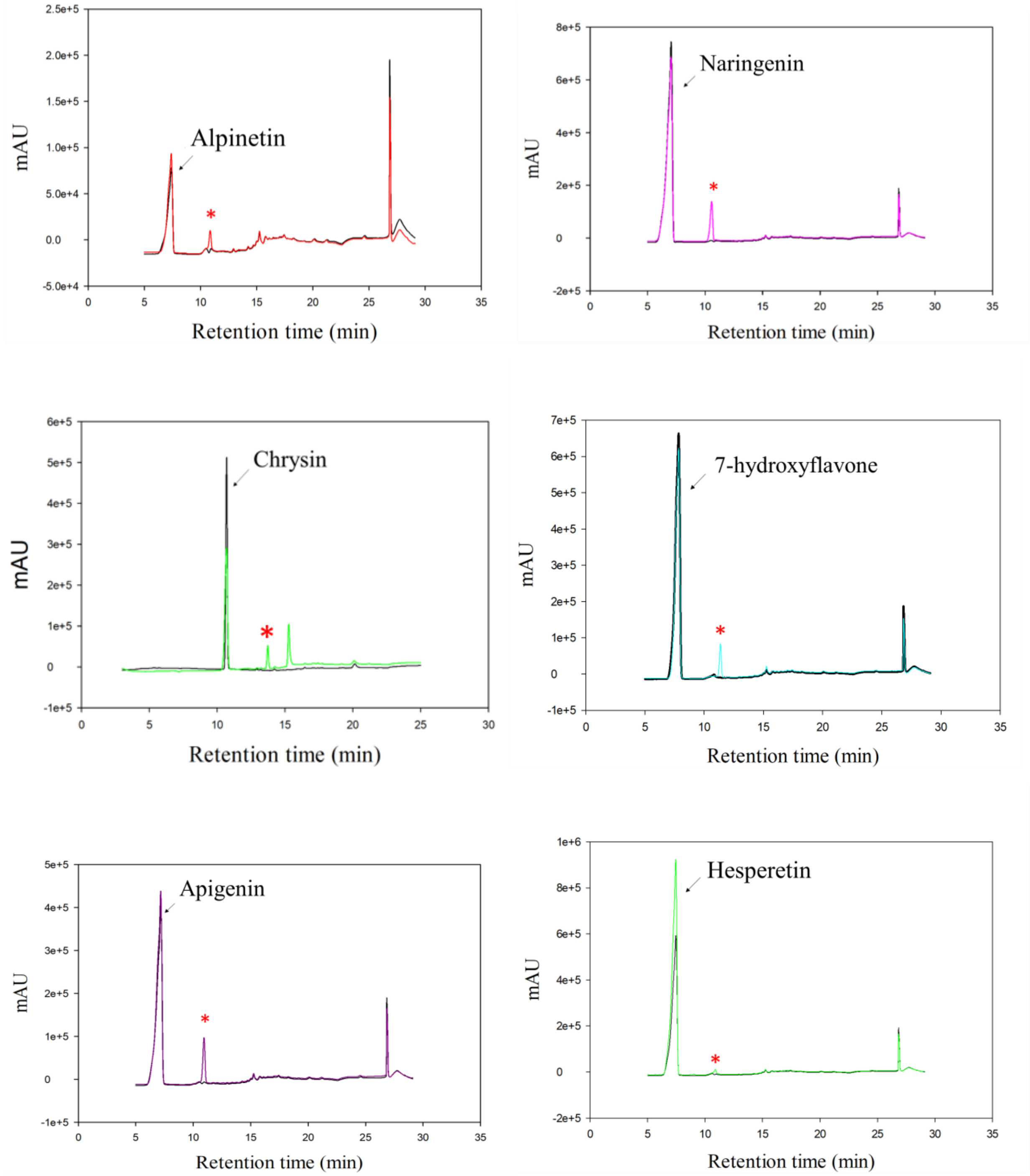


Figure 3.5: Assay products chromatogram of EnOMT1 with various flavonoid compounds as substrates (coloured) overlayed with the respective chromatogram of the control assay containing only the corresponding substrate (black) The peak marked with a red asterisk(*) denotes the corresponding methylated product.

Identities of newly formed methylated products were confirmed by comparing the retention time of the newly formed peak with that of authentic standards or by comparison with published information on relative retention times for methylated hesperetin (Qiao et al., 2012; Simirgiotis et al., 2015) and methylated chrysin (El-Hady et al., 2015; Markham et al., 1996).

Substrate specificity of EnOMT1 with seven flavonoid substrates are shown in **Figure 3.6**. A detailed kinetic analysis of EnOMT1 with two substrates chrysin and alpinetin was conducted. Chrysin was chosen because it showed a relative activity close to that of EnOMT1 with pinocembrin as a substrate; the two compounds are structurally similar, with an unsaturated bond in the C-ring the only difference. Alpinetin was selected due to its physiological relevance that it could be the intermediate compound in the formation of dimethylpinocembrin present in *O-methylated species 1*. The Michaelis-Menten (**Figure 3.7**) showed an apparent K_m value of 1.6 ± 0.09 µM for chrysin and 4.06 ± 0.26 µM for alpinetin.

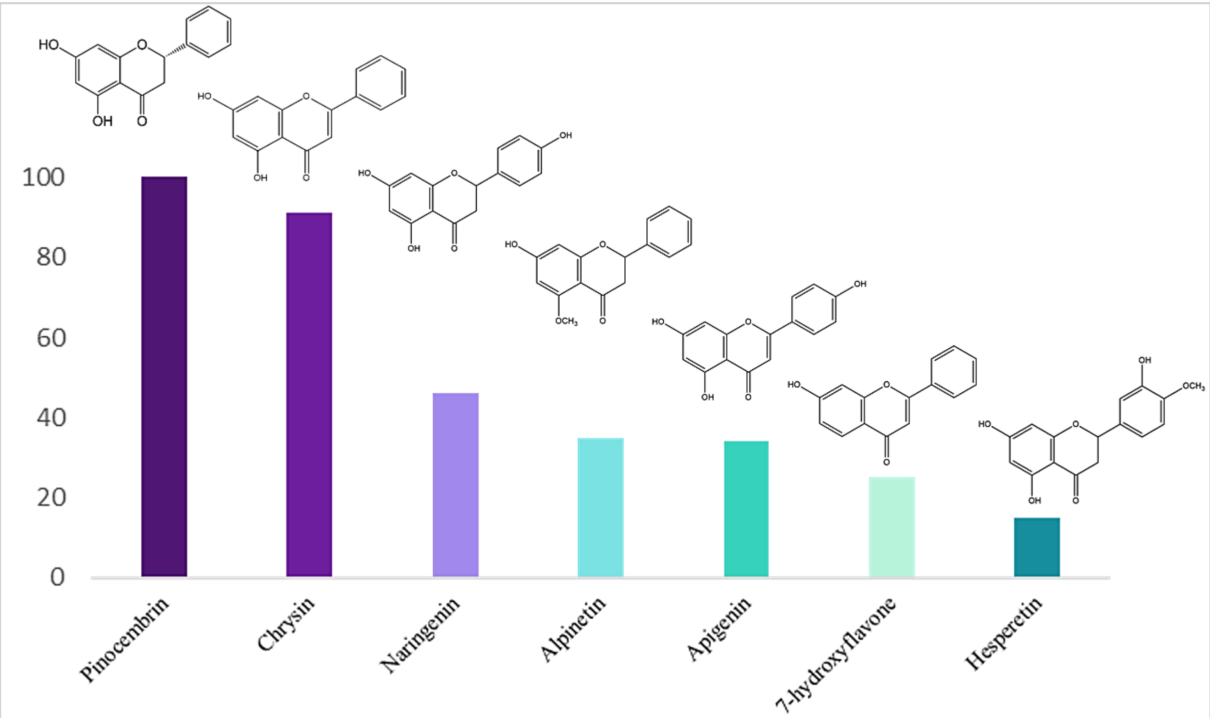


Figure 3.6: Relative activity (%) of EnOMT1 with other flavonoid compounds. Data were normalised to a pinocembrin activity of 100%. All substrates were at 10 µM and a saturating SAM (500 µM) was used in all assays.

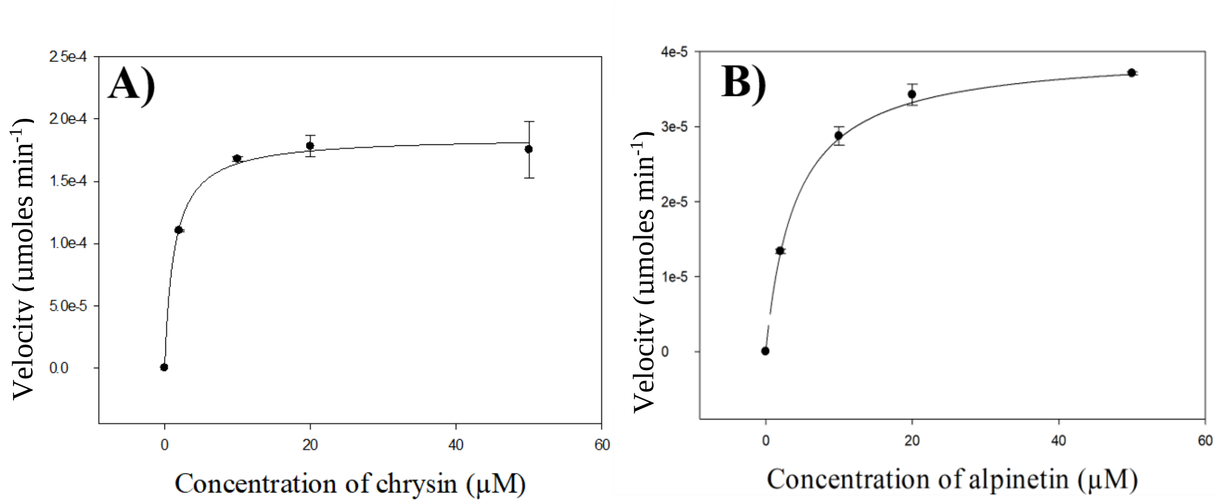


Figure 3.7: Michaelis-Menten plot of EnOMT1 activity against concentration of flavone chrysin (A) and flavanone alpinetin (B).

3.2.3 Isolation of EnOMT1 orthologue from *Eucalyptus croajingolensis*.

Many species in *Eucalyptus* subg. *Eucalyptus* have abundant amounts of methylated flavanones (Goodger et al., 2019). The driving force behind the accumulation of different complements of methylated flavanones in the different species is unknown. Successful isolation and characterization of pinocembrin 7-*O*-methyltransferase from *O-methylated species 1* in Chapter 2 demanded further interrogation into the active site chemistry of the enzyme to answer questions regarding how conserved the gene is across different species, and whether sequence comparison across species can shed some light into substrate binding as well as catalytic residues involved in methylation. To answer these questions, I attempted to isolate the EnOMT1 orthologues gene from another *Eucalyptus* species, *E. croajingolensis* (*O-methylated species 2)*, which had a similar flavanone profile to that of *O-methylated species 1* (**Figure 3.8**). Total RNA was successfully extracted from *O-methylated species 2* and subjected to subsequent cDNA synthesis. PCR utilizing the same oligonucleotide primers used for EnOMT1 isolation (*EnOMT1 F Kpn*; *EnOMT1 R Not1*; **Table 2.1**) and *O-methylated species 2* cDNA as template resulted in a band of approximately 1 kb. The amplified gene (hereinafter EcOMT1) showed 99.4% sequence identity to EnOMT1 with an additional glycine134 (highlighted in blue) insertion in EcOMT1 and substitution of amino acid valine to alanine on position 310 (highlighted in red, **Figure 3.9**).

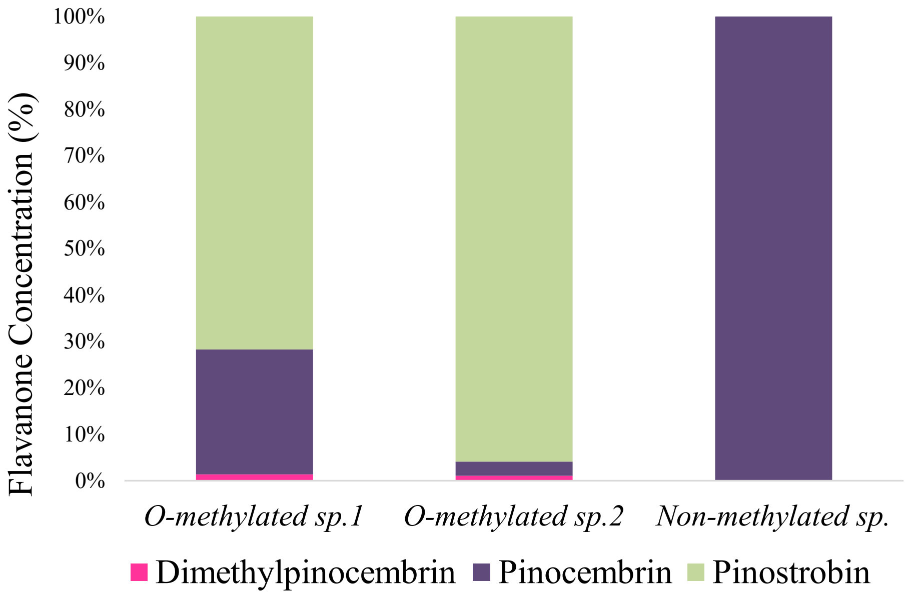


Figure 3.8: Flavanone profiles of *O-methylated species 1, 2 and non-methylated species* as a percentage of total foliar flavanones. Data are the average values of twelve individual trees. *E. croajingolensis* is named as *O-methylated species-2* for easy comparison with other species.

```
                           10        20        30        40        50
EnOMT1/1-349    1 MDPLDETLRGQAQAMKHMYAFADSMALKCAVELRIPDIIHSEGAGPVTLAQIASRIPSPS  60
EcOMT1/1-350    1 MDPLDETLRGQAQAMKHMYAFADSMALKCAVELRIPDIIHSEGAGPVTLAQIASRIPSPS  60

                           70        80        90       100       110
EnOMT1/1-349   61 PEMAYLASIMRLLVCKNVFSAVRNGGETLYGLTPTSRWLLQGGGHLSLAPAVLLVSHPAM 120
EcOMT1/1-350   61 PEMAYLASIMRLLVCKNVFSAVRNGGETLYGLTPTSRWLLQGGGHLSLAPAVLLVSHPAM 120

                          130       140       150       160       170
EnOMT1/1-349  121 MSPWHNLSDCIKDG-GSTAFQMAHGRGLWDLATENPEFNRTFREGMACLNKLMMKAIVDA 179
EcOMT1/1-350  121 MSPWHNLSDCIKDGGGSTAFQMAHGRGLWDLATENPEFNRTFREGMACLNKLMMKAIVDA 180

                          190       200       210       220       230
EnOMT1/1-349  180 YKDGFERVGSMVDVGGGTGDAVSEIVRLYPHIKGINFDRPHVIAEAPAHGGVSHVGGDMF 239
EcOMT1/1-350  181 YKDGFERVGSMVDVGGGTGDAVSEIVRLYPHIKGINFDRPHVIAEAPAHGGVSHVGGDMF 240

                          250       260       270       280       290
EnOMT1/1-349  240 EAIPSADAVFMKWVLHNWNDEDCVRILKNCKKALPKKNGVVIIADVVLQPDGDGLFEEAQ 299
EcOMT1/1-350  241 EAIPSADAVFMKWVLHNWNDEDCVRILKNCKKALPKKNGVVIIADVVLQPDGDGLFEEAQ 300

                          310       320       330       340       350
EnOMT1/1-349  300 VGLDLMMITVTRGKERSEPEWKKILEEGGFPRYNIIQTSCMLSIIEAFPL           349
EcOMT1/1-350  301 VGLDLMMITATRGKERSEPEWKKILEEGGFPRYNIIQTSCMLSIIEAFPL           350
```

Figure 3.9: Alignment of EnOMT1 with the orthologue from *O-methylated species 2* (EcOMT1). Amino acid insertions in EcOMT1is highlighted in blue and other amino acid substitutions are highlighted in red.

The coding sequence of EcOMT1 was successfully cloned and expressed in *E. coli* BL21 cells. With pinocembrin as substrate, EcOMT1 showed similar kinetic behaviour compared to

EnOMT1 with a $K_{m\ app}$ of 2.55 ± 0.3784 µM (**Figure 3.10**)

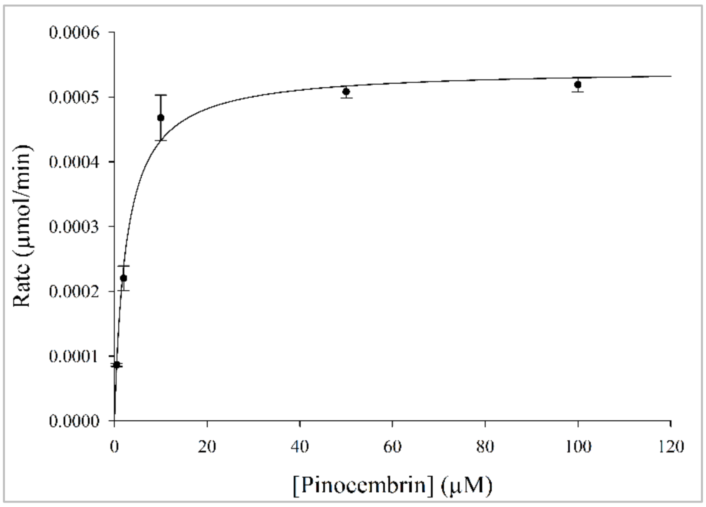


Figure 3.10: Michaelis-Menten curve of EcOMT1 with pinocembrin as substrate. Data points represent the initial velocities observed at different concentrations of pinocembrin at saturating SAM concentrations.

3.2.4 Isolation of EnOMT1 orthologue from *non-methylated species*

Non-methylated species accumulates the flavanone aglycone pinocembrin and lacks any methylated forms of pinocembrin at any stages of ontogenetic development (Goodger et al., 2019). The flavanone profile of *O-methylated species 1* and *non-methylated species* is shown in **Figure 3.8**. Given the difference in flavanone profile between *O-methylated species 1* and *non-methylated species,* it was of interest to investigate the activity of the OMT1 orthologue in *non-methylated species*.

A BLAST search of EnOMT1 sequence against *non-methylated species* transcriptome data revealed a homologous sequence with 94% sequence identity with a comparatively lower gene expression (**Section 2.3.2; Table 2.6; Appendix E**); however, a mildly reduced level of expression alone does not appear to explain the lack of *O*-methylation in *non-methylated species*. To investigate this issue, the EnOMT1 orthologue from *non-methylated species* (hereinafter ExOMT1), was cloned and expressed in *E. coli* BL21 cells. Alignment of the ExOMT1 protein sequence with EnOMT1 (**Figure 3.11**) showed 94% sequence identity. Just like EcOMT1, ExOMT1 was one amino acid longer than that of *O-methylated species 1* with an Alanine 247 insertion in ExOMT1 conserved motif B region (Joshi and Chiang, 1998). Motif A, C and 3 were highly conserved. . Most amino acid changes were in the C-terminal end of ExOMT1.

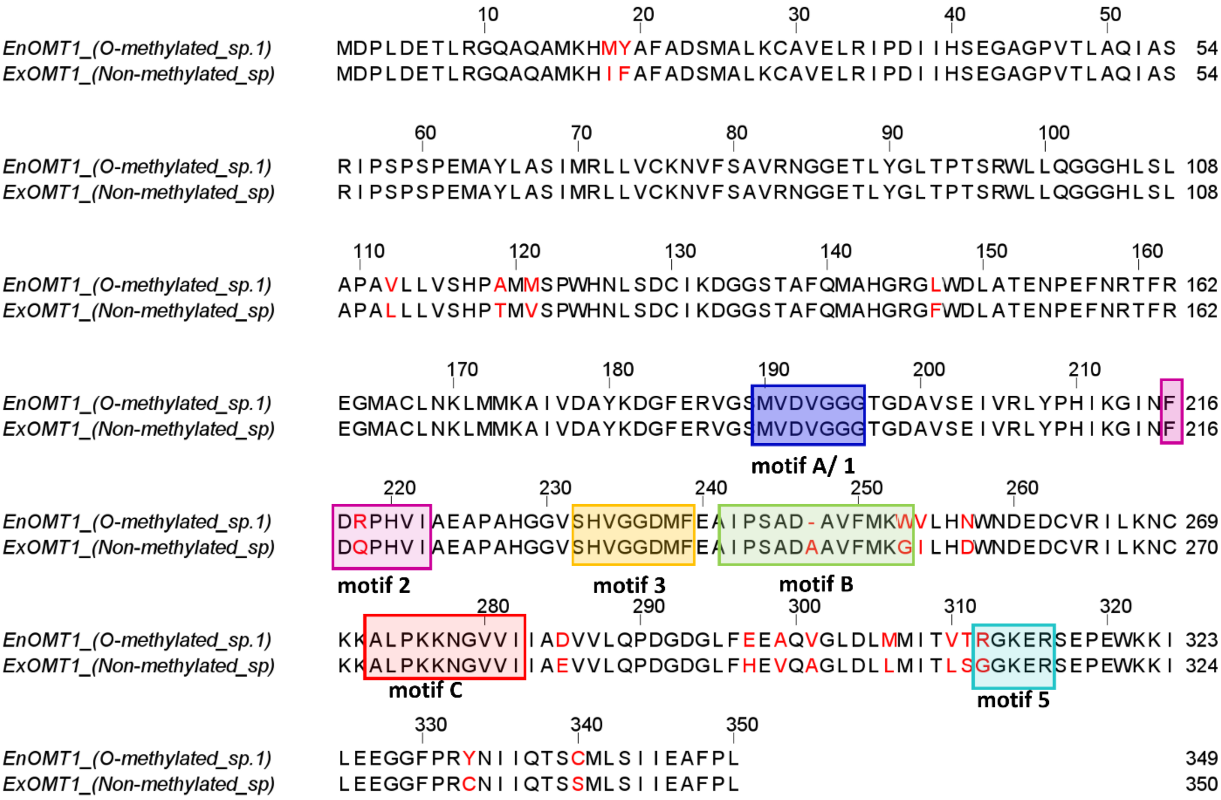


Figure 3.11: Alignment of EnOMT1 with the orthologue from *non-methylated species* (ExOMT1). The highlighted regions in red show amino acid changes between EnOMT1 and ExOMT1. Dash indicates gap. Conserved motifs are labelled by coloured boxes. Single amino acid changes were seen in motif 2 and 5 (Ibrahim et al., 1998). Motif B (Joshi and Chiang, 1998) had two changes including an amino acid insertion.

No methylated product was detected with ExOMT1 for the duration tested (3-10 min) with varying pinocembrin concentrations (2 -200 µM); however, a much longer incubation (~3 h) of the enzyme with 20 µM pinocembrin and 100 µM SAM resulted in the formation of a minor amount of methylated product, pinostrobin (**Figure 3.12**A). With same substrate concentrations EnOMT1 produced an equivalent amount pinostrobin after only three minutes of incubation (**Figure 3.12**B). An extrapolated non-linear regression analysis showed that the $K_{m\ app}$ is greater than 800 µM and the activity was two orders of magnitude less than that of EnOMT1. The striking reduction of activity can presumably be attributed to the sequence differences between EnOMT1 and ExOMT1. Contribution of these variant residues to the active site or substrate-binding pocket might impact the activity of the enzyme. Knowledge about the tertiary architecture of EnOMT1 is essential to understand the effect of these changes.

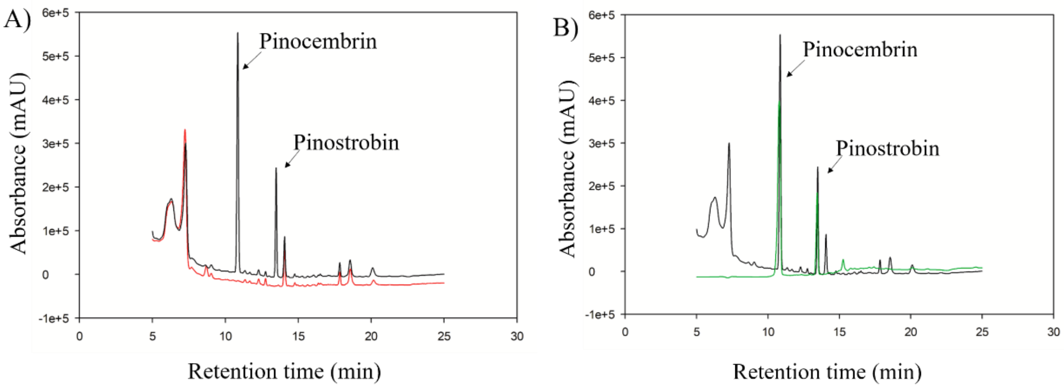


Figure 3.12: Pinostrobin production by ExOMT1. A) HPLC chromatogram showing the production of pinostrobin by ExOMT1 (black) after ~3 h incubation with 20 µM pinocembrin and saturating SAM concentration overlayed with control assay with empty vector extract (red). B) Overlayed HPLC chromatograms of ExOMT1 3h assay product and EnOMT1 five minute assay product, both with 20 µM pinocembrin and 100 µM. EnOMT1 formed an equivalent pinostrobin amount in just three minutes of incubation.

3.2.5 Phylogenetic analysis of *Eucalyptus* OMTs

A phylogenetic tree was constructed with Geneious Prime plugin PHYML, using 1000 bootstrap replicates, with rat catechol OMT as the outgroup. Candidates used in this evolutionary analysis was restricted to functionally characterised flavonoid and phenylpropanoid OMTs and two representative alkaloid NMTs. The same data set was subjected to Neighbour-Joining analysis (**Figure 3.13**) within the Geneious Prime package, using 1000 bootstrap replicates and the Jukes-Cantor substitution model. Virtually the same tree topology was produced if nodes of less than 80% were collapsed. Minor variations were seen in the substitution values on a few of the tree nodes.

The phylogenetic tree represented as cladogram showed separation into four distinct clades. Clade 1 contained Pl-OMTI class of cation dependant OMTs. Clade 2 represented the SABATH family of carboxyl-MTs. Two alkaloid NMTs from *Coffea canephora* were also clustered into clade 2. OMTs that belonged to the divalent cation independent Pl-OMTII class (Joshi and Chiang, 1998) showed a clear separation into two different clades (Clade 3 and 4). Although EnOMT1 was grouped with other flavonoid 7-OMTs in clade 3, a prenylated chalcone OMT was shown to be the closest evolutionary relative. All SAM dependant OMTs share specific regions in their amino acid sequence that are conserved throughout the evolutionary journey (Joshi and Chiang, 1998; Martin and McMillan, 2002; Vidgren et al., 1994). The presence of these conserved motifs in EnOMT1 (section **2.3.5**) and phylogenetic analysis showing close relatives of EnOMT1 belonging to class II plant OMTs point towards the likelihood of comparative modelling for enzyme structure prediction.

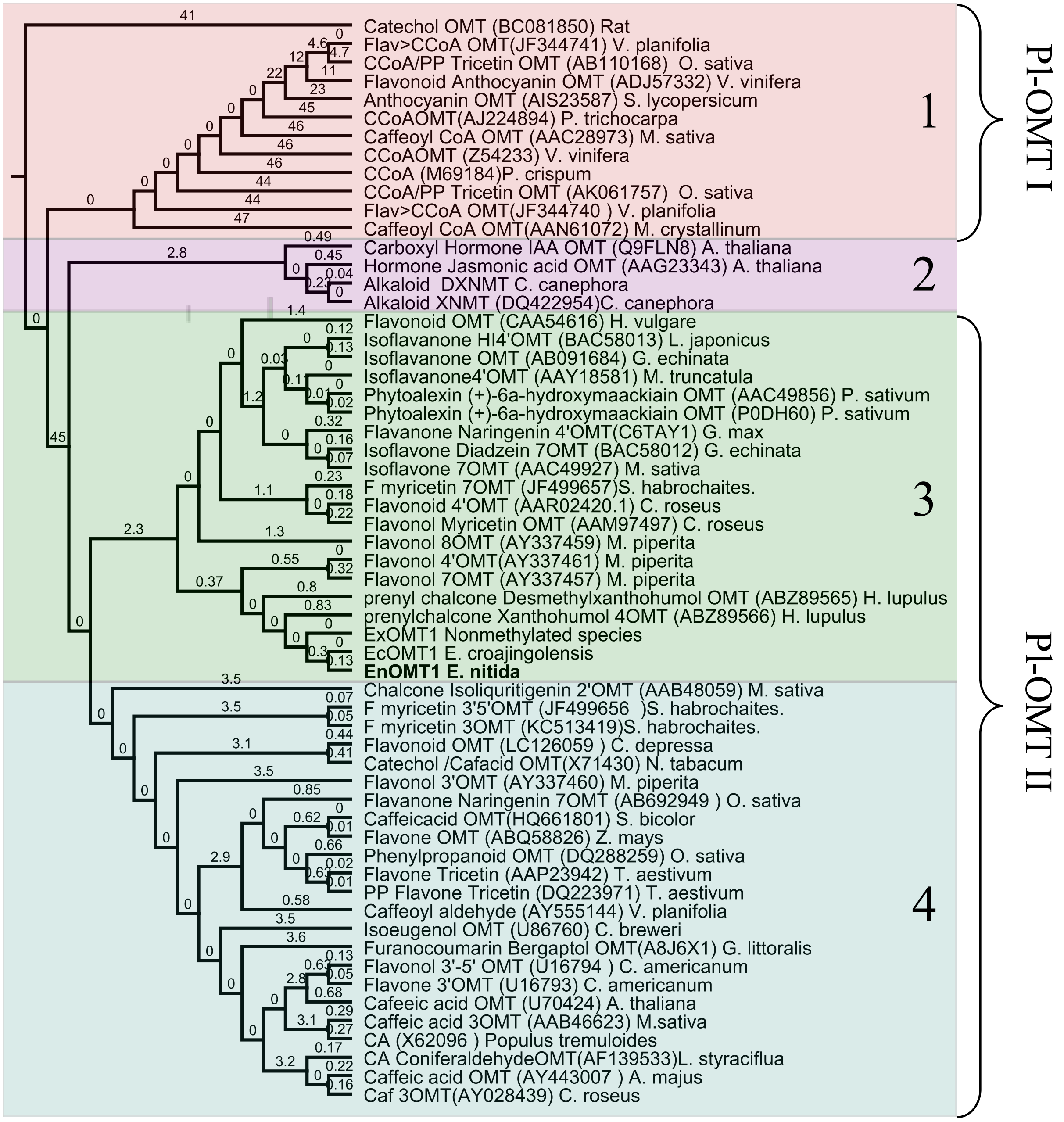


Figure 3.13: Phylogenetic relationship of *Eucalyptus* OMTs with other functionally characterized plant OMTs. Rooted similarity tree of plant flavonoid OMT proteins was constructed using the Neighbour-Joining method. The numbers at the branch denote phylogenetic distance of as amino acid substitutions per site. Branch labels contain Uniprot/GenBank accession numbers of the respective proteins in brackets. A consolidated list is given in Appendix A.

3.2.6 Protein structure analysis

The secondary structure of EnOMT1 was predicted using various bioinformatic tools, namely, PSIPRED, I-TASSER and PredictProtein. Predicted results were in close agreement with each other showing a composition of 44% helix and 9% β-sheets. The output of PSIPRED with the reliability indices (in 0–1 range) for all the three secondary structure states (H=helix, E=strand and C=coil) for each residue in the protein sequence is depicted in **Figure 3.14**. No transmembrane domains were identified in the protein. Subcellular localization prediction yielded no significant results. There were no nuclear localization signals, but a few amino acids (D182, E185, R186, K276, G291, R331) were predicted to be involved in protein-protein interaction (Yachdav et al., 2014). Solvent accessibility studies showed that only 45% of residues are exposed with more than 16% of their surface and the rest of the protein is buried and inaccessible.

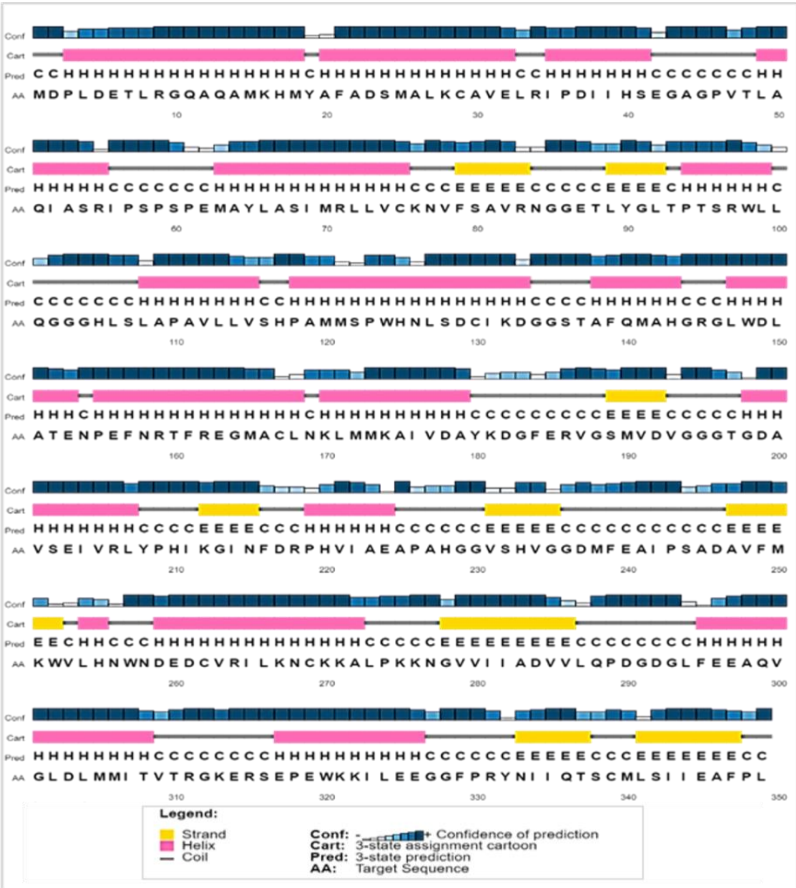


Figure 3.14: EnOMT1 coding sequence analysis showing predicted secondary structures of EnOMT1 using PSI-PRED bioinformatics tool. Conf stands for confidence of prediction colour coded from light to dark blue where dark blue corresponds to high confidence; Predicted secondary structures (Pred) corresponding to the amino acid sequence of EnOMT1 (AA)are depicted as cartoons (Cart) where helix region is denoted by H and pink rectangular boxes; E stands for β-strands and is shown as yellow rectangles. C stands for coil shown as grey lines. Protein structure prediction by homology modelling.

Protein 3D structure is very important to understanding function, interactions, and cellular localization. Therefore, the structure of EnOMT1 was predicted using three homology modelling packages: I-TASSER (Roy et al., 2010), Phyre2 (Kelley et al., 2015) and Swiss-Model (Waterhouse et al., 2018). The objective of using three different model-building tools was to compare and predict the accurate alignment to a 3D model. The quality of the different predicted models was compared using two different scores. The first is the C-score or confidence score, which measures the quality of predicted model, typically in the range of -5 to 2 where a higher value signifies a model with high confidence and vice-versa. The second is the TM-score, which measures the similarity of two protein structures, and has values between 0 and 1, where 1 indicates a perfect match between two structures.

I-TASSER, which ranked highest for predicting spatial structures of proteins over the last 10 years of experiments using CASP (Critical Assessment of protein Structure Prediction), gave the tertiary and quaternary structure model for EnOMT1 with a C-score of 1.07. Upon protein sequence submission to I-TASSER, top threading templates for query protein sequence were automatically identified by LOMETS threading meta-server (**Table 3.3**) and used to generate models from which the top five models, according to a C-score ranking, were kept as structure models. These were compared with solved homologous structures from the Protein Data Base (PDB) using the structural alignment program TM-align.

Table 3.3: The top four templates used by I-TASSER for threading alignment.

Rank	PDB hit	Iden1*	Iden2†	Cov‡	Norm Z-score§
1	3p9cA	0.33	0.35	0.99	2.64
2	5iccA	0.41	0.41	0.99	4.42
3	4e70	0.36	0.37	0.98	1.42
4	3reoA	0.35	0.37	0.98	6.21

*Iden1 is the percentage sequence identity of the template used by I-TASSER. †Iden2 is the percentage sequence identity of the whole template chain with a query sequence. ‡Cov represents the coverage of the threading alignment and is equal to the number of aligned residues divided by the length of query protein. §Normalised Z-score of the threading alignments. Alignment with a Norm Z-score >1 mean a good alignment and vice versa.

After comparison, model 1 (**Figure 3.15**A) was selected as the best-predicted model with a C-score of 1.07, TM-score of 0.86 ± 0.07, and an RMSD of 4.7 ± 3.1 Å. The top ten proteins from the PDB that have the closest structural similarity (i.e., the highest TM-score, to the predicted I-

TASSER model 1) are listed in **Table 3.4**. The programme identified norcoclaurine-6-*O*-methyltransferase (PDB ID: 5ICC), which is involved in benzylisoquinoline alkaloid synthesis as a best structural analogue of the top-scoring model of I-TASSER with the TM-score of 0.942.

In the second trial, isoflavone *O*-methyltransferase (PDB ID: 1FP2), a flavonoid OMT with a known crystal structure and having 32% sequence similarity to OMT1 was specified as a preferred template for homology modelling. The final model of this trial, model 2, had a C-score of 1.69 and a TM-score of 0.95±0.05. The score with higher value reflects a model of better quality. The structure was superimposed with Model 1 to check fold similarity and showed a similar fold with a TM-score of 0.91.

Table 3.4.Top 10 structural analogues of I-TASSER modelled EnOMT1 in PDB.

Rank	PDB Hit*	TM-score†	RMSD‡	IDEN§	COV¶
1	5ICC	0.942	1.76	0.403	0.991
2	1KYZ	0.942	1.50	0.355	0.977
3	3P9I	0.939	1.77	0.336	0.986
4	5CVU	0.935	1.75	0.379	0.989
5	4E70	0.921	2.03	0.344	0.989
6	5XOH	0.906	1.83	0.364	0.966
7	6I5Q	0.903	1.77	0.370	0.957
8	6NEG	0.894	2.13	0.345	0.969
9	1FP2	0.888	2.36	0.393	0.974
10	1FP1	0.881	1.86	0.314	0.937

*PDB Hit is the protein database identifier of the structural analogue;†TM-score measures the similarity of two protein structures, and has values between 0 and 1, where 1 indicates a perfect match between two structures; ‡RMSD is the root mean square deviation between residues that are structurally aligned by TM-align; §IDEN is the percentage sequence identity in the structurally aligned region; ¶COV is coverage of the alignment by TM-align and is equal to the number of structurally aligned residues divided by the length of the query protein.

Using the second modelling tool, Phyre2 (Kelley et al., 2015), 99% of residues were modelled with >90% confidence (**Figure 3.15**B). Superimposition of the two structural models predicted by I-TASSER and Phyre2 using TM align (Zhang and Skolnick, 2005) yielded a TM-score of 0.95. Out of 349 residues, 347 were aligned. In contrast, the superimposed structures of the Phyre2 model with the I-TASSER models show deviation in N-terminal helix fold (**Figure 3.15**C).

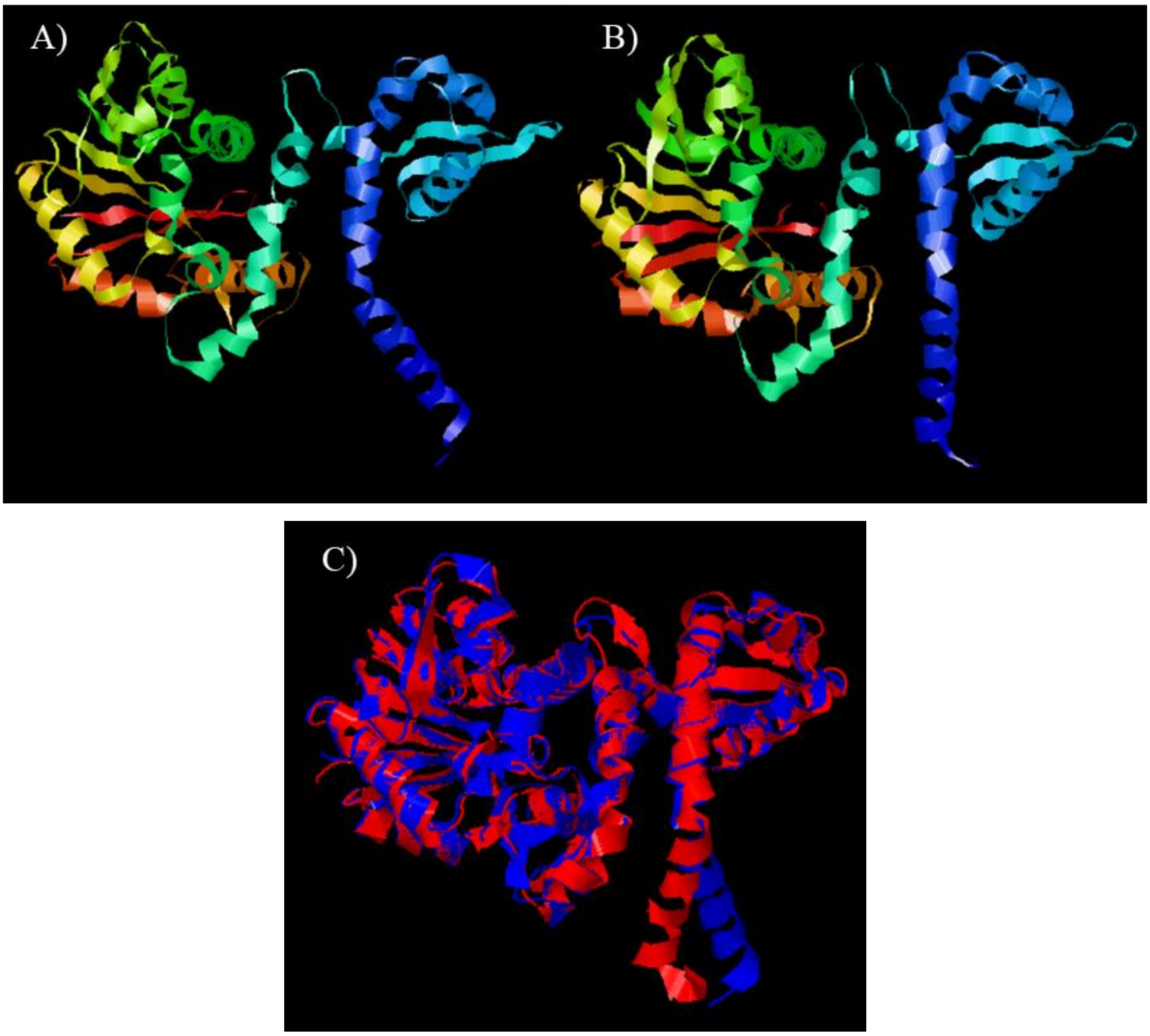


Figure 3.15: Predicted tertiary structure models of EnOMT1 using A) I-TASSER and B) Phyre2. Superimposed structural models predicted by the two different software applications (C) where blue represents I-TASSER model and red represents Phyre-2.

This may be because, to adopt the topology predicted, the enzyme was anticipated to have a small N- terminal transmembrane helix involving amino acids in position 20 to 30. MEMSAT-SVM (Nugent and Jones, 2012) gave a similar result indicating the possibility of a pore-lining region involving amino acids in position 17 to 32. Nevertheless, none of the other prediction software specific for transmembrane prediction, namely TMHMM (Krogh et al., 2001), TMpred and DAS (Cserzö et al., 1997), gave a positive result for transmembrane in that region in protein.

In all predicted models, EnOMT1 was predicted to be a homodimer. Most plant OMTs are known to form a homodimer in solution (Liscombe et al., 2012; Morris and Facchini, 2019). To confirm this, native protein gel electrophoresis was conducted with crude as well as a purified recombinant EnOMT1enzyme. Repeated trials were not very successful as the enzyme under native condition did not migrate into the Bio-Rad precast Mini protein gel (4 - 20% PAGE). The protein pI was estimated as 5.8, so at pH 6.8 (sample buffer) or pH 8.3 (running buffer), it should theoretically be negatively charged. The reason behind this is unclear. After several combinations of sample and running buffer pH, the EnOMT1 finally ran into the gel but with a varying Rf value (**Figure 3.16**). The predicted quaternary structure of EnOMT1 showing a homodimeric architecture is shown in **Figure 3.17**.

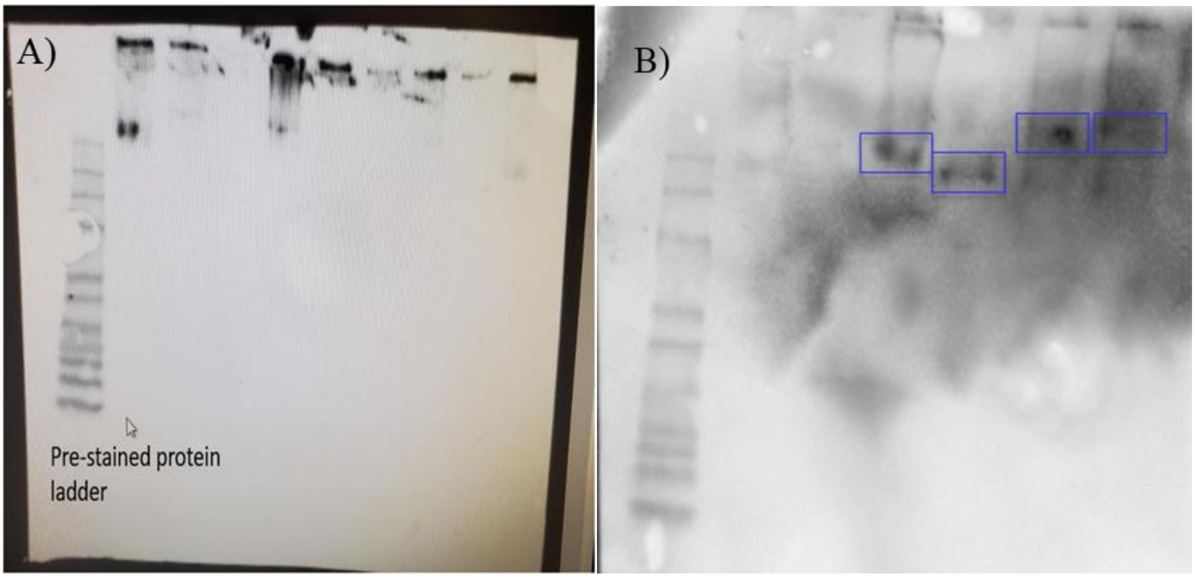


Figure 3.16: Native PAGE electrophoresis separation of recombinant EnOMT1 purified protein after western blot analysis using anti-His-antibody detection. A) The outcome of a trial following the manufacturer's instruction. B) After changing sample buffer pH, the bands inside the blue box indicate recombinant EnOMT1. In both panels, the well next to the protein ladder contains crude extract of EnOMT1 and all others contain purified fractions of EnOMT1.

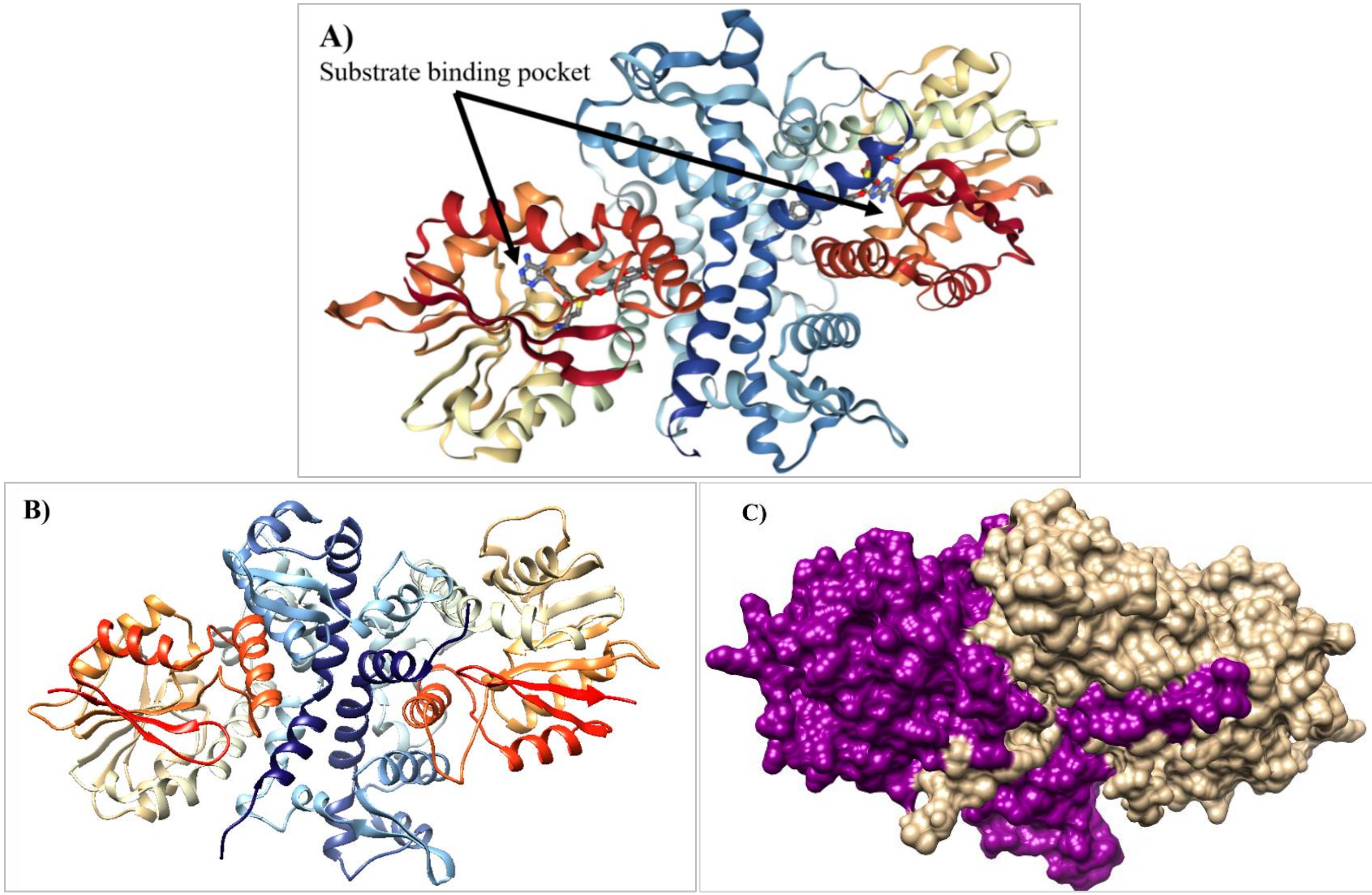


Figure 3.17 Quaternary structure comparison of EnOMT1 and Isoflavone OMT showing overall structure conservation and homodimeric formation. A) Quaternary stricture of Isoflavone OMT from *Medicago sativa* (PDB ID: 1FP2) with Isoflavone substrate and SAM shown inside the substrate binding pocket as ball-and-stick model. B) Quaternary structure of EnOMT1 predicted using Swiss-Model showing a homodimeric architecture. C) Surface model of EnOMT1 showing N-terminal helix insertion of a monomer (magenta) into the active site backwall region of neighbouring monomer (tan). All figures made in Chimera software by K. Chandran

Reliability of the predicted 3D-models was further checked utilizing Ramachandran plot. By analysing the backbone torsional/dihedral angles of each residue in the predicted protein structure, the plot gives information about the energetically favoured angle combinations and thus helps in the analysis of the stereochemical quality of a predicted model (Ramachandran et al., 1963). It is an indirect verification tool in which the quality of a 3D-model is derived based on the proportion of residues in favoured to that of the electrostatically forbidden or unfavourable region of the plot. All four models of EnOMT1 modelled with different parameters and servers were submitted to RAMPAGE Ramachandran plot server (Lovell et al., 2003) for stereochemical analysis.

The I-TASSER Model 1 Ramachandran plot (**Figure 3.18**) showed 91.6% residues in most favoured regions and 5.8% residues in additional allowed regions. Nine (2.6%) residues were in the outlier region, namelyGly43, Pro57, Gly103, Val115, Met190, Phe216, Lys251, Agr315, Arg331. A total of 97.4% of residues in allowed regions indicates that I-TASSER model 1 is a three-dimensional structure of reasonable quality.

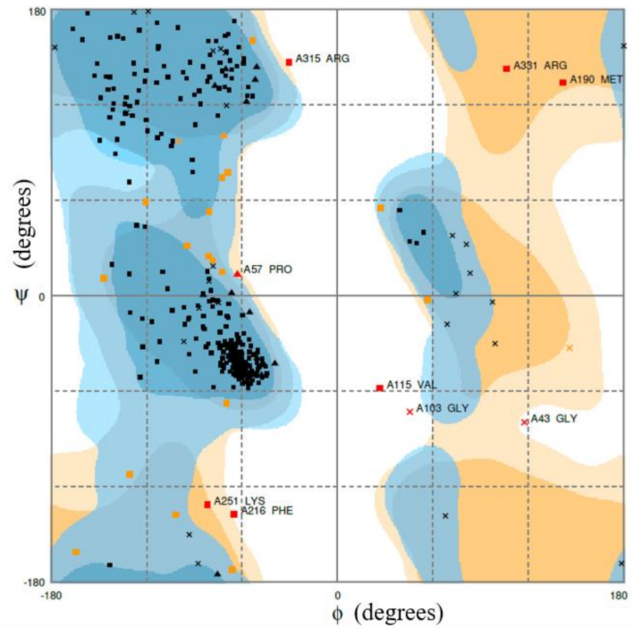


Figure 3.18: Ramachandran plot of I-TASSER model 1 showing torsional angles ψ and ϕ of different predicted amino acid in the protein backbone.

General/Pre-proline/Proline favoured | Glycine favoured

General/Pre-proline/Proline allowed | Glycine allowed

General refers to all eighteen amino acids except glycine and proline. Pre-proline refers to the residue that precedes proline in a peptide sequence. All these are represented by a square. Triangle represents proline and crosses represents glycine. The residues with unfavourable conformations are labelled and highlighted with red marks.

Ramachandran plot of tertiary structure of EnOMT1 modelled with *M. sativa* isoflavone OMT

(PDB ID: 1FP2) as the template (model 2) showed a greater number of residues in the outlier region compared to model 1. Therefore, model 2 was discarded.

As for the 3D structure predicted using the Phyre2 program, the number of residues in the favoured region was 94.5%, the number of residues in the allowed region was 4.6% and only 0.9% residues (His 210, Gln 288, Pro 289) were in outlier region (**Figure 3.19**).

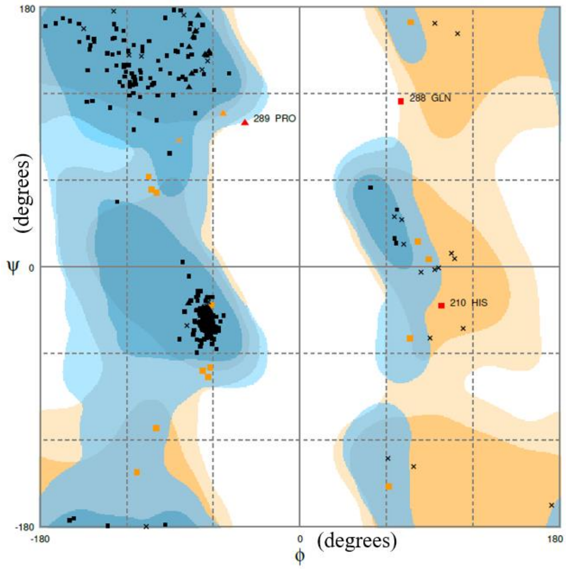


Figure 3.19: Ramachandran plot of Phyre2 model 1 showing torsional angles ψ and ϕ of different predicted amino acid in the protein backbone.

General/Pre-proline/Proline favoured — Glycine favoured

General/Pre-proline/Proline allowed — Glycine allowed

General refers to all eighteen amino acids except glycine and proline. Pre-proline refers to the residue that precedes proline in a peptide sequence. All these are represented by a square. Triangle represents proline and cross represents glycine. The residues with unfavourable conformations are labelled and highlighted with red marks.

Similarly, the number of residues in the favoured region in the model built by SWISS-MODEL was 97.1%, the number of residues in the allowed region was 2.9% and no residues were present in the outlier region. The Ramachandran plot generated by the inbuilt MolProbity tool in SWISS-MODEL is shown in **Figure 3.20** next to the one obtained from RAMPAGE server. The only difference between the two is that the plot obtained from RAMPAGE has more details about the glycine and proline residues whereas the MolProbity output only focuses on general (all except Gly and Pro) residues. The Ramachandran plot summary of all four three dimensional structures of EnOMT1 modelled with different servers is given in **Table 3.5**. All models showed good proportions of residues in favoured and allowed regions; however, in comparison, the quality of the model ranges from SWISS-MODEL structure> Phyre2 structure> I-TASSER model1.

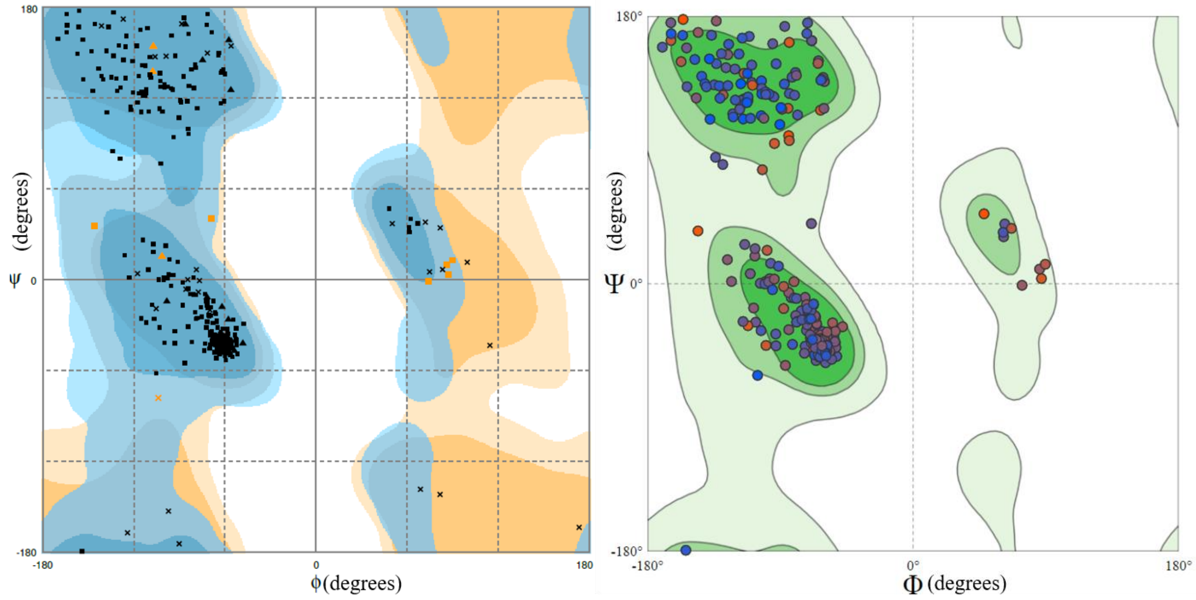


Figure 3.20: Swiss model Ramachandran plot showing torsional angles ψ and ϕ of different predicted amino acid in the protein backbone. A) Ramachandran plot generated using RAMPAGE server B) Ramachandran plot generated MolProbity server.

General/Pre-proline/Proline favoured — Glycine favoured

General/Pre-proline/Proline allowed — Glycine allowed

General refers to all eighteen amino acids except glycine and proline. Pre-proline refers to the residue that precedes proline in a peptide sequence. All these are represented by a square. Triangle represents Proline and cross represents Glycine. The residues with unfavourable conformations are labelled and highlighted with red marks.

Table 3.5: Ramachandran plot summary for different tertiary models predicted using various software applications showing the percentage of residues in favoured and outlier regions.

	Model name	Ramachandran plot values		
		Residues in the favoured region (%)	Residues in the allowed region (%)	Residues in the outlier region (%)
1	I-TASSER model 1	91.6	5.8	2.6
2	I-TASSER model 2	88.2	8.6	3.2
3	Phyre2 structure model	94.5	4.6	0.9
4	SWISS-MODEL structure	97.1	2.9	0

3.2.7 Ligand binding sites prediction

Based on the predicted tertiary structure of EnOMT1, biological annotations of the target protein were analysed using COFACTOR, which deduces protein functions (ligand-binding sites and gene ontology) using structure comparison and protein-protein networks (Zhang et al., 2017) and COACH, a meta-server that combines multiple function annotation results (on ligand-binding sites) from the COFACTOR, TM-SITE and S-SITE programs. Possible ligand-binding residues predicted by comparison with template proteins with similar binding sites are listed in (**Table 3.6**). SAM binding residues — namely, Trp148, Phe161, Met165, Asn169, Gln194, Phe216, Asp217, Arg218, Asp237, Met238, Phe239, Lyts251, Trp252, Val253, and Trp257 — were predicted with a higher degree of confidence by comparing with the similar templates (PDB ID:1KYW; caffeic acid OMT; *Medicago sativa*; **Table 3.6**-Row1; Figure 3.21). By comparison of the predicted EnOMT1 tertiary model to existing structurally characterised plant methyltransferases, only one residue (His255) was predicted as the active site residue with a confidence score of 0.72.

Table 3.6: Putative ligand binding sites of EnOMT1 predicted by COFACTOR.

Rank	C score*	PDB Hit[†]	TM-score[‡]	RMSD[§]	IDEN[¶]	Cov[#]	BS-score[††]	Ligand Name[‡‡]	Predicted binding site residues
1	0.61	1kyw	0.90	2.39	0.354	0.98	1.79	SAH	148,161,165,169,194,216,217,218,237,238,239,251,252,253,257
2	0.35	3p9k	0.92	2.12	0.338	0.97	1.45	C1Y	114,115,161,165,252,255,256,306,310
3	0.33	1fp2	0.89	2.57	0.395	0.97	1.47	HMO	115,120,121,124,139,161,252,255,256,305
4	0.32	1kyz	0.89	2.54	0.363	0.97	1.30	FER	115,120,147,161,309
5	0.28	3gxo	0.80	3.28	0.236	0.94	1.26	MQA	111,114,161,164,165,168,252,255,256,284,302,305,306,309
6	0.27	1tw3	0.78	3.39	0.218	0.93	1.02	ERT	112,147,161,165,168,173,252,256,301,302,305,309
7	0.19	1zgj	0.73	4.04	0.270	0.92	1.33	P1S	114,115,119,120,147,160,161,165,302,306,309
8	0.19	2qyo	0.77	3.96	0.310	0.96	1.27	QSO	115,120,121,124,306,309
9	0.17	3i58	0.74	3.79	0.225	0.90	1.10	7NA	112,161,165,252,255,302,305,306
10	0.10	1xdsA	0.76	3.49	0.210	0.91	1.03	DRA	112,147,161,165,200,252,256,298,302,305,306,310,337,341

*C- score is the confidence score of the predicted binding site [0-1]; [†]PDB Hit is the protein database identifier of the template proteins with a similar binding site; [‡]TM-score measures the similarity of two protein structures, and has values between 0 and 1, where 1 indicates a perfect match between two structures; [§]RMSD is the root mean square deviation between residues that are structurally aligned by COFACTOR; [¶]IDEN is the percentage sequence identity in the structurally aligned region; [#]COV is coverage of the global alignment and is equal to the number of structurally aligned residues divided by the length of the query protein; [‡‡]Ligand name -possible ligand name in three-letter code from PDB; [††]BS-score is the measure of local structural or sequence similarity of predicted binding site in EnOMT1 with that of the templates binding site. BS score>1 is significant; IDEN- percentage sequence identity in the structurally aligned region; COV- coverage of the global alignment. Top 5 ligands in the table are as follows: SAH- *S*-adenosyl-L-homocysteine, C1Y- 5'-deoxy-5'-Propyl-DADMe-Immucillin-A, HMO- 4'-hydroxy-7-methoxyisoflavone, FER- ferulic acid, MQA- Mitomycin A.

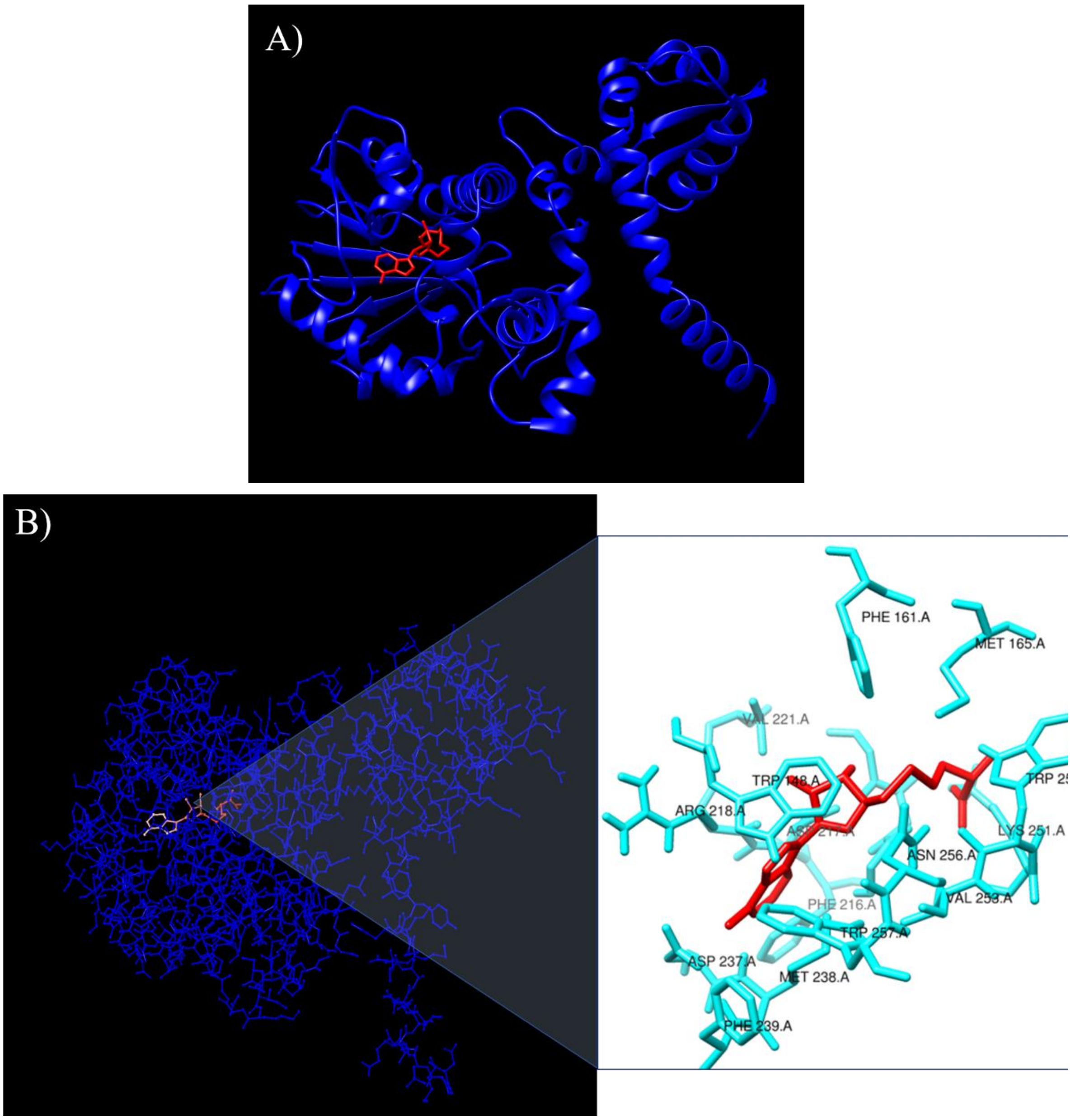


Figure 3.21: Binding of SAM to active site pocket of EnOMT1. SAM is shown as stick model and highlighted in red. A) EnOMT1 monomer bound to SAM. B) SAM bound monomer with the putative SAM binding residues highlighted (cyan) in the zoomed-in section

COFACTOR also predicted several other ligands as the possible substrate of EnOMT1 along with the residues that might involve in their binding (**Table 3.6**). These ligands are structurally distinct from pinocembrin except for 4'-hydroxy-7-methoxyisoflavone (HMO) and ferulic acid. Also, the C-score is considerably lower compared to predicted SAM binding residues. Therefore, residues involved in substrate/pinocembrin binding were further studied employing ligand docking analysis. Given our discovery that pinocembrin is likely the main biological substrate of EnOMT1, this substrate was docked into the 3D-model of the EnOMT1 protein using a web-based platform SwissDock (Grosdidier et al., 2011) and results were visualised and analysed in UCSF Chimera software (Pettersen et al., 2004). In general, for docking analysis, the viable conformation of the substrate is decided by a comparison of the binding free energy (ΔG). The lower the energy, the more favourable the binding mode.

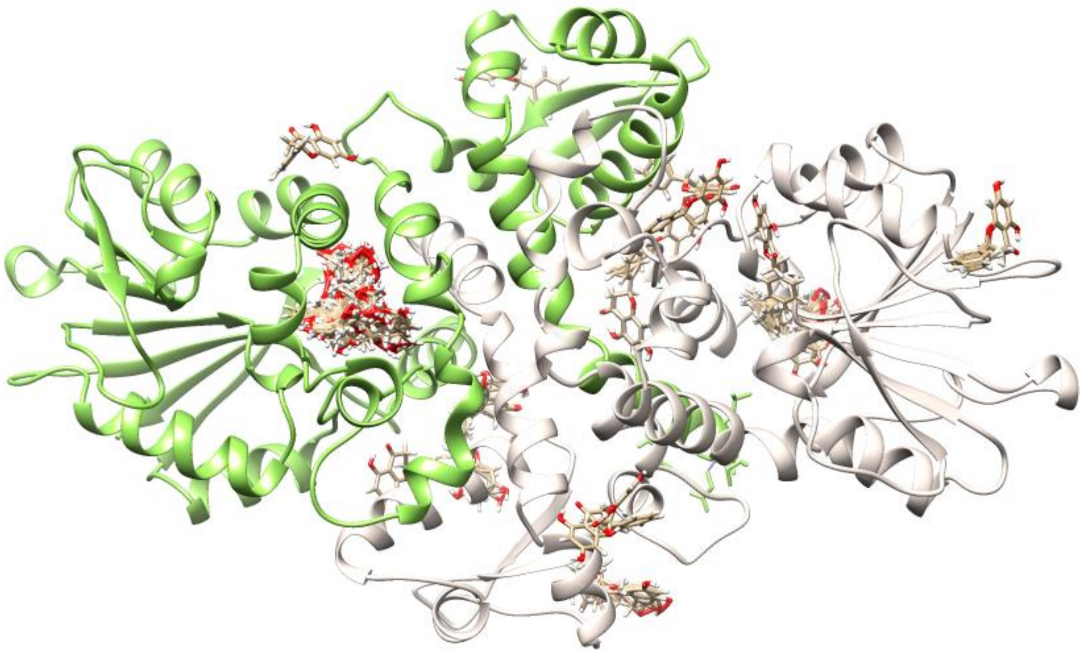


Figure 3.22: Predicted poses of pinocembrin in all possible binding pockets of EnOMT1 homodimer structure. Monomer-A is depicted in green ribbons and monomer-B is depicted in off-white.

The ligand pinocembrin was successfully docked on monomer-A of the EnOMT1 dimer model. For pinocembrin, 210 distinct binding poses were produced (**Figure 3.22**). The docking results are grouped as an optimised population of similar binding poses at a given pocket, called clusters (Grosdidier et al., 2007). The best binding conformation of pinocembrin in EnOMT1 was selected based on the receptor-ligand binding energy and the active site location in related methyltransferase crystal structures. Out of the thirty clusters predicted for pinocembrin docking, only fifteen were in the vicinity of the methyltransferase active site. These fifteen models of EnOMT1 with pinocembrin docked into the active site were superimposed with the EnOMT1 model docked with S-adenosyl-homocysteine (SAH). Amino acid residues interacting with pinocembrin were the same for all fifteen clusters; however, pinocembrin conformation in cluster-11 showed the most favourable position for methylation with the 7-*OH* in the A-ring positioned near the sulphur entity on the S-adenosyl-homocysteine (SAH) and the distance between these two is 4.5Å (**Figure 3.23**). Predicted interacting residues within a distance 2Å from pinocembrin — namely, Met15, Tyr19, Ala111, Met120, Trp124, Trp148, Leu166, Asn256, His255 — are shown in **Figure 3.23**. Met15, Tyr19 belongs to the N-terminal helix of the neighbouring monomer (monomer-B).

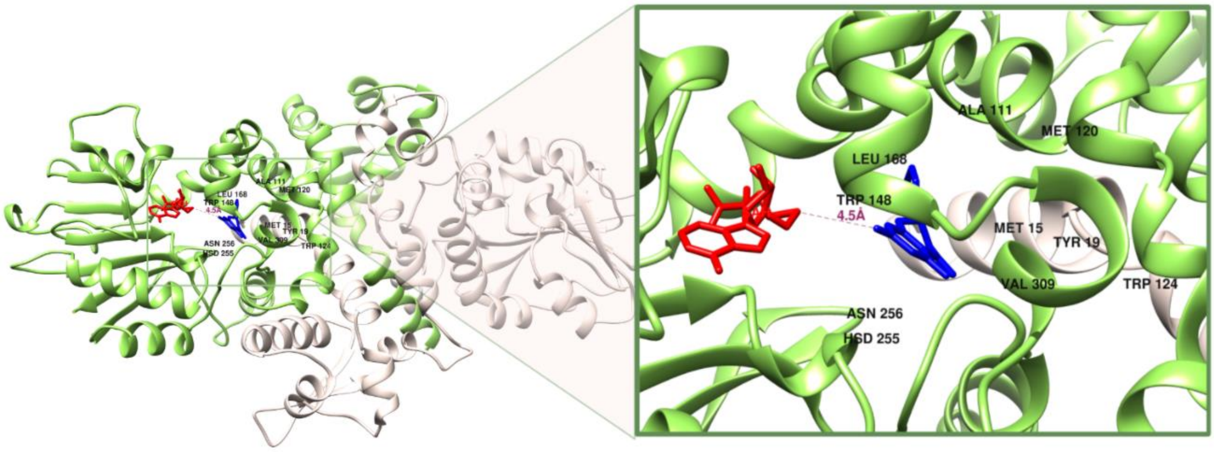


Figure 3.23: Pinocembrin (blue) and SAH (red) docked into the predicted active site of EnOMT1. All residues within 2Å of pinocembrin are labelled. Figure made with UCSF Chimera software.

Since there are no crystal structure studies available for any flavanone OMT, blindly depending on prediction data for substrate binding (pinocembrin) is unwise, but nevertheless provides useful background data. Crystal structure analysis is needed to confirm the docking as well as homology modelled predictions.

3.2.8 Site-directed mutagenesis

To confirm the importance of the putative residues involved in substrate binding, they were subjected to site-directed mutagenesis and the expressed mutant proteins assayed for enzyme activity. Functional effects of single amino acid substitutions in EnOMT1 were predicted with the bioinformatic tool SNAP2 (Hecht et al., 2015). Amino acid changes in the C-terminal region of EnOMT1 region which is believed to be involved in SAM/substrate binding ability were predicted to have a strong negative impact on its activity (**Figure 3.24**).

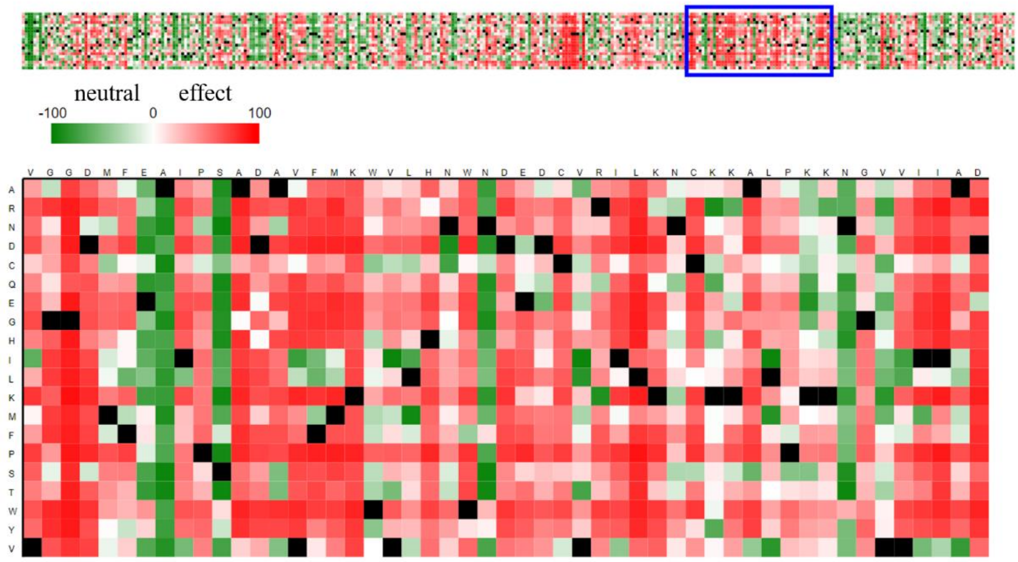


Figure 3.24: Heatmap showing predicted functional effect of amino acid substitutions in EnOMT1. The C-terminal end spanning amino acid positions 234-284 is zoomed in to show the predictions in detail. The horizontal axis of the mutational matrix depicts the wild type amino acids of EnOMT1 and the vertical axis depicts twenty amino acid to correspond to all possible single mutations. Effect of each amino acid substitution independently for each position of EnOMT1 protein are coloured by their score where red indicates a high score (score > 50; strong signal for effect), white and green (score < 50) indicate neutral effect and black marks the corresponding wildtype residues.

Predicted functional effects of the amino acid changes in ExOMT1 to that of EnOMT1 were analysed with the SNAP2 results. Out of the twenty amino acid changes, only two 252 W→G and 284 D→E were predicted to have a functional effect (**Table 3.7**). All others were predicted to be neutral with 53-97% confidence. For validation of these predicted effects, site-directed mutagenesis experiments were performed. Amino acids in EnOMT1 were mutated towards the ExOMT1 sequence. Mutations were introduced in the EnOMT1 gene in three different regions. Mutation regions, containing three amino acid changes each, were named as region I, II and III (**Figure 3.25**A). Initially, four different mutants (EnOMT1_M1 to EnOMT1_M4; **Figure 3.25**B) were designed, cloned and expressed in *E. coli*. *In-vitro* assays of the variants in the presence of various concentrations of pinocembrin were carried out to find the relative activity of mutants compared to EnOMT1. EnOMT1_M1 with mutations in the region I (harbouring changes 252 W→G, 253 V→I, and 256 N→D) completely lost catalytic activity (**Figure 3.26**). EnOMT1_M2 with mutations in region II (harbouring changes 296 E→H, 298A→V, and 300 V→A) had no major effect and in fact, appeared to be a slightly improved version of the wild type EnOMT1.

Table 3.7: Predicted functional effect of sequence variants on ExOMT1 to that of EnOMT1

Wild amino acid	Position	Variant amino acid	Predicted effect	Score*	Expected accuracy†
M	18	I	neutral	-86	93%
Y	19	F	neutral	-82	93%
V	112	L	neutral	-55	78%
A	119	T	neutral	-83	93%
M	121	V	neutral	-55	78%
L	147	F	neutral	-74	87%
R	218	Q	neutral	-58	78%
W	252	G	**effect**	20	63%
V	253	I	neutral	-87	93%
N	256	D	neutral	-95	97%
D	284	E	**effect**	8	53%
E	296	H	neutral	-8	53%
A	298	V	neutral	-88	93%
V	300	A	neutral	-46	72%
M	305	L	neutral	-66	82%
V	309	L	neutral	-90	93%
T	310	S	neutral	-86	93%
R	311	G	neutral	-15	57%
Y	332	C	neutral	-39	66%
C	339	S	neutral	-39	66%

*SNAP prediction score that correlates with severity of effect. Values range from 100 to +100 and a score > 0 depicts signal for effect. Severe effects have a score>50; †Accuracy of predicted functional effect.

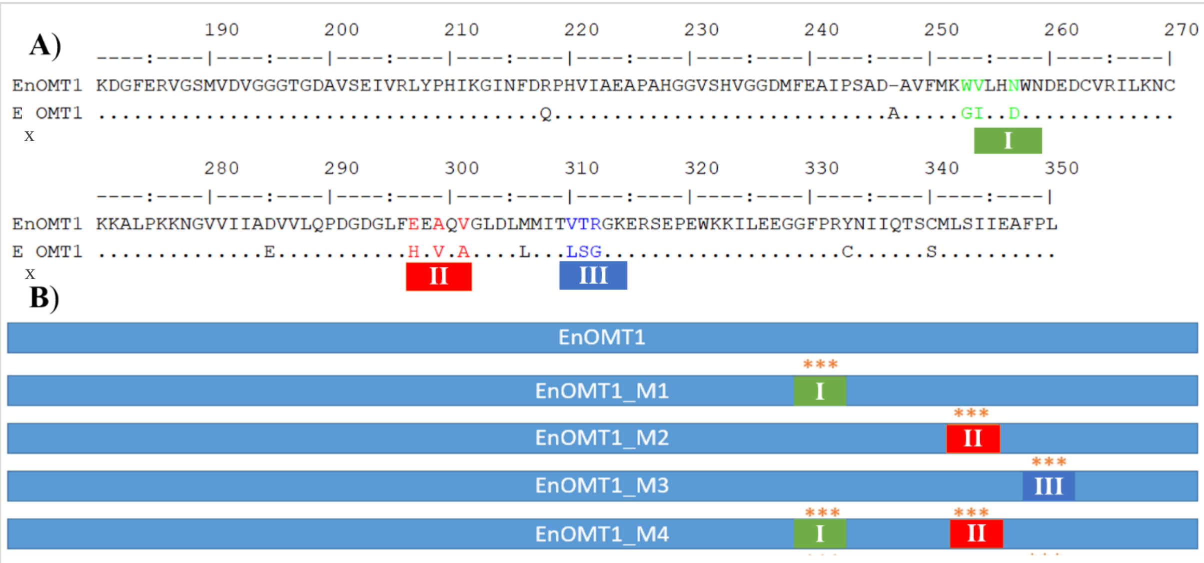


Figure 3.25: Site-directed mutation of EnOMT1. A) Coding sequence EnOMT1 aligned with ExOMT1. All agreements to EnOMT1 sequences are denoted as dotes and only amino acids diagreements in ExOMT1 is shown. Three regions (I- III) each harbouring three amino acid mutations are also shown. Highlighted amino acids of EnOMT1 in respective regions are mutated to corresponding amino acids of ExOMT1. B) Schematic showing mutants M1-M4 with multiple mutations in comparison to EnOMT1.

EnOMT1_M3, with mutations in region III (harbouring changes 309 V→L, 310 T →S, and 311 V→A), and EnOMT1_M4, with the region I as well as region II mutation, showed a decrease in activity compared to the other mutants, but still produced some pinostrobin.

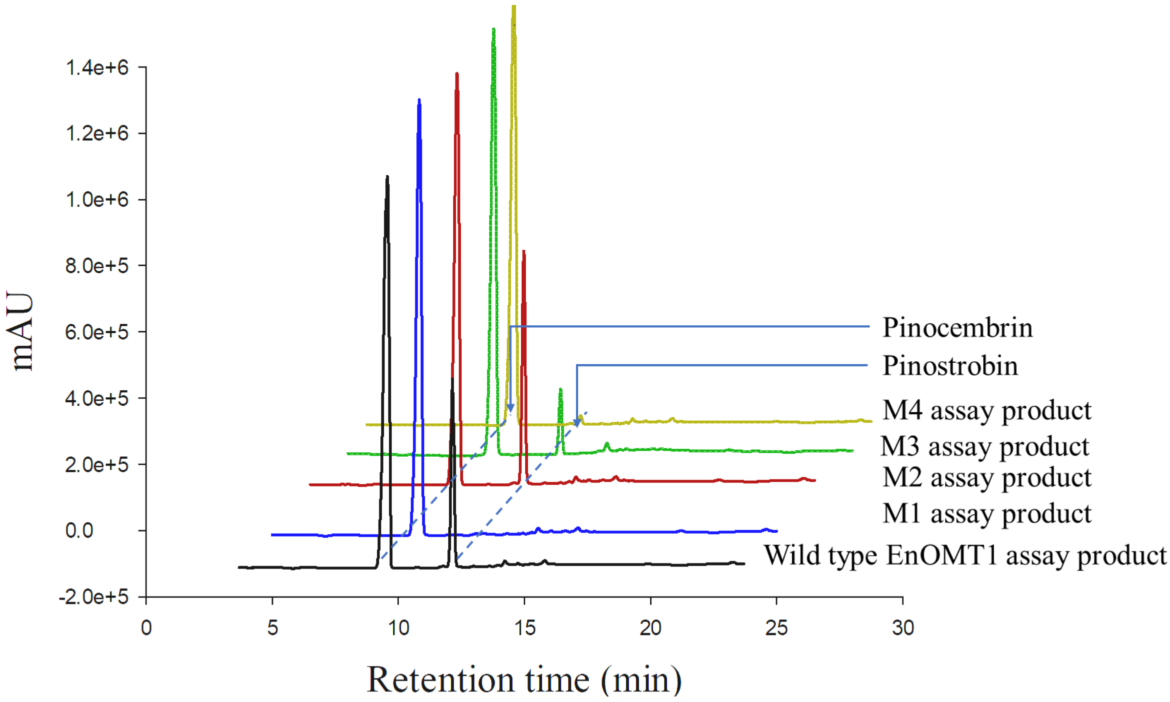


Figure 3.26: Chromatogram of EnOMT1 mutant assay products. Each assay was incubated with 50µM pinocembrin for 10 min.

To examine further the contribution of individual amino acid changes in that region, single mutants EnOMT1_W252G, EnOMT1_V253I, and EnOMT1_N256D (hereinafter referred to as single mutant A, B and C respectively) were designed. The single mutant A showed almost the same strong effect as M1 whereas mutants B and C showed a smaller effect as shown in **Figure 3.27**.

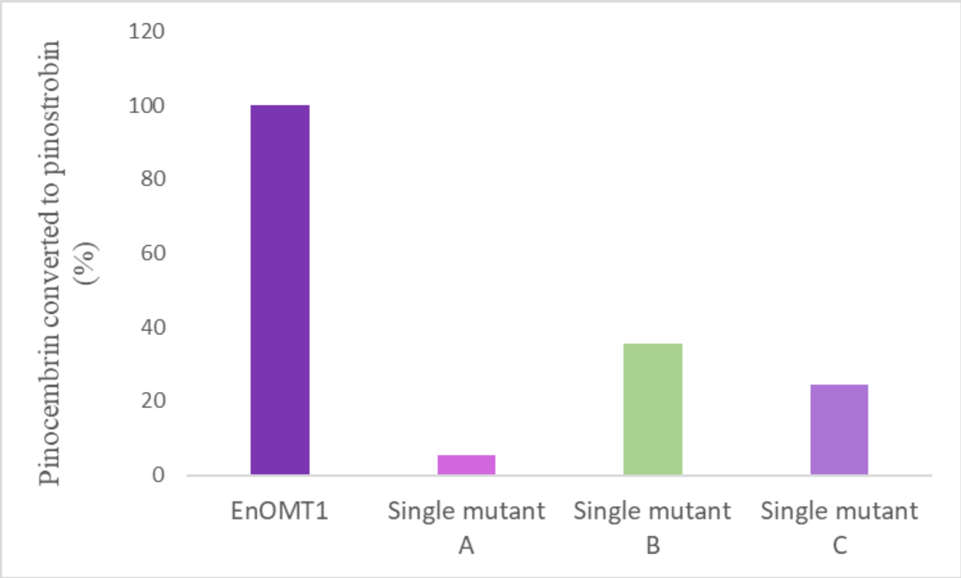


Figure 3.27: Activity of single mutants compared to that of EnOMT1 expressed in terms of the percentage of pinocembrin converted into product.

3.2.9 Search for 5-*O*-methyltransferase in *O-methylated species 1*

The flavanone profile of *O-methylated species 1* shows the presence of small amounts of a dimethylated form of pinocembrin (**Figure 3.8**). The methyl groups on dimethylpinocembrin are on 7-*OH* and 5-*OH* of the A ring. From the results so far, it is evident that EnOMT1 accounts for the 7-*OH* methylation. Given my results suggest EnOMT1 is 7-*OH* specific, there should be another methyltransferase that catalyses methylation of the 5-*OH*. Therefore, we continued to search for the 5-OMT in *E.nitida*. Since the 5-*O*-methylation is likely to happen on pinocembrin, we hypothesised that the 5-OMT sequence might be similar to that of EnOMT1.

Blast analysis of the EnOMT1 sequence in the *O-methylated species 1* transcriptome database resulted in a sequence with 67% sequence identity to EnOMT1 (named EnOMT7; **Figure 2.15**; **Figure 3.28**). The gene was successfully cloned into the pHUE vector system and subsequently expressed in E. coli. Unfortunately, *in vitro* enzyme assay did not show any formation of 5-*O*-methylated alpinetin.

```
                10        20        30        40        50        60        70        80        90
       ----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|-
EnOMT7 MESLDETEATLRGQEPIWKLIYAYADSMALKCAVELRLPDIIHSYGGGPVTLAQIASRIPSPSPGTAYLARIMRLLVRKNIFCANRCQADG
EnOMT1 .DP....---....AQAM.HM..F.............I......E.A.................EM....S......C..V.S.V.-----

                  100       110       120       130       140       150       160       170       180
       ---:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|--
EnOMT7 GHEGHEILYSLTPSSRWLLQGPGSELSLAPLVLMENHPWLMSPWHCLGDCVRTG-SVAFEMAHGRKIWDFASENPEFNRLFNDGMACLSKI
EnOMT1 --N.G.T..G...T.......-.GH.....A..LVS..AM.....N.S..IKD.G.T..Q.....GL..L.T.......T.RE.....N.L

                 190       200       210       220       230       240       250       260       270
       --:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|---
EnOMT7 LVRGIVEVYKDGFKTIGSLVDVGGGVGRTLAEIIKSFPHIQGINFDLPHVVATAPMHDGVTHVGGDMFETIPQADAIFMKGIMHDWGDEDC
EnOMT1 MMKA..DA.....ERV..M......T.DAVS..VRLY...K.....R...I.E..A.G..S........A..S...V...WVL.N.N....

                280       290       300       310       320       330       340       350       360
       -:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|-
EnOMT7 VKILKNCRKAIPEKNGKVIIADMVLKPEGDGMFDETGLVLDLVMLVHSSGGKERDELEWKKILEEGGFPRYNIIKTLSFLSIIEAYPL
EnOMT1 .R.....K..L.K...V.....V..Q.D...L.E.AQVG...-.MITVTR....S.P.................Q.SCM......F..
```

Figure 3.28: Sequence comparison of EnOMT1 with EnOMT7. All agreements to EnOMT1 are denoted as a dot.

Another approach is to undertake transcriptome sequencing of an additional species abundant in 5-*O*-methylated derivatives of pinocembrin. Extensive screening of different *Eucalyptus* species accumulating flavanones in foliar glands has been done to search for such a species (described in detail in Chapter 4). This approach was unsuccessful as no candidate species that produce high levels of alpinetin have been detected (Goodger et al., 2019; Goodger et al., 2016; Marsh et al., 2019).

3.3 Discussion

Eucalyptus subgenus *Eucalyptus* is known to contain at least three major *O*-methylated flavanone compounds – pinostrobin, dimethylpinocembrin and alpinetin in lower abundance. During this investigation to uncover the OMTs involved in the methylation of the flavanone aglycone pinocembrin, a flavanone 7-OMT, EnOMT1, was successfully isolated and characterized.

3.3.1 Biochemical characterization shows that EnOMT1 is an efficient methyltransferase enzyme.

The kinetic properties of EnOMT1 were comparable to the published values observed with isoflavone 4′OMT from garden pea (*Pisum sativum*; K_m = 3μM for 2,7,4′-trihydroxyisoflavanone; (Ibrahim et al., 1987), naringenin 7-OMT from rice (*Oryza sativa*;

K_m=1.9 ± 0.1 µM for naringenin) and a flavonol OMT from *Chrysosplenium americanum* (K_m=6.4 µM for 3′,5-dihydroxy-3,4′,7-trimethoxyflavone) (Gauthier et al., 1996). In general, a lower K_m demonstrates a considerably higher affinity of the enzyme towards the substrate. Along with a very low K_m value of 2.05 ± 0.16 µM, EnOMT1 shows higher steady-state kinetic parameters, k_{cat} of about 0.1 s^{-1} and k_{cat}/K_m of about 0.05 $\mu M^{-1}s^{-1}$, than some characterised plant *O*-methyltransferases (k_{cat}/K_m Norcaclaurine-OMT = 7.4 $mM^{-1}s^{-1}$; Ounaroon et.al., 2003)(Ounaroon et al., 2003; Schmidt et al., 2012). The kinetic parameters of EnOMT1 are in very close agreement with an improved variant of *T. flavum* (S)-scoulerine 9-OMT (K_m for alkaloid (*S*)-scoulerine of 2.91 ±1.20 µM and a k_{cat} of 0.102 ± 0.009 s^{-1}) recently produced by structure-guided engineering of the wild type enzyme (Valentic et al., 2020). On the other hand, two 7-*OH* specific flavonoid OMTs from *O. basilicum* has been shown to have a significantly higher catalytic efficiency [k_{cat}/K_m of 0.73 $\mu M^{-1}s^{-1}$ and 2.81 $\mu M^{-1}s^{-1}$ with apigenin respectively]. In the case of the co-substrate SAM, most of the previously characterised OMT enzymes show a higher K_m value for SAM [K_m = 20 µM (Gauthier et al., 1996); 21.7 µM (Levac et al., 2008); 67 µM (Chen et al., 2003); 99.8 µM (Liu et al., 2006)]. Nevertheless, a low K_m of 4.99± 0.29 µM of EnOMT1 with SAM shows that the catalytic efficiency of EnOMT1 for SAM is multiple folds higher than that of other OMTs. The flavanone profiling of *Eucalyptus nitida* showed a pinostrobin concentration of 10.48 mg g^{-1}_{DW}, which is comparatively higher than that of other flavanones in the species. To achieve such a high level of conversion, the enzyme should have a high affinity for both its substrates, pinocembrin and SAM. Thus, a comparable low K_m value of EnOMT1 for both substrates suggests the recombinant enzyme is a faithful reproduction of its biological counterpart.

3.3.2 EnOMT1 is a regiospecific methyltransferase with strict positional specificity to the 7-hydroxyl of flavonoids

During the *in vitro* substrate specificity studies presented here, only flavonoid substrates with a free hydroxyl at the 7th position were methylated by EnOMT1, although to different degrees, indicating EnOMT1 is 7-*OH* specific methyltransferase of flavonoids. Such stringent specificity has been described as a general attribute of plant flavonoid 7-OMTs (Berim et al., 2012; Christensen et al., 1998; Shimizu et al., 2012). Surprisingly, no products were obtained from assays of EnOMT1 with the 7-*OH* substrates quercetin and luteolin. Both flavonoids possess two B-ring hydroxyls, suggesting these moieties preclude EnOMT1 action.

The majority of the flavonoid *O*-methyltransferases characterised so far accept quercetin as a substrate and show a specific pattern in the methylation of hydroxyls on the *meta* (6-, 8-, 3′-), *para* (7-, 4′-) and 3-positions of the flavonoid ring system i.e., some catalyse *meta O*-methylation while others catalyse *para O*-methylation (Brunet and Ibrahim, 1980). The C-ring hydroxyl of quercetin is known to influence the substrate binding and the substitution of its 3-*OH* affects the enzyme activity (De Luca and Ibrahim, 1985; Itoh et al., 2016). Methylation of 3′-*OH* is a prerequisite for 7-*OH* as well as B-ring methylation of quercetin in *Chrysosplenium americanum* flavonoid glucoside biosynthesis (Ibrahim et al., 1987). Therefore, it is possible that with EnOMT1, the 3-*OH* or 3′-*OH* of quercetin may be interfering with substrate binding. Conversely, previous experiments have suggested the presence of 3-*OH* stabilizing the flavonoid structure and enhancing the reactivity of both ring hydroxyls towards methylation (Brunet and Ibrahim, 1980).

The reduction/absence of activity of EnOMT1 with B-ring substituted flavonoids such as luteolin, apigenin, and quercetin demonstrates the drastic effect B-ring hydroxyl configuration has on the activity of the enzyme. Consistent with the results from studies of other flavonoid OMTs, this study further strengthens the relationship between a substrate's hydroxylation pattern and its methyl-accepting ability. In *Chrysosplenium americanum*, the nucleophilicity of A-ring hydroxyls and, consequently, their reactivity towards *O*-methylation was affected by B ring substitution (Ibrahim et al., 1987). Also, a 4′-*O*-methyltransferase isolated from *C. roseus* only methylated flavonoids with the same B-ring hydroxyl configuration as that of its substrate homoeriodictyol (3′-*OMe* 4′-*OH*) and a specific A-ring substitution pattern (Schröder et al., 2004). In the case of EnOMT1, substrates with only 4′-*OH* on the B-ring showed a reduced level of activity but no activity was detected for compounds with vicinal hydroxyl groups on B-ring – i.e., both 3′ and 4′ hydroxyl. In a study of *in vitro O*-methylation, cell-free extracts of calamondin orange tissues catalysed methylation of a number hydroxyls on quercetin and its derivatives (Brunet and Ibrahim, 1980). The authors consequently suggested that various interactions of A- and B- ring meta (6-, 8-, 3′-) and para (7-, 4′-) hydroxylation patterns influence methylation. The rigid para-quinoid structure, which results from hydroxylation at the 4′-position, results in resonance of the flavonoid skeleton and thus a limit to A ring methylation when there is no vicinal hydroxyl in position 6 or 8 (Brunet and Ibrahim, 1980). Therefore, the lack of hydroxylation at 6 and 8 position of the flavonoid structures tested in this study might account for the reduced activity of EnOMT1 towards substrates with 4′-*OH* methylation. Whether EnOMT1 will methylate the 7-*OH* of a substrate with hydroxyl at 6, 8 and 4′-positions needs to be investigated. Similarly, hydroxylation at both 3′-and 4′-positions, as in luteolin and quercetin, results in an ortho-quinoid, structure, which is suggested to promote the A-ring methylation of 6-*OH* and 8-*OH* positions (Brunet and Ibrahim, 1980). On the contrary, substrate hydroxylation at both 3′-and 4′-positions completely removed activity of EnOMT1 towards the A-ring 7-*OH* in the compounds examined here. Even though EnOMT1 did not accept luteolin, hesperetin (a monomethylated derivative of luteolin at 4′-*OH*) showed approximately 15% activity compared to pinocembrin (**Table 3.2**). Thus, it appears the methylation of the 4′-*OH* enabled EnOMT1 to further methylate hesperetin at the 7-*OH* position. The degree of methylation of the functional hydroxyl groups is known to regulate the biosynthesis of various secondary metabolites (Phillipson et al., 2012; Rueffer and Zenk, 1985). Therefore, the methylation of the 4′-*OH* appears to have enabled the enzyme to accommodate hesperetin into its active site.

Recombinant EnOMT1 exhibited equivalent activity towards the flavone backbone structure chrysin *in vitro*. Detailed kinetic studies comparing chrysin and alpinetin (both compounds are known to be present at low levels in *Eucalyptus*) showed that chrysin has a slightly lower K_m value to that of pinocembrin. This is not surprising as chrysin and pinocembrin are structurally and biochemically very similar. The result show that the C-ring double bond between carbon 2 and carbon 3 (the only difference between chrysin and pinocembrin 2D-structure) does not affect the 7-*O*-methylation of flavonoids by EnOMT1. Then again, caution should be shown when using *in vitro* data for the prediction of the physiological substrates and products of OMTs (Liu and Dixon, 2001). For example, an isoflavone OMT methylates B-ring 4′-*OH* of daidzein *in vivo*, but *in vitro* studies show that it acts on 7-*OH* of daidzein which is on the A ring (Zubieta et al., 2001). Dimethyl chrysin and glycosylated chrysin have been reported to be present in various *Eucalyptus* species (Goodger et al., 2016) and so are flavone aglycones (Wollenweber and Kohorst, 1981). Therefore, in species with chrysin, it may be possible that the same enzyme is responsible for the synthesis of methylated forms. EnOMT1 may catalyse the conversion of both substrates, if present, *in vivo*, and convert them into respective methylated forms. Nevertheless, as flavone derivatives metabolites are not detected in the *O-methylated species 1* gland extract (Goodger et al., 2016), this activity will likely not play a role in the intact *O-methylated species 1* plant. Therefore, for *O-methylated species 1*, the activity with chrysin and other flavones might not have any physiological relevance.

An important finding is the activity of EnOMT1 with alpinetin, a 5-*O*-methylated flavanone present in *Eucalyptus* species. The relative activity of EnOMT1 with alpinetin was around 30%

compared to EnOMT1 activity with pinocembrin as substrate. Most of the species with a high amount of 7-*O*-methylated pinostrobin failed to show any detectable amount of alpinetin. Instead, 5, 7-di-methyl pinocembrin was commonly detected. It is possible that the highly active EnOMT1 convert most of the alpinetin that is present in the cell to di-methylated pinocembrin, which then gets transported and sequestered into subcellular secretory glands (Goodger et al., 2016).

Multiple methylations of flavonoids are usually catalysed by distinct enzymes (Ibrahim et al., 1987); however, sequential methylation of flavonoids by the same enzyme has also been observed previously (Cacace et al., 2003). Particularly, CdFOMT5 from *Citrus depressa* methylates four different hydroxyl positions on quercetin A and C-ring (Itoh et al., 2016) in no particular sequence. In contrast, TaOMT2 from wheat exhibits strict sequential methylation of B-ring 3′-,4′- and 5′-*OH* of tricetin proceeding in an order of 3′-mono-→3′,5′-di-→3′,4′,5′-trimethyl ether derivatives (Zhou et al., 2006). Furthermore, an altered regiospecificity was shown by Poplar OMT7 due to two different conformations of substrate binding during molecular docking studies (Kim et al., 2008). On the other hand, I did not find sequential methylation of pinocembrin by EnOMT1. Longer incubation did not result in dimethyl pinocembrin (a major di-*O*-methylated flavanone in *Eucalyptus* species) formation. Therefore, this result is in accordance with the hypothesis that a separate flavanone OMT, specifically methylating an A-ring 5-hydroxyl on pinocembrin, is expected to be involved in the synthesis of alpinetin and dimethyl pinocembrin.

The quest for isolation of a methyltransferase acting on 5-*OH* of pinocembrin in *O-methylated species 1* was not successful. I was expecting the 5-OMT to show very high primary sequence identity to EnOMT1, as observed previously with different OMTs acting on similar compounds in the same or related species in the genera *Humulus lupulus* and *Thalictrum flavum* (Nagel et al., 2008; Robin et al., 2016; Valentic et al., 2020). The leading candidate from transcriptome analysis, with 68% sequence identity to OMT1, did not produce alpinetin. Additionally, none of the other OMT candidates demonstrated 5-*O*-methyltransferase activity. Many factors such as amino acid composition, protein folding and post-translational modifications can influence the expression of functional plant enzymes in *E. coli* (Itoh et al., 2016). The possibility of a post-translationally modified form of EnOMT1 acting to methylate at the 5-*OH* position also needs to be investigated. Transgenic expression of EnOMT1 in a model plant system may enable further investigation of translational modifications leading to the reversal of 7-OMT to a 5-OMT. For example, co-expression of two OMT genes involved in the biosynthesis of noscapine in opium poppy (Park et al., 2018) resulted in an unusual heterodimer which catalysed a distinct substrate compared to their homodimer counterparts. Although, there are no other reports of heterodimeric OMTs in plants, provided the leading candidate from transcriptome analysis had only 68% sequence identity with EnOMT1, it is worthwhile investigating the heterodimer possibility for a 5-OMT in *O-methylated species 1*

3.3.3 Amino acid changes identified as important catalytic residues

Searching for EnOMT1 orthologue in *non-methylated species* and *O-methylated species 2* resulted in an interpretation of residues that might have a role in catalysis. Despite having a comparably high quantity of pinocembrin in glands, *non-methylated species* accumulates no methylated flavanones. Gene expression analysis by quantitative PCR showed that the EnOMT1 orthologue in *non-methylated species,* ExOMT1, has a lower activity compared to EnOMT1 **(**Appendix E). The observed difference in activity is likely due to the amino acid difference of EnOMT1 sequence to ExOMT1. Out of the twenty amino acid changes, *in-vitro* validation of nine amino acid changes were performed by site-directed mutagenesis.

Methyltransferases are thought to proceed with the direct transfer of a methyl group to a substrate in an S_N2-like reaction facilitated by a basic catalytic residue (i.e. histidine, lysine) (Woodard et al., 1980). Comparative modelling results of EnOMT1 indicated His-255 as the key active site residue. In the case of plant OMTs, there is considerable evidence that a histidine around this position is the residue that deprotonates the substrate nitrogen or hydroxyl (Bennett et al., 2018; Zubieta et al., 2002). The importance of the His residue imidazole ring, for electron flow between adjacent residues, has been well demonstrated by the complete loss of activity after mutation (Zhou et al., 2010). While the His255 residue was conserved in ExOMT1, the substitution of three adjacent amino acid residues (W252, V253, N256) in EnOMT1 to mimic the sequence of ExOMT1 resulted in a protein that exhibited complete loss of activity. Reduction in catalytic efficiency of single mutants further supports the importance of these individual amino acids in the catalytic process. Therefore, in addition to the catalytic His residue, the adjacent residues — namely, Trp252, Val253, and Asn256 — play critical associative roles, possibly forming a channel for methyl passage (Cabry et al., 2019). Furthermore, structural studies have shown that in alkaloid OMTs, His-Asp acts as the catalytic dyad (Cabry et al., 2019). An Asp residue is present adjacent to His in a majority of the isolated flavonoid and phenylpropanoid OMTs. In flavonoid or phenylpropanoid OMTs, the involvement of this Asp residue in hydrogen bonding network in enzyme-substrate complex has been demonstrated through crystal structures analysis (Green et al., 2014). In EnOMT1, the change of asparagine to aspartic acid resulted in a reduction of activity. Hence, His-Asn might be the catalytic dyad of EnOMT1.

Substitutions at three amino acids in the 310 regions (Val 309, Thr 310, and Val311) were predicted to have a neutral effect, but they were in fact shown to be minor determinants of catalytic efficiency. The side chains of polar Thr301 might be involved in hydrogen bonding formation as seen in the case of isoflavone OMT from *Medicago sativa* (Zubieta et al., 2001). The hydrogen bonding via polar sidechains is necessary to stabilize the hydroxyl group of the methyl acceptor (Zubieta et al., 2001). Surprisingly, EnOMT1_M4, with region I as well as region II mutations, showed greater activity than mutations in region I alone. While the reason behind this is not clear, it is possible that the mutations in region II (Glu 296, Ala 298, and Val 300) have in a small way counteracted the drastic effect of region I mutation. This is supported by the results observed for mutations in region II alone (M2), which showed improved activity in comparison to wild-type EnOMT1 (**Figure 3.26**).

The results of crystal structure studies suggest that the general structure of the active site of plant OMTs is very similar (Zubieta et al., 2001; Zubieta et al., 2002). The active site pocket is lined with majorly hydrophobic residues whose side-chain positions and orientation control the shape and size of the pocket and thus help in differentiating between similar methyl acceptors (Zubieta et al., 2001; Zubieta et al., 2002). Thus, it is likely that these hydrophobic amino acids are involved in substrate binding. Nevertheless, based on the mutagenesis study we cannot definitively state whether the loss of activity is due to inefficient pinocembrin binding or SAM binding. The amino acid changes in EnOMT1 were directed toward the sequence of *non-methylated species* OMT1. The fact that ExOMT1, at a higher pinocembrin concentration and longer incubation, showed minor activity and produced pinostrobin shows that the enzyme ExOMT1, although inefficient, can bind pinocembrin. Hence, it is possible that ExOMT1 lost the ability to efficiently utilise SAM, so that even after having access to a very high concentration of pinocembrin *in vivo*, ExOMT1 fails to methylate it. Another possibility is that the binding strength (K_d) of pinocembrin was reduced significantly. Activity in that case would proportionally reduce unless pinocembrin was at very high levels.

Another possible explanation for the lack of methylated forms of pinocembrin in *non-methylated species* is substrate sequestration before the enzyme can access it. Two biosynthetic enzymes, chalcone synthase and chalcone isomerase, responsible for the production of flavanones, have been shown to be located not only in the cytoplasm but also in the nuclei of some cells (Saslowsky et al., 2005). This differential localization of enzyme systems may be a conserved mechanism for controlling the site-specific accumulation of flavonoid end products. In *Eucalyptus*, the flavanone pinocembrin and its methylated derivatives are localized to the extra-cellular lumen of sub-dermal secretory glands in the leaf (Heskes et al., 2012). The methyltransferase enzymes, on the other hand, are known to be a component of an aggregated multienzyme system on the surface of the endoplasmic reticulum within cells (Ibrahim et al., 1987). Tabersonine 16-*O*-methyltransferase encoded by 16OMT in *C. roseus* has been localised specifically in abaxial and adaxial epidermal cells of young leaves, and GFP imaging analysis showed that the enzyme is organised *in vivo* as a homodimer in the cytosol of these epidermal cells (Murata and Luca, 2005). Hence, from a physiological point of view, it is possible that the metabolic channelling of pinocembrin into the gland lumen would prevent its efficient uptake by the cytosolic or ER-anchored EnOMT1. The enzymes involved in plant specialized metabolism often have low turnover numbers (Dang and Facchini, 2012; Park et al., 2018; Salim et al., 2018). To overcome this lower catalytic efficiency, plants often express such enzymes at very high levels in the appropriate tissue and cell type to achieve the required level of catalytic activity (Gang et al., 2002). The transcriptomics data shows that the expression level of ExOMT1 is substantially lower than EnOMT1. The reduced expression of the ExOMT1 gene, along with the reduced activity of the ExOMT1 enzyme combined with pinocembrin sequestration to the gland lumen immediately after synthesis, might explain the absence of any detectable amounts of methylated derivatives of pinocembrin in *non-methylated species* gland extracts.

Unfortunately, due to time constraints, I did not test the effect on activity of the remaining amino acid differences between EnOMT1 and ExOMT1. Given the results of mutagenesis, the majority of the reduced activity of ExOMT1 seems to be explained by changes in the region I near the catalytic His255 region. So, the remaining amino acid changes are unlikely to have a large effect. Further investigation of the functional effect of these remaining amino acid changes is needed to prove the compelling hypothesis regarding the role of N-terminal residues have in catalysis. In particular, future work should aim to explore the 284 D→E region, which is predicted to have at least a minor functional effect.

3.3.4 EnOMT1 is closely related to class II OMTs

The overall tertiary structure of EnOMT1, with an N-terminal dimerization domain and C-terminal catalytic domain, is consistent with those previously reported plant SAM OMTs: *Medicago sativa* Caffeic acid OMT (MsCOMT; PDB 1KYZ), *M. trunculata* isoflavonoid OMT (MsIOMT; PDB 1FP2), and *M. sativa* chalcone OMT (MsChOMT; PDB 1FP1) (Zubieta et al., 2001; Zubieta et al., 2002).

The homology modelling predicted EnOMT1 to be a dimer. Most plant OMTs form a dimer in *vitro* and all structurally characterised OMTs to date show dimerization in solution (Liscombe et al., 2012). Dimer formation confirmation was attempted with the recombinant EnOMT1 using native PAGE electrophoresis, but definitive results were not forthcoming. Considering the nature of catalysis described at the dimer interface (Liscombe et al., 2012), it is reasonable to suggest that EnOMT1 forms a homodimer *in vivo*. Molecular docking analysis results showing the involvement of N-terminal residues of the neighbouring monomer (Met15, Tyr19) in the binding of substrate pinocembrin further supports the importance of homodimer formation for catalysis. Also, the deep burial of substrate binding sites inside the predicted active site via dimer formation seems apt for the binding of highly hydrophobic flavanone substate.

The topology predicted by the Phyre2 and MEMSAT-SVM program showed the presence of a small N-terminal pore-lining transmembrane helix. This result is confounding as there is no evidence of any pore-lining regions mentioned in the structurally characterised flavonoid and alkaloid OMTs (Liscombe et al., 2012). The solved crystal structure of plant OMTs has an N-terminal region involved in protein dimerization with the dimer interface playing a critical role in the activity (Ferrer et al., 2005; Zubieta et al., 2001). Structural comparison of EnOMT1 with *Medicago sativa* IOMT (PDB: IFP2) quaternary structure (**Figure 3.17**-A) showed that the predicted pore-lining helix region is not directly involved in substrate binding. Moreover, previous studies have suggested that an N-terminal transmembrane region of plant sterol *C*-methyltransferase is responsible for its localization to and association with the endoplasmic reticulum (Husselstein et al., 1996). This, along with the notion that OMTs are a part of a multienzyme complex on the surface of the endoplasmic reticulum (Ibrahim et al., 1987), strengthens the likelihood of such a region being present. Furthermore, since we know that the methylated flavanones are localised in gland cavities (Goodger et al., 2019), the possibility of having a pore region that can act as a transporter channel for these molecules in and out of the cells cannot be disregarded.

Phylogenetic analysis of the three isolated *Eucalyptus* OMTs was performed with characterised representatives from a different class of plant phenolic compound methyltransferase to see if there is a trend in the evolution of these genes. Gene duplication and subsequent divergence are considered as the main driving forces behind the evolution of enzymes with novel functions (Ohno, 1999). Specifically, in the case of plant OMTs, there have been studies suggesting that plant OMTs involved in floral scent production appear to have evolved by gene duplication followed by mutation (Gang, 2005). The phylogenetic analysis of the enzymes with 56 biochemically characterized plant OMTs accepting a broad variety of substrates (**Figure 3.13**) showed that xanthohumol 4-*O*-methyltransferase (X4OMT; Uniprot ID- B0ZB56.1) from *Humulus lupulus* is the closest evolutionary relative. Just like in *Eucalyptus*, the glandular trichomes and lupulin glands of *H. lupulus* are composed of secretory cells which make xanthohumol, a prenylchalcone (a terpenophenolic) that accumulates it in the extracellular gland lumen along with the essential oils (Nagel et al., 2008; Oliveira and Pais, 1990). This striking resemblance of the biosynthetic systems of EnOMT1 and X4OMT may be the reason behind the minimal evolutionary distance between the two. Phylogenetic analysis of plant methyltransferases and their classification has focused on the nature of methyl acceptors (Lam et al., 2007; Liscombe et al., 2012). Perhaps the metabolic specialization of different cell types and the physiological location of methyl transfer might be the driving force of the evolution of enzymes with a new function.

3.4 Conclusion

The research described in this chapter reports biochemical and structural characterisation of EnOMT1 and isolation and comparison of homologous enzymes of EnOMT1 from related species. Kinetic parameters of EnOMT1 were estimated with pinocembrin as the substrate and the estimated values are comparable to existing characterised plant MTs. Substrate specificity analysis showed that EnOMT1 is a strictly position-specific flavonoid OMT methylating 7-*OH* of flavonoid compounds. Methylation was not achieved with any of the simple phenolics, phenylpropanoids and chalcone tested. Sequence comparison of the EnOMT1 to the orthologues genes from related species with different complements of methylated flavanones enabled the identification of amino acid residues, that I propose, influence the catalytic efficiency of EnOMT1 by involving in ligand binding. The impact of some these suggested residues on catalysis is confirmed through site-directed mutagenesis and *in vitro* enzyme assays. A further study could assess the structural as well as functional effect of the remaining amino acid changes.

The findings of substrate specificity assays suggest presence of another 5-*OH* specific OMT in the synthesis of alpinetin from pinocembrin. This would be a fruitful area for further work. Further research could explore whether post-translational modifications of EnOMT1 alter its regioselectivity. Transgenic expression of EnOMT1 in a model plant system may enable further investigation of translational modifications. Homology modelling of EnOMT1 along with molecular docking showed the involvement of dimerization domain in catalysis. Therefore, additional parameters like binding efficiency as well as *in vitro* dissociation constant of the homodimer complex should also need to be considered for future studies on flavanone OMTs in *Eucalyptus*.

CHAPTER 4

FLAVONONE *C*-METHYLATION IN *EUCALYPTUS*

4.1 Introduction

Plant methyltransferases are generally grouped according to the target atom they transfer the methyl group to, such as *O*-, *C*-, *N*- and *S*- methyltransferases. In plants *N*-methylation generally occurs in nitrogen-containing alkaloids and *S*-methyltransferases act on the sulphur atom of specific thiols (Coiner et al., 2006; Zhao et al., 2012). *O*-methylation (the transfer to an oxygen atom or carboxyl group) is also a widespread mechanism in plants, whereas *C*-methylation is very rare and appears restricted to certain plant families. *C*-methylation involves the transfer of a methyl moiety directly to a carbon atom of the substrate backbone by the enzyme *C*-methyltransferase (CMT).

Several *C*-methylated flavonoids have been isolated and identified from various plants. Courtney et al. (1983) suggested *C*-methylation as a distinctive feature of the family *Myrtaceae*, which was later supported by several other studies (Wollenweber et al., 2000) (Dao et al., 2010). Subsequently, additional plant families such as *Nyctaginaceae* (Sutthivaiyakit et al., 2013) and *Myricaceae* (Malterud et al., 1996) have also been reported to possess *C*-methylated flavonoids. Nonetheless, a detailed assessment of published studies (section **1.3.2.iv**) indicates that the majority of *C*-methylated flavonoids are found in the family *Myrtaceae*. Recently, several *C*-methylated flavanones have been found in plants from the Myrtaceous genus *Eucalyptus* subgenus *Eucalyptus*, including unsubstituted B-ring flavanones cryptostrobin, strobopinin, desmethoxymatteucinol, and *5*-hydroxy-*7*-methoxy-*6,8*-dimethylflavanone (Goodger et al., 2016; Marsh et al., 2019).

Likely due to the limited number of species that contain them, the process of flavonoid *C*-methylation has not been well studied and there remains several aspects about which little is known. Some information is available on the isolation and detection of the *C*-methylated compounds from particular plant sources, but no sequence data or cDNA clones are available for CMTs catalysing methylation of flavonoids or any other related phenolic compounds. The limited sequence information for CMTs from plants include the *C*-24 sterol methyltransferase from *Zea mays* and *Glycine max* and a γ-tocopherol methyltransferase from *Arabidopsis thaliana* and *Capsicum annuum* (Grebenok et al., 1997; Koch et al., 2003; Shi et al., 1996a). The biosynthetic route of *C*-methylation in flavonoids is therefore unclear.

Some information on the biochemical pathway for *C*-methylation is available from other systems. For example, bacterial CMTs catalyse a direct transfer of the *C*-methyl group to phenolic compounds (Tengg et al., 2012). Therefore, it is possible that the *C*-methyl group in plants is also incorporated into flavanone aglycones or *O*-methyl compounds through a direct enzymatic transformation (Goodger et al., 2019). Another possible route is the one proposed by Schroder et al. (1998) for *Pinus strobus* where a methyl group is incorporated into a chalcone backbone, through a condensation reaction of methylmalonyl-CoA with a diketide intermediate, subsequently forming *C*-methylated flavanones circumventing aglycone formation (Schröder et al., 1998). Nevertheless, in plants, the presence of methylmalonyl-CoA *in vivo* has not yet been detected. With our success in flavanone *O*-methyltransferase isolation using the transcriptomic approach described in Chapter 2, we decided to follow a similar approach as an initial test to investigate flavanone CMTs in *Eucalyptus*. The results from chapters 2 and 3 show that an OMT acted on the flavanone aglycone pinocembrin, therefore the possibility of CMTs following a similar mechanism in *Eucalyptus* is investigated in this chapter.

Various *Eucalyptus* subgenera also accumulate several β-triketones, a rare cyclic metabolite with three ketonic carbonyl groups often with multiple *C*-methylation or *C*- and *O*-methylations (Goodger et al., 2018; Hellyer, 1968). β-triketones are also found in other *Myrtaceae* genera such as *Leptospermum*, *Darwinia*, *Calytrix*, *Backhousia*, *Baeckea, Xanthostemon* and *Melaleuca* (Ghisalberti, 1996). The most common β-triketones in *Eucalyptus* are those derived from a tetra-methylated phloroglucinol ring including agglomerone, tasmonone, grandiflorone, and conglomerone (Hellyer, 1968; Jones and Lahey, 1939).

Another aspect of the *O*- and *C*-methylated flavanones that requires more research attention is their specific function in *Eucalyptus* as well as *ex planta* production for commercial applications. Synthesis of the medicinally important natural products *ex planta* is a good approach to utilise their biological activities in the pharmaceutical field. High-yielding cell lines of many medicinal plant cell suspension cultures have been successfully used as an *ex planta* production platform for harvesting plant metabolites, and different strategies like precursor feeding, elicitor mediated induction, and UV irradiation are currently used to enhance production (Yue et al., 2016). Increased production of a metabolite in response to external stimuli implies its involvement in a particular stress response (Lois, 1994). For example, an increase in activity of flavonoid *O*-methyltransferase was seen in *Petroselinum crispum* cell cultures exposed to UV light stress (Hahlbrock and Grisebach, 1979). Similarly, fungal elicitors of cell cultures of *Pisum sativum* and *Glycine max* resulted in the accumulation of *O*-methylated flavonoids.

The primary objective of research described in this chapter is to understand flavonoid *C*-methylation in *Eucalyptus* by isolation and identification of the *C*-methyltransferases involved. A transcriptomics approach for gene isolation necessitates finding a suitable candidate species rich in *C*-methylated flavanone compounds, therefore, in the first section of this chapter an extensive screening of various *Eucalyptus* species is described, and *E. diversifolia* was chosen as the candidate species for transcriptome sequencing. Analysis of the sequenced transcriptome enabled the identification of putative CMT candidates, and these candidates are tested for *C*-methylation activity. Additionally, this chapter aims to establish *Eucalyptus* callus and cell suspension cultures to explore the feasibility of *ex planta* production of methylated flavonoids. The developed cultures will be used to explore the ecological significance of methylated flavanones through their response to abiotic and biotic stress and to see if such stresses can be used to enhance production in a commercial setting.

4.2 Materials and methods

Plant materials: All plant tissues used in this study were collected from plants grown at the University of Melbourne Dookie campus except for three individual *E. agglomerata* trees that were grown at the University of Melbourne Parkville campus. *E. diversifolia* seedlings (seeds collected from Portland area, VIC) were purchased from ERA Nurseries, Hamilton, and planted at Dookie in May 2017. Seed source, provenance and propagation details of the rest of the species used in this study are given in Appendix D.

Figure 4.1: *Eucalyptus diversifolia* plant used for *C*-methyltransferase study. A) Three-year-old *E. diversifolia* tree in The University of Melbourne Dookie Campus plot. B) A close-up picture of flowering *E. diversifolia* showing leaf structure.

Chemicals*:* All chemicals used were of analytical grade unless specified otherwise. Milli-Q purified water was used for HPLC analysis. HPLC grade organic solvents were purchased from Merck. Chloroform and isopropanol were purchased from Chem-Supply and 2% CTAB extraction buffer was purchased from TEKnova. NaCl was purchased from Merck Australia. All other reagents and materials used in molecular biology experiments were purchased from Thermo Fisher Scientific Australia Pty Ltd, if not specifically stated otherwise.

Bacterial culture media: The media composition used to grow bacteria were as described in Chapter 2.

4.2.1 Profiling of non-volatile compounds from *Eucalyptus* species

The same procedure described in Chapter 2 was followed for flavanone screening, analysis, and detection. A fully developed leaf was ground to powder and 0.1g of the ground material was used for extraction. When required, individual foliar oil glands were isolated following the procedure described by Goodger et al (2010) using the enzyme pectinase from *Aspergillus niger* (Sigma-Aldrich, St. Louis, USA). Briefly, a fully developed leaf was cut into small strips approximately 20 mm length, immersed in 1.5 mL of diluted pectinase enzyme (1:2 v/v with distilled water) and incubated at 30 °C in a thermomixer shaking at 600 rpm for 30 s at 1 min intervals. The digestion was stopped when the epidermis peeled off easily from the rest of the leaf tissue (approximately after 12-16 hr depending on the species). Oil glands, carefully detached from leaf tissue using fine forceps and brush under a dissecting microscope, were sieved through, washed, and collected using micropipettes into a clean Eppendorf tube. An equal number of glands was collected in two different tubes and ground using a micro-pestle and extracted in 200µL of 100% acetonitrile for non-volatile analysis and 200 µL of 100% hexane for volatile component analysis, respectively.

Volatile components in the extract were analysed by Gas Chromatography with Flame Ionisation Detection (GC-FID) prior to the mass spectrometry. Ground oil glands were extracted with hexane, dehydrated with anhydrous Na_2SO_4 and then injected into the GC-FID (Perkin Elmer Autosystem XL; Perkin Elmer, Melbourne, Australia) fitted with a Zebron ZB-5 low polarity column (30 m × 0.25 mm i.d., Phenomenex, USA). The detector temperature was 220 °C, injector temperature was 250 °C and helium (1 mL min^{-1}) was the carrier gas. The column temperature was held at 120 °C for 1 min following injection, increased to 180°C at 7°C min^{-1} and then held isothermally for a further 10 min.

The hexane extracts were also subjected to GC-MS analysis for compound identification following the procedure described in Goodger et al. (2018) using an Agilent 7890A GC and 5975C MS (Agilent Technologies, Santa Clara, USA). Mass spectra acquisition and analysis was carried out using Agilent MSD ChemStation (version E.02.02.1431). β-triketones were identified based on the Adams (2017) mass spectra library.

4.2.2 Native enzyme extraction and assay

Native methyltransferase extraction was carried out following the method described previously (Scheid et al., 1982). Fully grown leaves of respective *Eucalyptus* species (1g) were ground to powder in the presence of liquid nitrogen and polyvinylpolypyrrolidone. The ground powder was extracted with 1 ml of ice-cold extraction buffer (0.1M potassium phosphate buffer, pH 7.5, containing 1 mM EDTA, 10 mM β-mercaptoethanol, and 1X protease inhibitor cocktail tablet). The protein extract was separated from leaf debris via centrifugation at 6000 x *g* for 8 min. A fraction of the extract was taken as a crude extract for analysis. The remaining extract was passed through a desalting column to remove all small molecule contaminants like secondary metabolites. The resulting mixture was then used as a native enzyme extract. The enzyme assay mixture composition was the same as used for OMT recombinant assay with a slight modification. To prevent dilution, enzyme mixture in buffer solution was added instead of fresh buffer i.e., for a 250 µl reaction mixture 225 µl enzyme mixture was mixed with substrate and SAM.

4.2.3 Establishment of callus/cell suspension cultures of *Eucalyptus species.*

Leaf and internodal explants from *in vitro* germinated seedlings of *Eucalyptus croajingolensis* and *E. nitida* and *E .diversifolia* were used to develop cell suspension cultures. Callus was induced on woody plant medium (Fernando et al., 2016) supplemented with 0.02 mg L^{-1} NAA, 0.66 mg L^{-1} TDZ, 25 g L^{-1} sucrose and 0.25% Gerlrite (Media MM2). Callus generation was also done with glasshouse-grown fully developed leaves. Produced calli were multiplied by transferring to fresh media of the same composition in 8 weeks interval. Approx. 5g of callus was transferred to a 250 ml Erlenmeyer flask containing 100 ml liquid MM2 (same composition as above without Gelrite) and incubated on a shaker at 100 rpm, 26 ± 1˚C in dark to initiate cell suspension culture. For subculture, 5 ml of cells from the parent culture were pipetted out to a new 250 ml Erlenmeyer flask with 25 ml of spent/conditioned medium (medium in which the cells had been growing), and 25 ml fresh liquid medium. The cell suspension was maintained by subculturing every 4 weeks.

The growth of the cells in suspension cultures was monitored in every four days interval for thirty days. The total culture was poured into a sterilized measuring cylinder and the settled cell volume was measured by allowing the cells to settle in graduated cylinders for one hour. A growth curve was plotted with cell settle volume along the Y-axis and the number of days after subculture along the Y-axis. Microscopic examination of cell suspension cultures of *E. croajingolensis* and *E. nitida* was done using a stereomicroscope (Nikon SMZ 1500) to analyse the cell type and condition of the suspension cultures.

4.2.4 UV light exposure studies

A group of 24 healthy-looking glasshouse grown *non-methylated species* seedlings (average height=18.5cm) were selected for the study. Half were kept in a chamber with normal light (control) and the other half were kept in a chamber with both normal and UV-A light. Both chambers were under 16 h photoperiod. Chamber conditions (**Table 4.1**) were monitored continuously. Plants were watered manually twice per week with 400 mL of tap water. Leaves grown under these conditions were extracted and analysed for flavanone content following the same procedure described in Chapter 2.

Table 4.1: Growth chamber conditions used for UV-light exposure studies

	Control	UV
Average temperature	26.19 ˚C	26.28 ˚C
Average% relative humidity	40.98	41.67
Average light intensity (µmol $m^{-2}s^{-1}$)	107.42	109.2

4.3 Results

4.3.1 Screening of *Eucalyptus* species for an ideal *C*-methyl flavanone rich species

A detailed screening of ten species of *Eucalyptus* was performed to identify a species with a high quantity of *C*-methylated flavanones. The selected representatives covered the *subgenus Eucalyptus* (monocalypts; **Table 4.2**). Initially, twelve individual plants from each species namely, *E. diversifolia; E. agglomerata; E. preissiana; E. muelleriana; E. stellulata; E. mitchelliana; E. haemastoma; E. croajingolensis; E. apiculata;* and *E. nitida* were analysed for their leaf contents and their metabolic profile is given in **Figure 4.2**

Table 4.2: List of species selected for flavanone profiling and their taxonomic classification (Brooker, 2000).

Species	**Subgenus**	**Section**	**Series**
E. croajingolensis	*Eucalyptus*	*Aromatica*	*Radiatae*
E. nitida	*Eucalyptus*	*Aromatica*	*Insulanae*
E. muelleriana	*Eucalyptus*	*Capillulus*	*Pachyphloius*
E. agglomerata	*Eucalyptus*	*Capillulus*	*Pachyphloius*
E. apiculata	*Eucalyptus*	*Eucalyptus*	*Strictae*
E. haemastoma	*Eucalyptus*	*Cineraceae*	*Psathyroxylon*
E. mitchelliana	*Eucalyptus*	*Longitudinales*	*Longitudinales*
E. stellulata	*Eucalyptus*	*Longitudinales*	*Longitudinales*
E. diversifolia	*Eucalyptus*	*Longistylus*	*Diversiformae*
E. preissiana	*Eucalyptus*	*Longistylus*	*Preissianae*

In all ten species studied here, total flavanone concentration varied significantly between species. *E muelleriana* showed the least average concentration of flavanones with traces of pinocembrin. Pinostrobin dominated the profiles of *E. apiculata*, *E. nitida* and *E. croajingolensis*. Alpinetin was the least common flavanone detected and trace amounts were seen in several individual trees of *E. agglomerata*, *E. haemastoma* and *E. stellulata*. The observed difference in flavanone profile between species, as well as individuals within a given species, is detailed in section 4.3.5. This section will focus on *C*-methyl flavanone accumulating species. Previous research from our lab has identified cryptostrobin (CS) and desmethoxymatteucinol (DMM) in acetonitrile extracts of *E. agglomerata* leaf samples (harvested from Currency Creek arboretum trees; Goodger et al., 2019). Therefore, initial identification was done by the retention time comparison with this *E. agglomerata* extract (which had been stored at -20 °C). When analysed using the HPLC method described in Section 2.2.2, CS and DMM eluted at a retention time of 11.9 min and 12.9 min, respectively. Nevertheless, acetonitrile extract of *E. agglomerata* samples collected from Dookie campus or Parkville campus plots showed no significant peaks corresponding to CS or DMM, and instead they showed β-triketone in abundance and traces of *O*-methylated flavanones in some individuals which are detailed in section 4.3.5. *E. diversifolia* was chosen to be the candidate species for study as it had a considerably high quantity of *C*-methylated flavanones CS and DMM (**Figure 4.2**). Further confirmation of the identity of the *C*-methylated flavanones was aided by comparison of UV absorbance and retention times with authentic standards.

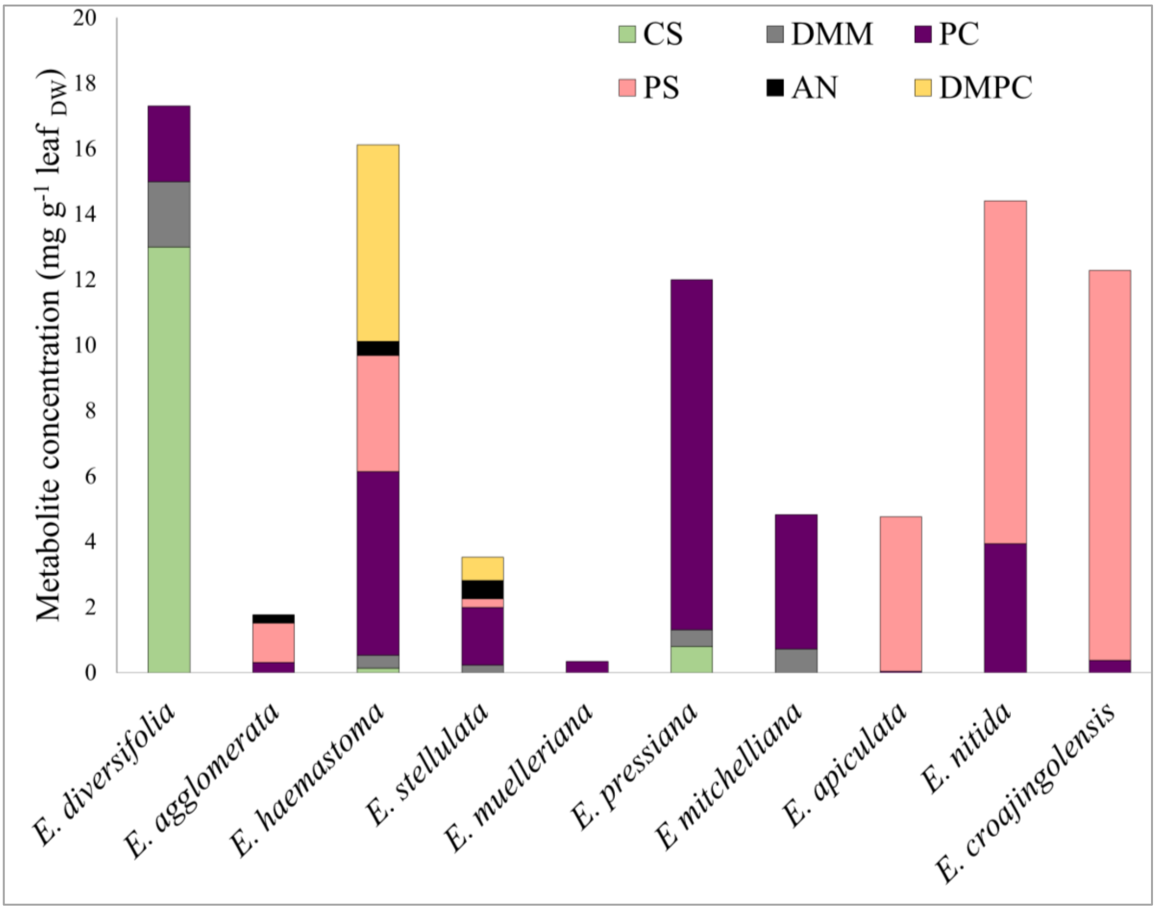


Figure 4.2: Foliar flavanones in candidate species. The depicted data are the average values of twelve individual plants except for *E. apiculata* which was based only on two plants. CS- cryptostrobin; DMM- desmethoxymetteucinol; PC- pinocembrin; PS- pinostrobin; AN- alpinetin; DMPC- dimethylpinocembrin. The concentration of CS and DMM were calculated in PC equivalence.

4.3.2 Full-length transcriptome analysis of *E. diversifolia*

To isolate *C*-methyltransferases from *Eucalyptus*, transcriptome sequencing of *Eucalyptus diversifolia*, a species with a significantly high concentration of the *C*-methylated flavanone cryptostrobin was performed (**Figure 4.2**). Total RNA was successfully extracted from leaf tips. The concentration of RNA sample analysed on a NanoDrop spectrophotometer was 487 ng/µL and the A260/A280 ratio was greater than 2.1 for all the samples sent for sequencing. The RNA sample passed the quality check for library construction and sequencing with an RNA integrity number of 8.9.

i. <u>Transcriptome reconstruction</u>

The *E. diversifolia (C-methylated species)* sequencing raw data had 38,285,060 raw read counts. The error rate distribution analysis showed that the quality of the sequence across all base reads is high with an average sequencing error rate of 0.01%.

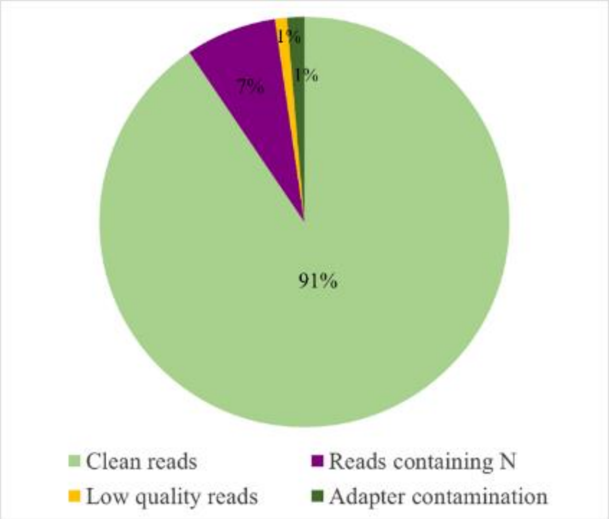


Figure 4.3: Classification of raw reads for *C*-methylated species highlighting clean reads.

Reads that contained more than 10% N (uncertain bases) and adaptor contamination were removed to improve the quality and 91% of *E. diversifolia* raw reads were selected as high-quality reads (clean reads) for further analysis (**Figure 4.3**). A summary of the raw data quality check is given in **Table 4.3**.

Table 4.3:Transcriptome sequencing quality control summary showing raw reads and clean reads

Sample Name	Raw Reads[1]	Clean Reads[2]	Clean Bases[3]	Error (%)	Q20[4] (%)	Q30[5] (%)	GC Content(%)
E. diversifolia	38285060	34651996	5.2G	0.01	96.16	90.43	49.69

[1]Raw Reads-the original sequencing reads counts; [2]Clean Reads-number of reads after filtering; [3]Clean Bases-clean reads number multiplied by read length in giga bases (G); [4] Q20 and [5] Q30-percentages of bases whose correct base recognition rates are greater than 99% and 99.9% respectively.

Clean reads were *de novo* assembled by Trinity software (Grabherr et al., 2011). Assembled unigene length ranged from 200-9100 bp with a mean length of 1278 bp (**Figure 4.4**A).

ii. Functional annotation

The assembled unigenes were then annotated based on their sequence similarity with entries in seven public databases [NCBI nucleotide sequences (Nt), NCBI non-redundant protein sequences (Nr), Swiss-Prot, Protein family(Pfam), euKaryotic Orthologous Groups (KOG), and Gene Ontology databases(GO), Kyoto Encyclopedia of Genes and Genome(KEGG/KO)].

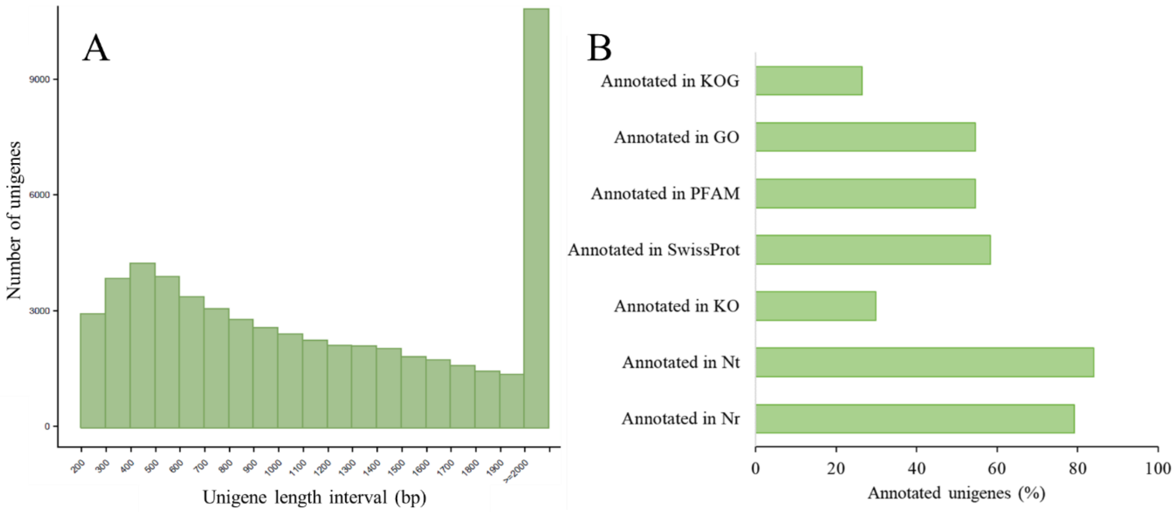


Figure 4.4: Unigene characteristics: A) unigene length distribution, B) Annotation summary showing the percentage of unigenes annotated via seven different databases.

Out of the total 56,772 *E. diversifolia* unigenes, 84% were annotated to the NCBI-Nt database and 79% to the NCBI-Nr database (**Figure 4.4**B). Among the NCBI-Nr and SwissProt annotated unigenes, 89.3% of unigenes had matching sequences in *Eucalyptus grandis* followed by *Vitis vinifera* (1.4%) and *Theobroma cacao* (1.3%) (**Figure 4.5**A). BLASTx similarity analysis demonstrated a very low E-value distribution which emphasises significant homology and a high annotation accuracy (**Figure 4.5**B). In summary, a total of 87.4% of unigenes were annotated with significant blast score in at least one database and 22% of *E. diversifolia* unigenes were annotated in all databases (**Figure 4.5**C).

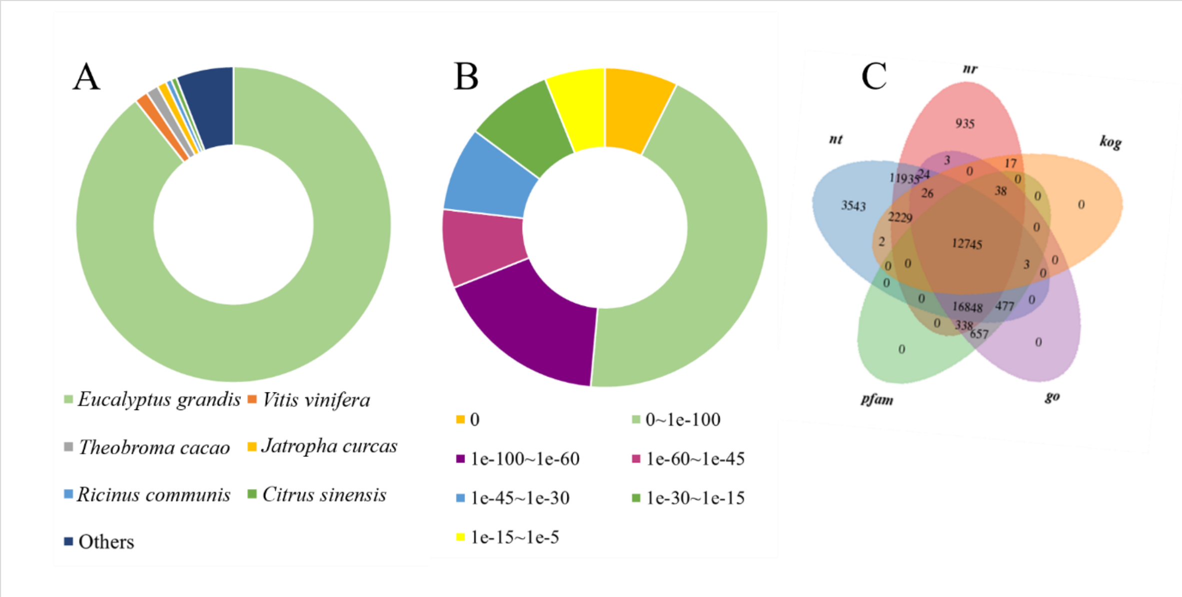


Figure 4.5: Characteristics of functional annotation by unigene similarity against various databases. A) Species distribution as seen in BLASTx top hit of *E. diversifolia* unigenes against NCBI-Nr and Swiss-Prot database. B) E-value distribution of BLASTx hits against NCBI-Nr database. C) Venn-diagram showing the number of unigenes annotated to NCBI-Nt, NCBI-Nr, PFAM, KOG, and GO databases. Abbreviations are as follows: Nr- NCBI non redundant protein database, Nt- NCBI nucleotide database, KOG- eukaryotic orthologous groups, GO- gene ontology database, Pfam- Protein family database.

GO enrichment analysis was used to classify the functions of the full-length transcripts. GO is an international system that standardises the gene function of orthologous genes across different organisms to facilitate global analysis (Consortium, 2008). Based on functional attributes, the 31,159 GO annotated genes were grouped into three main ontology domains: Biological Process, Cellular Component, and Molecular Function (**Figure 4.6**). A second level grouping shows the specific functional category of the putative genes in the main domain. Each unigene may fall into more than one functional category. Notably represented categories under the biological process domain include cellular process, metabolic process, and single organism process. Specifically,16,365 genes were suggested to be involved in the metabolic processes category, which potentially includes the flavanone biosynthesis genes under investigation (**Figure 4.6**).

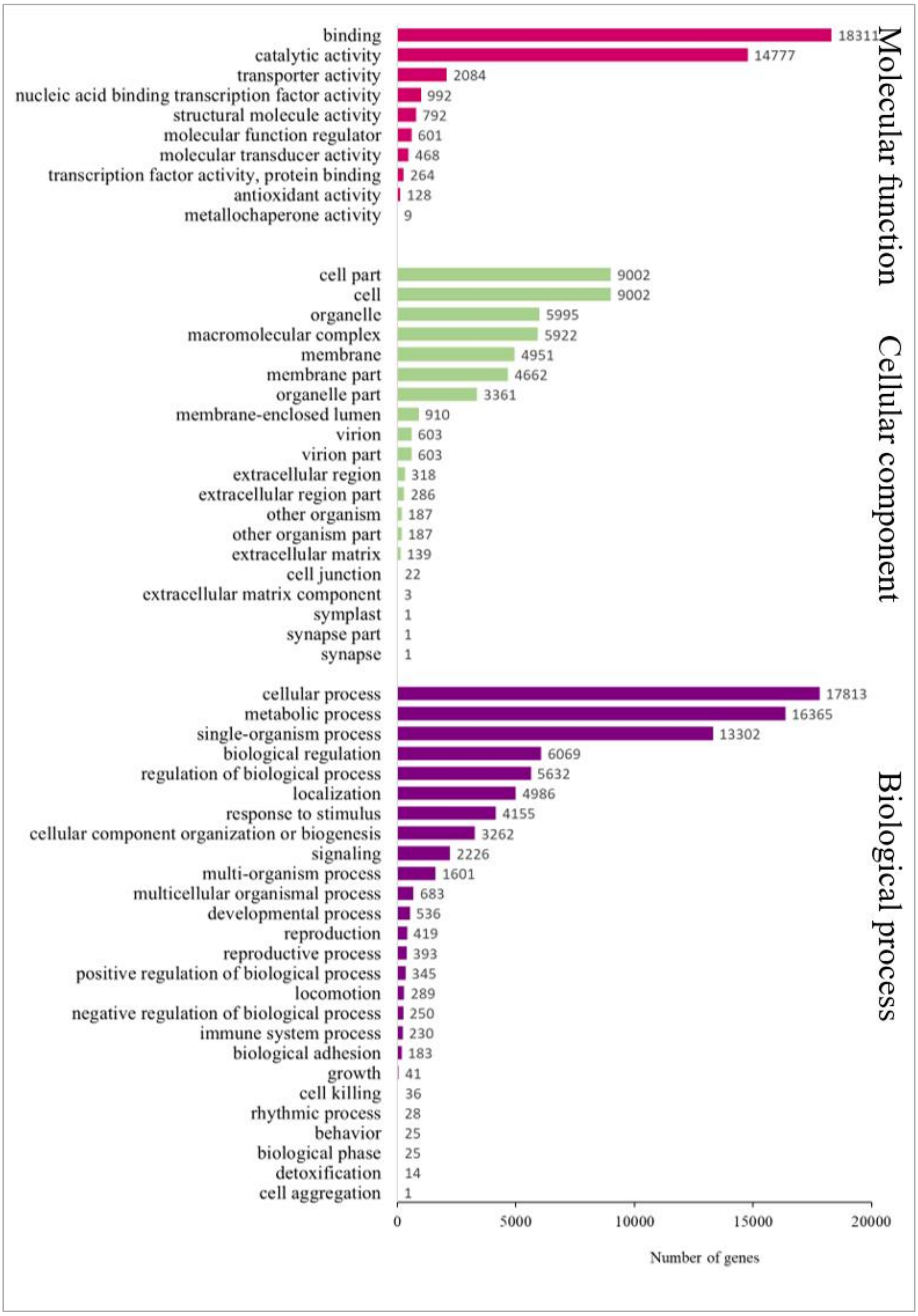


Figure 4.6: GO classification of *E. diversifolia* unigenes arranged into three main domains: molecular function, cellular component, and biological processes.

After KOG functional annotation, 15060 unigenes were grouped into the twenty-five KOG functional categories. More than 450 unigenes were found to be putatively involved in secondary metabolite biosynthesis, transport, and catabolism; and 780 genes were defined as having an unknown function; and 225 unigenes were predicted to be involved in CoA transport and metabolism (**Figure 4.7**).

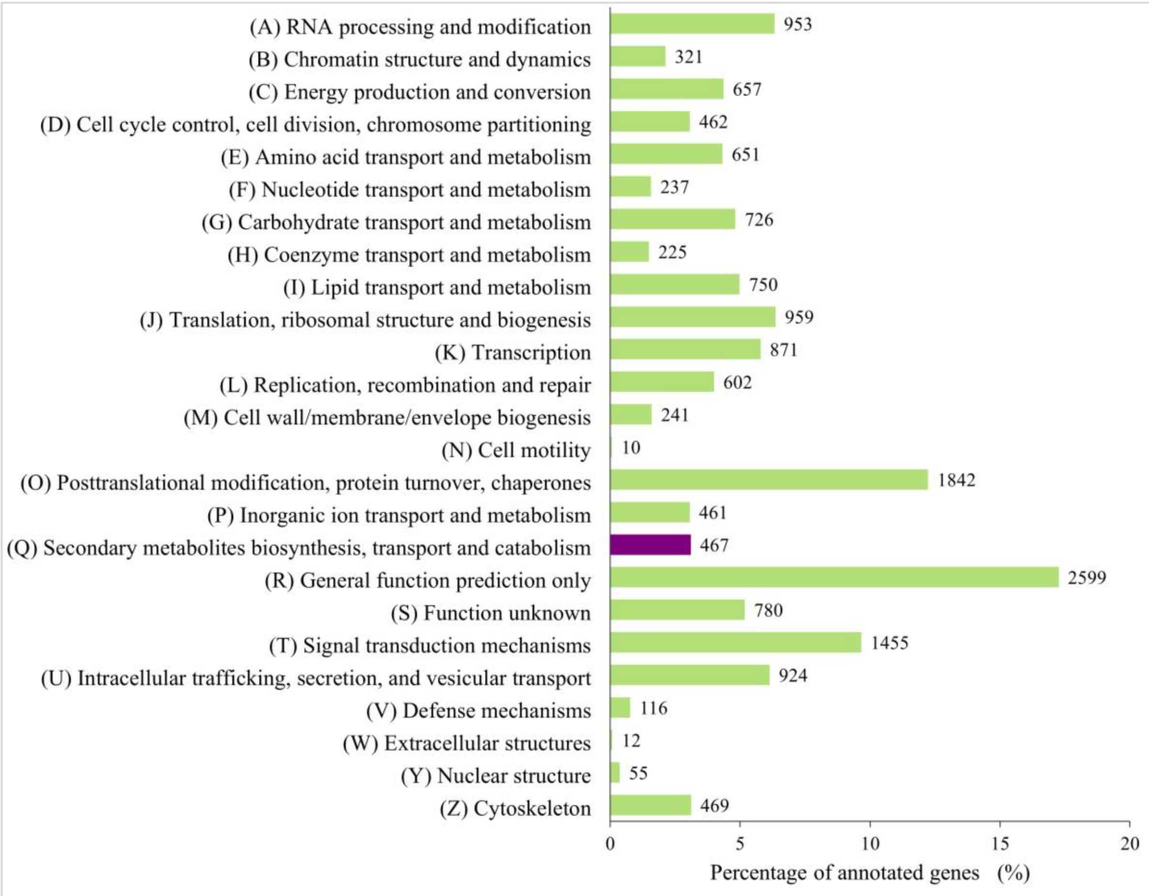


Figure 4.7: KOG function classification of unigenes from *E. diversifolia* leaf tip transcriptome. Different categories of KOG function classification (A-Z) are listed along the Y-axis and the percentage of annotated genes along the X-axis. The numbers of genes in a particular category are denoted alongside the bar.

Biological processes involve the cooperation of many genes in a metabolic pathway and pathway-based analysis helps in understanding the biological function of genes. The KEGG pathway database enables systematic evaluation of biological functions of genes. In total, 17,000 unigenes were classified into various KEGG pathways clustered in five main functional categories: cellular processes, environmental information processing, genetic information processing, metabolism, and organismal systems. The majority fell into the metabolism category followed by genetic information processing. Over 330 unigenes were found to be specifically involving in terpenoid and polyketide metabolism (**Figure 4.8**).

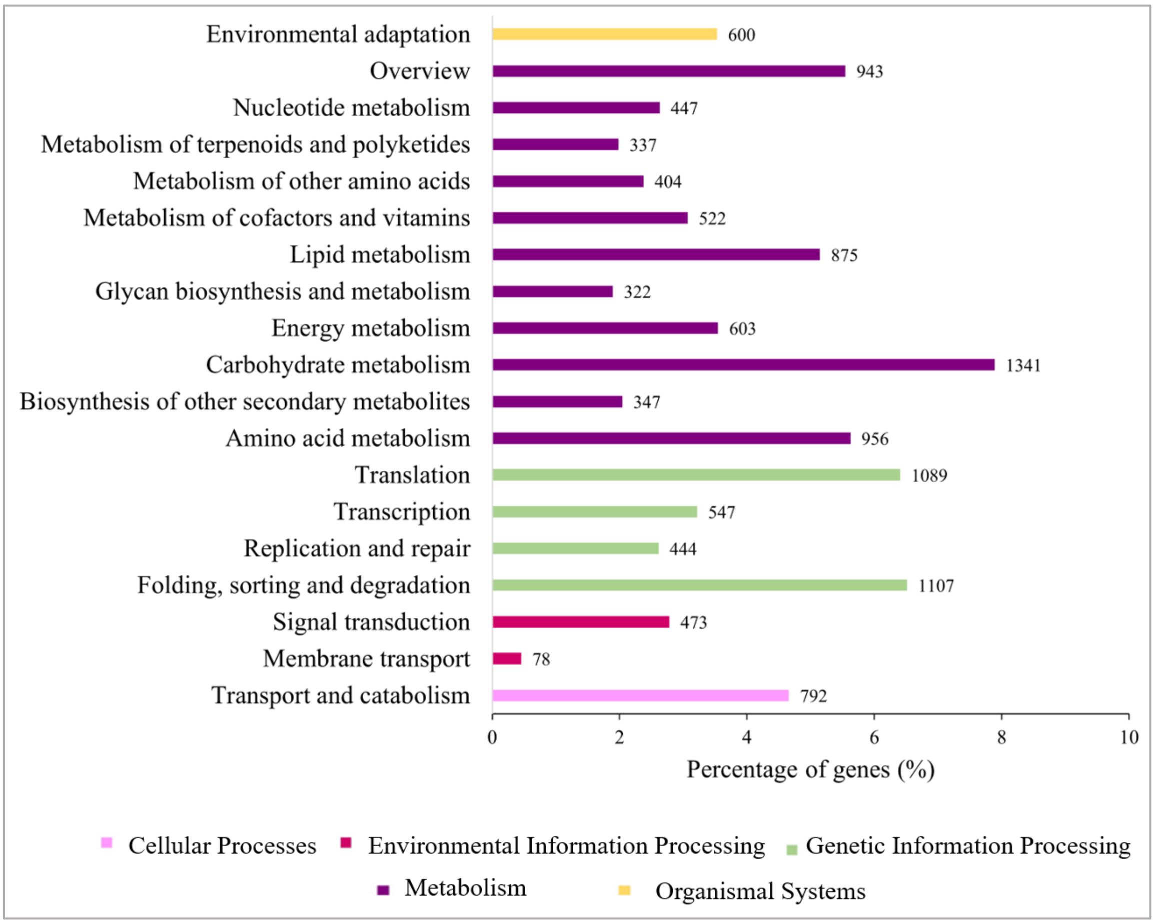


Figure 4.8: KEGG pathway allocation of *E. diversifolia* leaf tip transcriptome. Percentage of genes based on a total annotated gene of 17000 (X-axis) involved in various biological processes (Y-axis). All human-disease-related results were removed.

4.3.3 Isolation, expression, and analysis of putative genes involved in *C*-methyl flavanone biosynthesis

Following transcriptome sequencing, putative candidates were chosen using the same approach used for *O*-methyltransferase identification in Chapter 2. After functional annotation by the Nr database, 88 unigene clusters were identified to contain the methyltransferase domain. The highly expressed methyltransferase genes in *C-methylated species* were compared against the homologous gene in *O-methylated* as well as *non-methylated species* (**Figure 4.9**). CMT gene expression is expected to be very low in the species with no *C*-methyl flavanones.

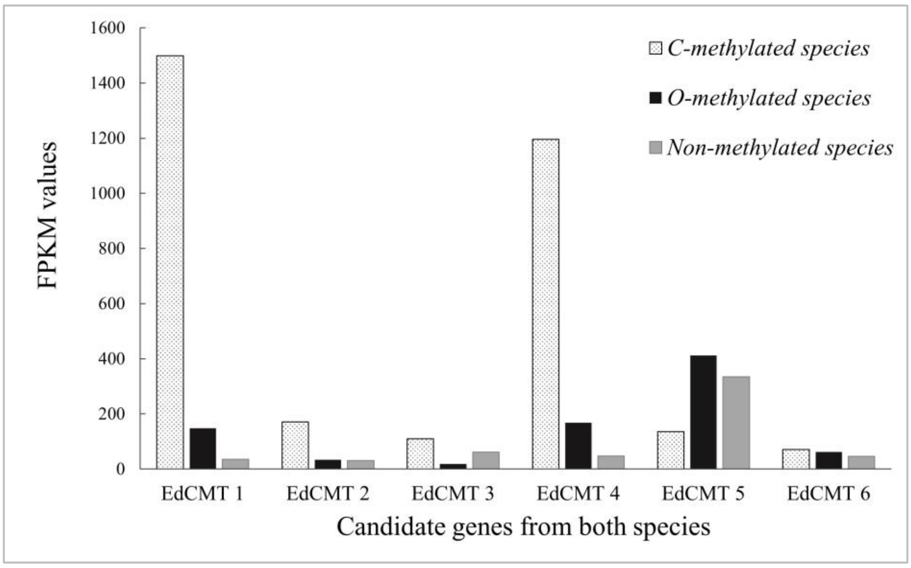


Figure 4.9: Putative CMT candidates from *E. diversifolia (C-methylated species)* with corresponding FPKM values compared to respective orthologous gene from *E. nitida* (*O-methylated species 1*) and *non methylated species*. Unigene cluster names of the putative genes are as follows: EdCMT 1- Cluster-3874.9282; EdCMT 2- Cluster-3874.13060; EdCMT3- Cluster-3874.2318; EdCMT 4- Cluster-4240.0; EdCMT 5- Cluster-3874.7814; EdCMT6- Cluster-3874.8563. Full nucleotide sequence of these clusters from the transcriptomics data is given in Appendix C.

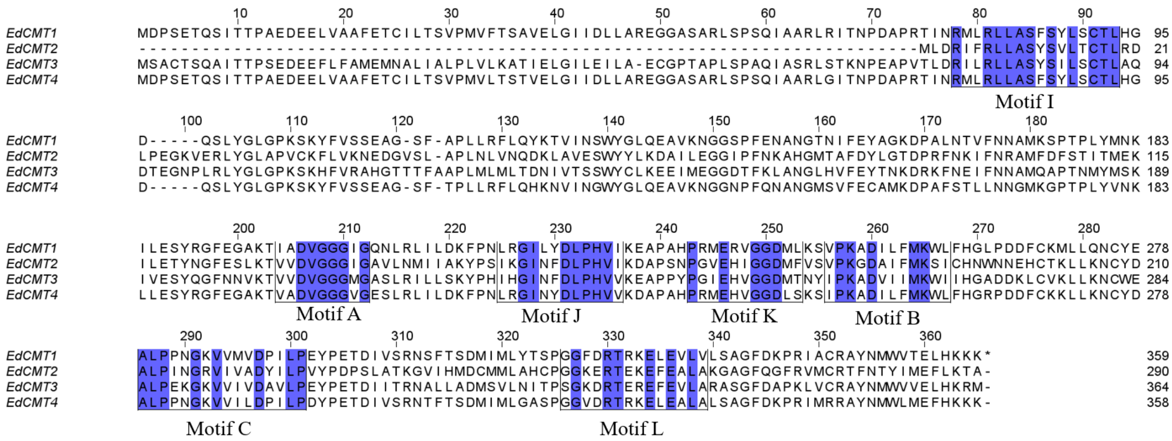


Figure 4.10: Predicted coding sequence alignment of CMT candidates showing methyltransferase conserved motifs. Motif regions suggested by Joshi and Chiang (1998) are highlighted with identical residues in each motif shaded in blue.

Based on the sequence similarity and differential expression pattern, initially, four putative CMT candidates (EdCMT 1-EdCMT 4) were selected. **Figure 4.10** shows multiple sequence alignment of these candidates with conserved binding motifs. The SAM-binding motifs A, B and C were present with at most two mismatches to the suggested consensus motif (Joshi and Chiang, 1998). Four additional motifs (I, J, K, L) that are specific to Mg^{2+} independent Pl-OMTII class methyltransferases were also present. Furthermore, the four CMT candidates also have the spatial relationship of 52 amino acids between the first amino acid of motif A and motif B and 30 amino acids between motifs B and C. All four putative CMT genes were successfully cloned into the pHUE vector system and expressed in *E. coli* BL21 strain. Expression was confirmed with dot-blot assays. A set of flavanone compounds (pinocembrin, pinostrobin, alpinetin, pinocembrin chalcone) were trialled as substrates in an activity assay. Unfortunately, the first four candidates did not convert the substrates into a *C*-methylated product. Therefore, the search was expanded to two more putative CMT candidates. EdCMT 5 and EdCMT 6 were selected based on their sequence similarity (42% and 86%, respectively) with a previously characterised plant sterol *C*-methyltransferase from *Glycine max* (Shi et al., 1996b). Unlike the other four selected CMT candidates, the spatial separation between motif A, B and C differed in EdCMT 5 and 6. There are 121 amino acids between motif A and B (**Figure 4.11**). Despite showing sequence identify to a plant *C*-methyltransferase, EdCMT 5 and EdCMT 6 gene products also failed to produce any *C*-methylated flavanones *in vitro*. Their activity was not tested with any sterol compounds as it is not within the scope of this study.

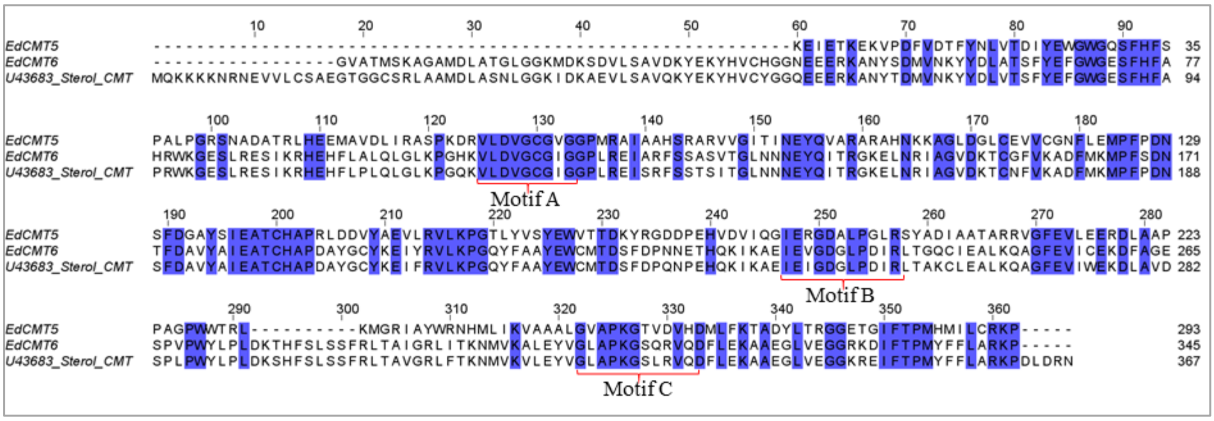


Figure 4.11: Sequence alignment of EdCMT 5 and 6 with sterol *C*-methyltransferase (accession number U43683) of *Glycine max*. Identical residues are highlighted in blue. Methyltransferases motifs suggested by Joshi and Chiang (1998) are labelled.

Choosing more putative CMT candidates was challenging due to the limited sequence information on *C*-methylation in plants. Lack of activity of all six putative CMTs with flavanones and pinocembrin chalcones raised a question of an actual substrate of *C*-methylation in *E. diversifolia*. To address this, a more straightforward approach of native CMT extraction and assay was employed. Feeding the flavanone or other possible substrates to a crude native enzyme or partially purified leaf extract would be ideal in confirming the actual substrate of native *E. diversifolia* CMT. Therefore, I attempted to extract the native enzyme from *E. diversifolia* and test for the activity of the crude extract against the flavanone compounds. For initial studies, fresh cell-free leaf extract passed through a desalting column, to remove all small molecule contaminants like secondary metabolites, was used as a crude enzyme extract. Extract of *O-methylated species* leaf was used as a positive control and HPLC fractionation of the reaction content of pinocembrin incubation with cell-free extract of *O-methylated species* leaf showed an identifiable additional peak. The product was not present in assays with boiled protein extracts. No product was detected with *C-methylated species* extract fed with flavanones. Due to the difficulties in lab access and sample collection from the Dookie campus, this experiment was interrupted, and I was unable to test the activity of the native enzyme with upstream substrates like cinnamate, cinnamoyl-CoA or phenylalanine.

4.3.4 Identification of additional putative genes involved in *C*-methyl flavanone biosynthesis

Further analysis of the transcriptome sequence data was performed to compensate for the discontinued lab-based experiments due to Covid restrictions disallowing laboratory access. More than 600 unigenes were annotated as methyltransferases in the *E. diversifolia* transcriptome annotated using the NCBI-Nr database. After excluding unigenes annotated as histone, DNA or RNA methyltransferases, eighty-eight unigenes were considered for further analysis. Predicted translation sequences of only seventy-seven unigenes were used for phylogenetic analysis as the remaining sequences were partial. The evolutionary relationship of these unigenes with characterised methyltransferases from various plant genera catalysing *O*-, Carboxy, *C*-, *N*-, and *S*-methylation was inferred using PHYML software (Guindon et al., 2010). A maximum-likelihood tree (JTT matrix model; 1000 bootstrap replicates), constructed with malonyl-ACP OMT from *E. coli* as an outgroup, showed association of twenty-six clusters with Pl-OMTII class of methyltransferases. Fourteen genes were grouped with Pl-OMTII class containing CCoAOMTs and thirty genes were categorized as closely related to the SABATH superfamily which majorly includes carboxy-MTs and some alkaloid NMTs (D Auria et al., 2003; **Figure 4.12**). Seven unigenes clustered with plant sterol C-methyltransferase. Cluster-13398.0 was grouped with the isolated pinocembrin 7-OMTs that are described in Chapter 1 and 2.

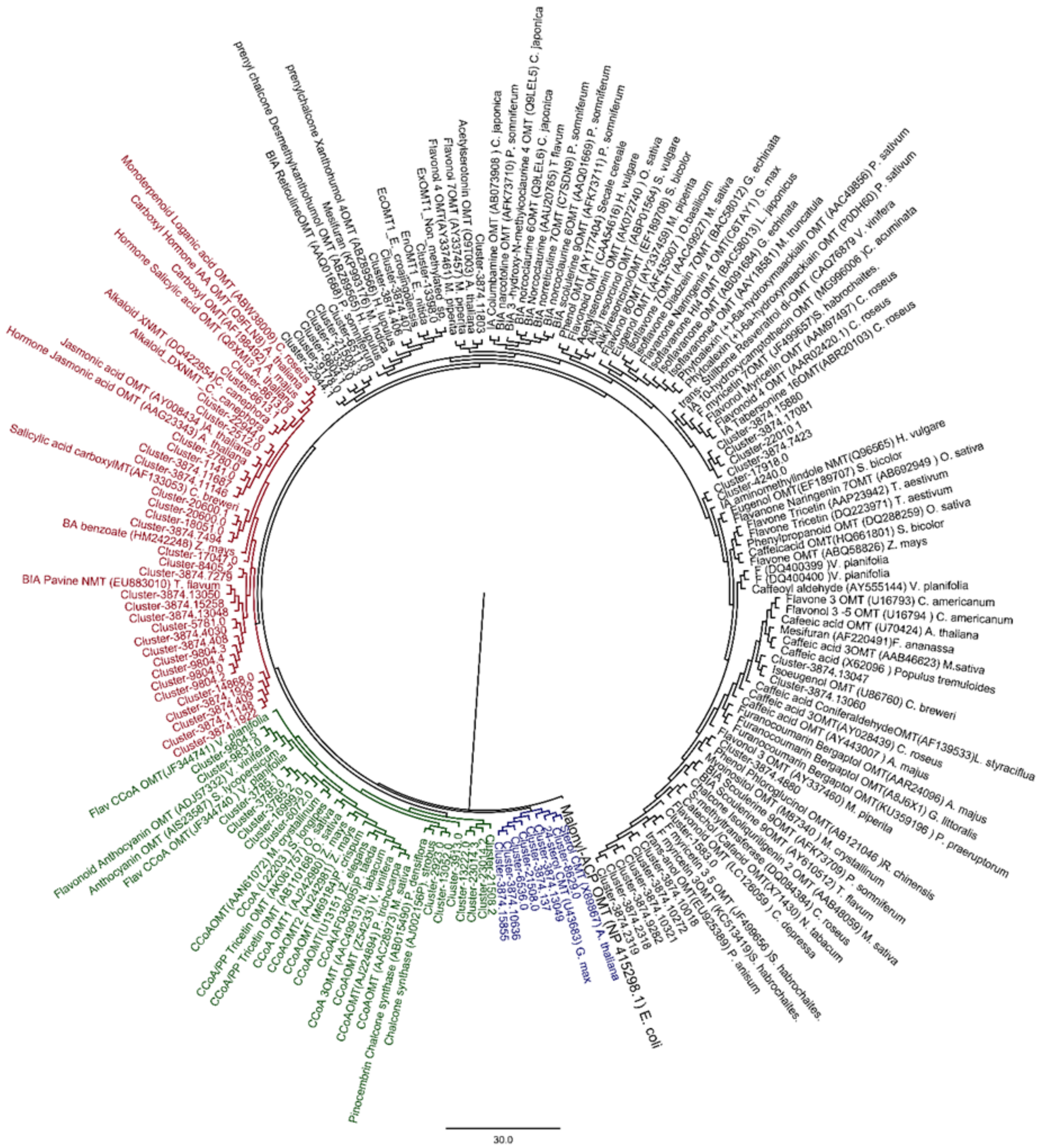


Figure 4.12: Interspecific phylogenetic analysis of putative methyltransferases genes from *E. diversifolia* with characterized plant methyltransferases catalyzing various phenolic substrates. The figure shows grouping into PL-OMTII class (caffeic acid OMT like compounds; black), Pl-OMTI class (CCoAOMTs like compounds; green), SABATH family (red). Unigenes that clustered with sterol CMTs are highlighted in blue. The accession number of the methyltransferases are listed in brackets. Branch lengths reflect the degree of divergence of each sequence. The scale bar at the bottom indicates the genetic distance. Detailed list of functionally characterised MTs used in this analysis is given in Appendix A and unigene clusters are listed in Appendix C.

Detailed gene function annotation analysis was done to identify more putative CMTs. As mentioned earlier, KEGG pathways are divided into five main functional categories: cellular processes, environmental information processing, genetic information processing, metabolism, and organismal systems. Each of these main categories is subdivided into several groups (see

Figure 4.8 for category names). Under metabolism, the subcategory, biosynthesis of other secondary metabolites, is grouped into ten pathways, including phenylpropanoid, flavonoid, isoflavonoid, anthocyanin, flavone and flavonol biosynthesis (**Figure 4.13**A). A total of seventy-three unigenes were categorised to specific flavonoid pathways including isoflavonoid biosynthesis and anthocyanin biosynthesis (**Figure 4.13**A); however, out of these seventy-three unigenes only one was annotated as putative methyltransferase (**Figure 4.13**B). A total of 160 unigenes were putatively categorised as phenylpropanoid biosynthetic genes and specifically included eighteen putative methyltransferases predicted as caffeoyl-CoA *O*-methyltransferases (**Figure 4.14**).

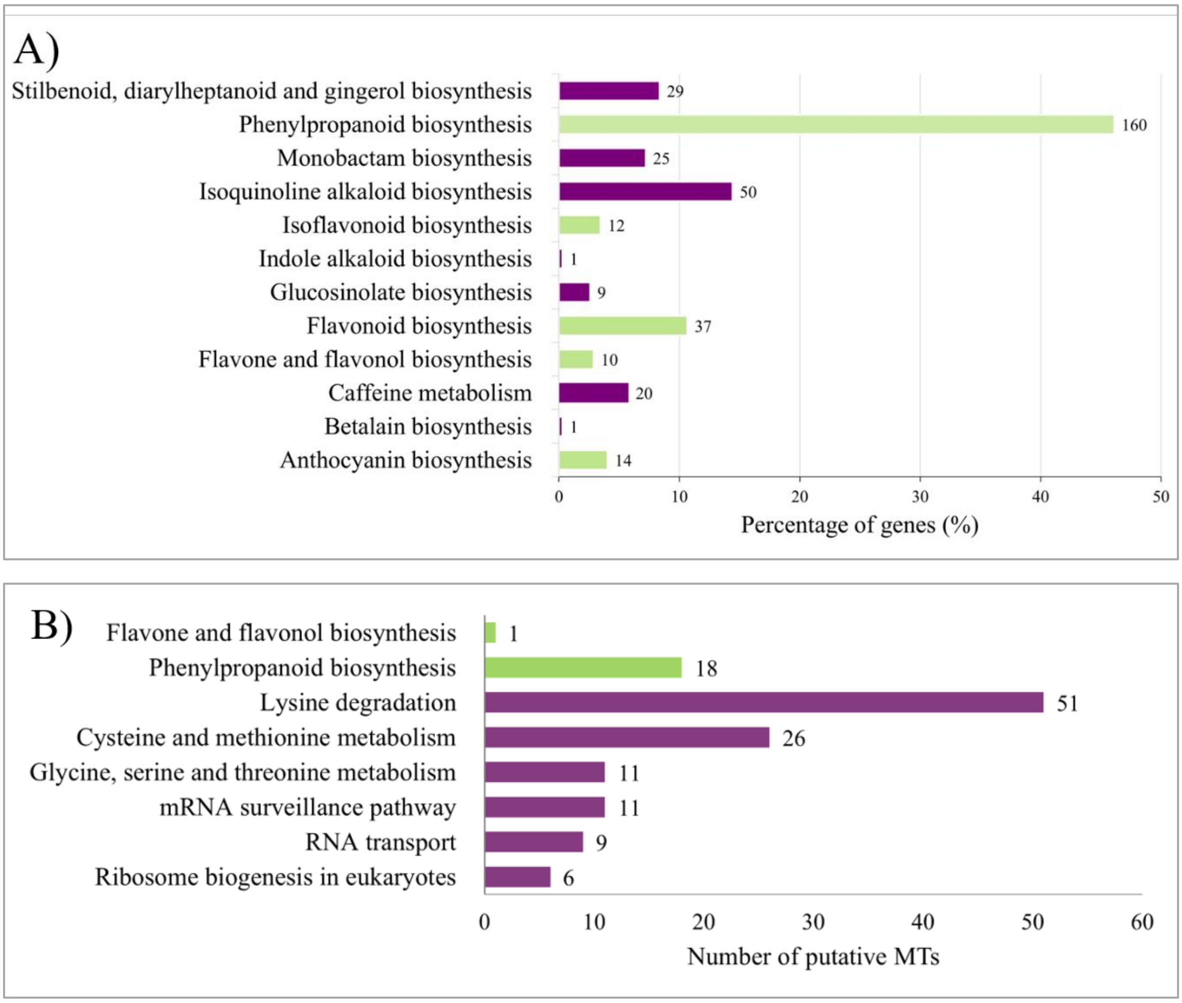


Figure 4.13: KEGG pathway subcategorization within the "metabolism" category. A) Grouping of unigenes that fell into the "biosynthesis of other secondary metabolites" sub-category into various pathways. Phenylpropanoid and flavonoid biosynthesis pathway groups are depicted in green bars. B) Number of unigenes annotated as MTs in KEGG main category of "metabolism". Eighteen unigenes were predicted as MTs involved in phenylpropanoid biosynthesis and one was predicted to be a flavonoid MTs.

Based on the KEGG pathway analysis, several flavanone biosynthesis related genes were identified. These include malonyl-CoA decarboxylase (EC 4.1. 1.94; two unigenes), acetyl-CoA carboxylase (EC 6.4.1.2; fourteen unigenes), phenylalanine ammonia lyase (EC: 4.3.1.24; six unigenes), 4-coumarate-CoA ligase (EC: 6.2.1.12; nine unigenes), chalcone synthase (EC: 2.3.1.74; five unigenes), chalcone isomerase (EC:5.5.1.61; five unigenes). The genes identified downstream of pinocembrin were flavanone 3-hydroxylase (EC:1.14.11.9; 2 unigenes), flavone synthase (EC:1.14.20.6; 2 unigenes), bifunctional dihydroflavonol 4-reductase/flavanone 4-reductase (EC:1.1.1.219 1.1.1.234; 1 unigene), flavonol synthase (EC: 1.14.20.6; 2 unigenes), flavonoid 3', 5'-hydroxylase (EC: 1.14.14.81; 1 unigene), flavonoid 3'-monooxygenase (EC:1.14.14.82; 5 unigene), flavonol 3-*O*-glucosyltransferase (EC:2.4.1.91; 1 unigene), anthocyanidin synthase (EC:1.14.20.4; 1 unigene) and several other anthocyanin biosynthetic genes (**Figure 4.15**). The only unigene annotated as a putative flavonoid *O*-methyltransferase was Cluster-9831.0 in the flavone and flavonol biosynthetic pathway (KO00944). This unigene is orthologous to the *Catharanthus roseus* flavonoid 3',5'-methyltransferase [EC:2.1.1.267], which catalyses myricetin methylation. Based on the KEGG ontology, a total of 133 unigenes were annotated as methyltransferases. Specifically, 26 unigenes were annotated as methyltransferases to translation subcategory under the main functional category of genetic information processing

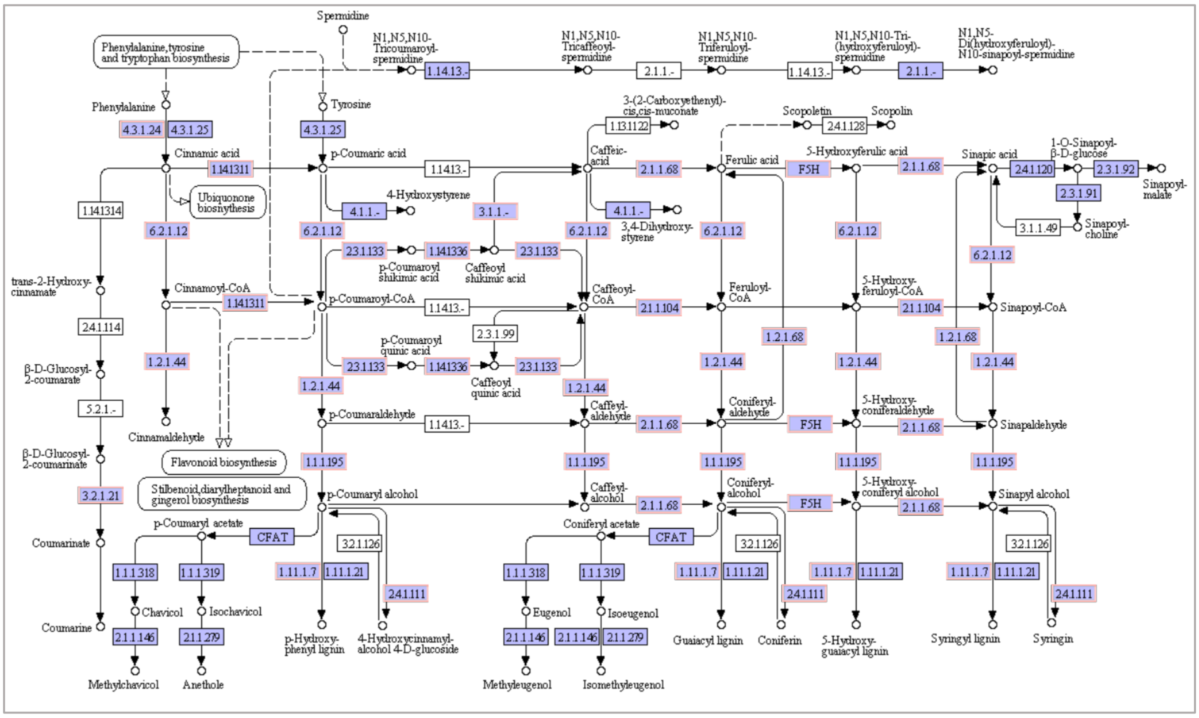


Figure 4.14: KEGG pathway of phenylpropanoid biosynthesis with enzymes labelled with EC number. Those annotated to *E. diversifolia* unigenes are highlighted with a pink outline. The general pathway can be accessed via https://www.genome.jp/kegg-bin/show_pathway?map00940.

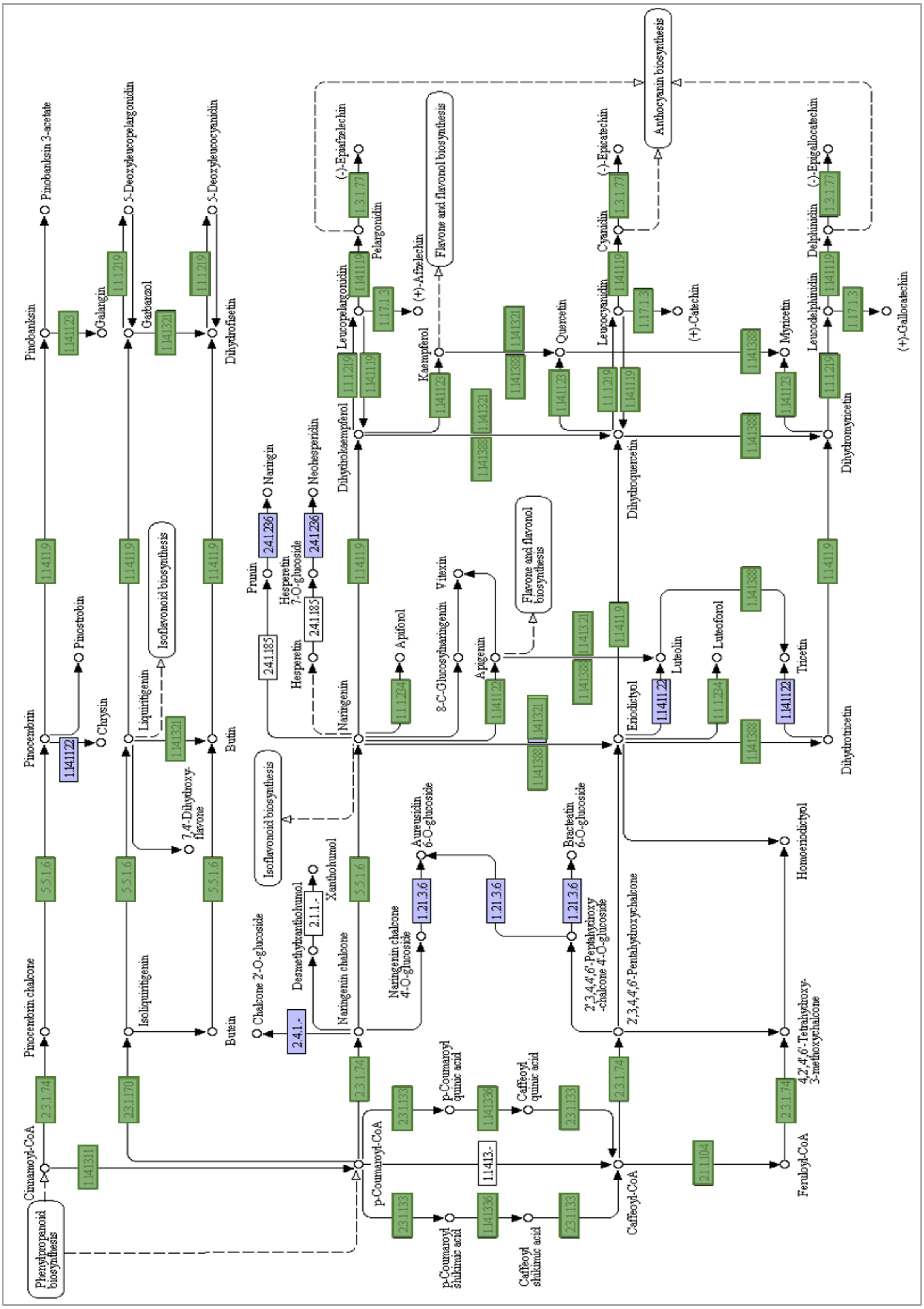


Figure 4.15: KEGG analysis of flavonoid biosynthesis pathway genes in *E. diversifolia* leaf tip transcriptome. Transcripts identified in the pathway are labelled in green colour. All enzymes are labelled with Enzyme Commission number according to the IUBMB nomenclature.

Table 4.4: Unigene clusters putatively related to flavonoid *C*-methylation

Enzyme	**Associated putative unigenes***
Chalcone synthase	Cluster-3874.8829, Cluster-3874.8832, Cluster-3874.9670 Cluster-3874.8754, Cluster-3874.8755
Chalcone isomerase	Cluster-3874.9533
4-coumarate-CoA ligase	Cluster-5607.0, Cluster-3874.7167 Cluster-3874.16413, Cluster-12908.0, Cluster-2293.0, Cluster-21773.0, Cluster-3874.13993, Cluster-3417.0, Cluster-11665.0
Phenylalanine ammonia lyase	Cluster-14799, Cluster-3874.6310, Cluster-20740.0, Cluster-14799.0, Cluster-14799.4, Cluster-3874.13844
Acetyl-CoA carboxylase	Cluster-23433.0, Cluster-3874.4217, Cluster-3874.4216, Cluster-16650.0, Cluster-16650.1, Cluster-3874.6671 Cluster-1651.0, Cluster-3874.12825, Cluster-7597.0, Cluster-3874.5586, Cluster-3874.12823, Cluster-3874.12824 Cluster-3874.11770, Cluster-12795.0
Malonyl-CoA decarboxylase	Cluster-16272.0, Cluster-7420.0
Malonyl-CoA/methylmalonyl-CoA synthetase	Cluster-3874.7344, Cluster-3874.10805

*Genes downstream of pinocembrin involved in flavone, flavonol and anthocyanin synthesis, are not included.

To assess the possibility of an alternate biosynthetic route for *C*-methylation in *Eucalyptus*, enzymes that may catalyze reactions upstream of flavanone formation like chalcone synthase (CHS) and methylmalonyl-CoA biosynthetic enzymes were assessed in detail. Five unigene clusters were annotated as chalcone synthase (CHS; **Table 4.4**). Two of them (Cluster-3874.8754, Cluster-3874.8829) were partial sequences, therefore they were discarded. Predicted translation sequences of the remaining clusters were named EdCHS 1 (Cluster-3874.8755), EdCHS 2 (Cluster-3874.8832), and EdCHS 3 (Cluster-3874.9670). EdCHS 1 has 94% sequence identity to EdCHS 2, but only 76% to EdCHS 3. All three putative CHS unigenes were compared to the chalcone synthase variant PStrCHS2 that catalyze the formation of methylated chalcones in *Pinus strobus*. EdCHS 1 and 2 have 79% sequence identity to PStrCHS2, whereas for EdCHS 3 it is 66 %.

Ten unigene clusters were found to likely encode methylmalonate semialdehyde dehydrogenases, which are a part of propanoate metabolism. Two clusters were annotated as malonyl-CoA/methylmalonyl-CoA synthetase. Further in silico analyses like structure prediction and modelling were not attempted as there is no plant CMT enzymes whose structure has been studied in detail. Chalcone synthase predicted clusters could have been modelled to see if they are similar to *Pinus strobus* CHS, but similarly there are no structural studies on the latter.

4.3.5 Identification and sequence analysis of EnOMT1 homologue in *E. diversifolia*

Phylogenetic analysis of putative methyltransferases from *E. diversifolia* grouped a unigene cluster together with the 7-OMTs identified in Chapter 2. *E. diversifolia* reported in this study lack *O*-methylated flavanones, therefore it was interesting to see the characteristics of homologous gene of *EnOMT1* in *E. diversifolia*. To investigate the sequence similarities, EnOMT1 was 'Blasted' against the *E. diversifolia* unigene database, which resulted in the identification of an EnOMT1 homologous gene (Cluster-13398.0). The sequence was partial and corresponded to the C-terminal end of the *O*-methyltransferase enzyme. Using the primers used for EnOMT1 isolation, the full sequence of the homologous gene, EdOMT1, was retrieved. The C-terminal sequence alignment of EdOMT1 with EnOMT1 and ExOMT1 is shown in (**Figure 4.16**). Surprisingly, the C-terminal amino acid changes in EdOMT1 were identical to the *non-methylated species* enzyme ExOMT1. Relative expression levels were quantified using qPCR and showed a lower level of expression compared to *EnOMT1* and similar expression level to that of *ExOMT1* from *non-methylated species* (Appendix E).

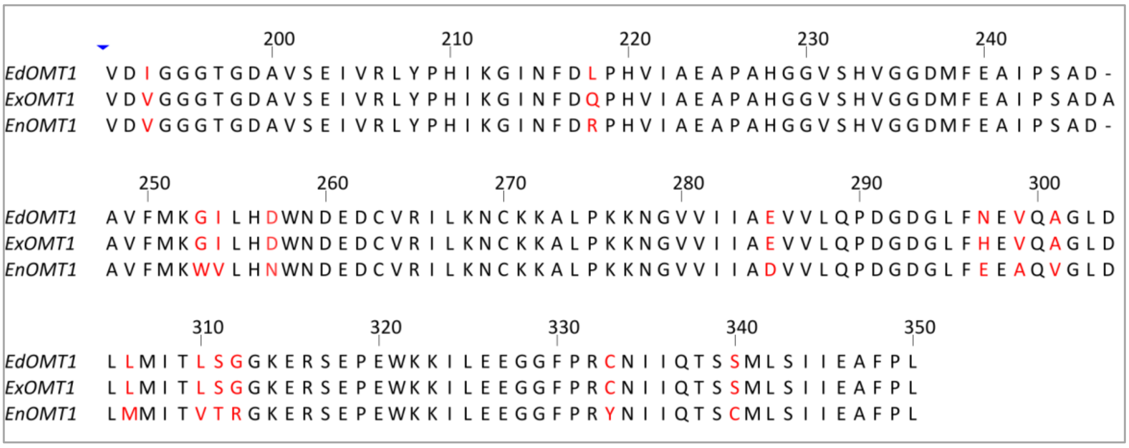


Figure 4.16: C-terminal amino acid sequences alignment of EnOMT1 (*O-methylated species*) with orthologues from *E. diversifolia* (EdOMT1) and *non-methylated species* (ExOMT1). All mismatches are highlighted in red.

4.3.6 Localisation of methylated tri-ketone, conglomerone, to foliar oil glands of *E. agglomerata*

While doing foliar metabolic analysis of *E. subgenus Eucalyptus* species for an ideal *C*-methyl flavanone rich candidate, some interesting and tangential observations were made that are worth reporting here. As mentioned earlier in section 4.3.1, an unidentified compound at RT 11.6 min dominated in acetonitrile leaf extracts of all *E. agglomerata* individual trees. To examine the location of the detected metabolites, secretory glands from *E. agglomerata* (three individual plants that showed quantitatively different profiles) were isolated and analysed. No visual differences were observed in the isolated oil glands examined by light microscopy (**Figure 4.17**-A). Acetonitrile extract of glands analysed using RP-HPLC showed the peak for the unidentified compound confirming it is at least partially localised in glands (**Figure 4.18-A)**. Hexane extracts of glands subjected to GC-FID followed by GC-MS showed that the unknown compound is a volatile β-triketone conglomerone with a characteristic fragment of *m/z* 195 (**Figure 4.18**-B; Adams, 2017). Identity was further confirmed by ESI-LCMS/MS with MS2 characteristic [M+H]+ ion fragments *m/z* (% intensity): 206(100), 191(23), 207(20), 151(18) and 221(15) as well as retention time comparison with conglomerone from *E. brevistylis* extract (**Figure 4.18**-A; Senaratne, 2020).

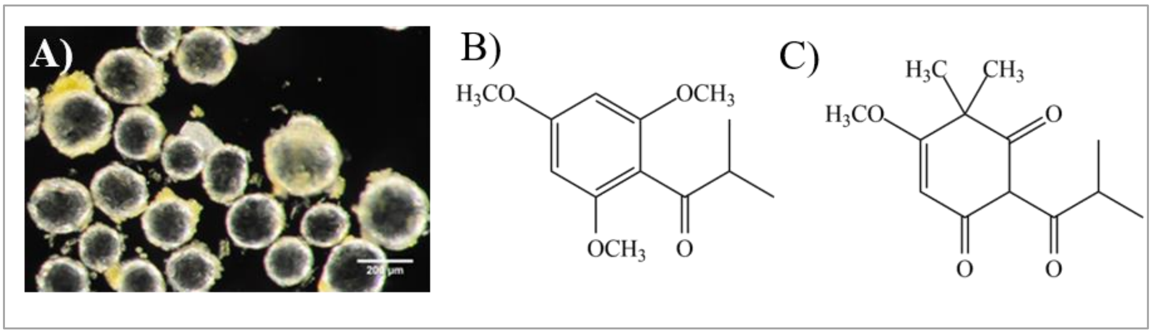


Figure 4.17: Structures of volatile triketone compounds localized to the foliar secretory cavities of *E. agglomerata.* A) Isolated foliar glands *of E. agglomerata;* B) Structure of conglomerone; C) Structure of agglomerone.

GC-MS analysis of the *E. diversifolia* extract also showed the presence of several triketones, of which the most abundant was agglomerone (**Figure 4.18**C, D). Putative triketones with molecular formulae $C_{11}H_{14}O_4$ (Trik-1), $C_{12}H_{16}O_4$ (Trik-2), and $C_{13}H_{18}O_4$ were also observed (**Figure 4.18**D).

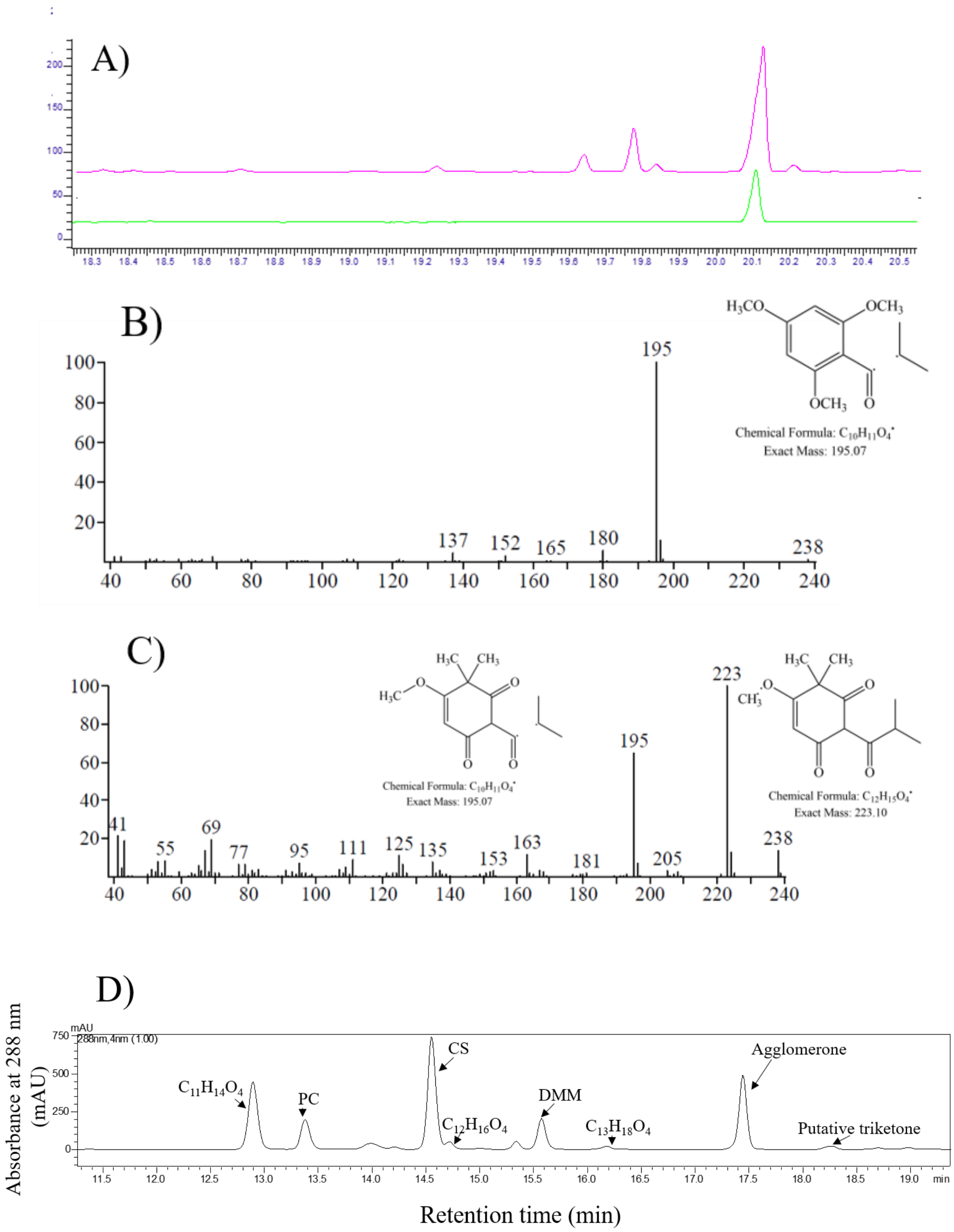


Figure 4.18: Structural elucidation of β-triketones from *Eucalyptus*: A) GC-FID overlayed chromatogram of *E. agglomerata* gland extract (pink) and conglomerone from *E. brevistylis* (green) showing the abundant compound as conglomerone. B) Mass spectra of conglomerone from GC-MS analysis. C) Mass spectra of agglomerone from GC-MS analysis. D) HPLC-PDA chromatogram of *E. diversifolia* extract showing flavanones and triketones. PC- pinocembrin; CS- cryptostrobin; DMM- desmethoxymatteucinol.

4.3.7 Intra-specific variation in the metabolic profile of *Eucalyptus*

The occurrence and composition of metabolites in foliar acetonitrile extracts were analysed in detail. Representative species belonging to *subgenus Eucalyptus* used in this study are listed in **Table 4.2** and their flavanone profile is given in **Figure 4.2**. Metabolic profile of the *Eucalyptus* species that showed a presence of triketones is given in **Figure 4.19**-A. Based on the presence or absence of the identified triketones and flavanones in the individual plants, *E. diversifolia* and *E. agglomerata* were grouped as two chemotypes (**Figure 4.19**-A). One of the *E. diversifolia* chemotypes contained the triketone agglomerone and Trik-1 in abundance (*E. diversifolia-I;* 6 individual plants) while the other did not (*E. diversifolia-II;* 6 individual plants). The total mean concentration of *C*-methyl flavanones in *E. diversifolia-I* was higher than that of the other. In the case of *E. agglomerata*, out of the 15 plants analysed, one chemotype (*E. agglomerata-I*; eight individuals) exclusively accumulated the triketone conglomerone whereas *E. agglomerata-II* (six individuals) showed the presence of *O*-methylated flavanones (16%) along with conglomerone (65%) and Trik-1(19%). One specific *E. agglomerata* plant collected from the University of Melbourne, Parkville campus showed different complements of triketones along with *O*-methylated flavanones, where a decrease in conglomerone (25%) and Trik-1 (2%) concentration was balanced with higher Trik-2 (51%) and the flavanone PS (Plant 3; **Figure 4.19**B).

A significant quantitative variation in flavanones was observed between individual trees within a species. For example, in *E. diversifolia*, cryptostrobin concentration varied from 4-15 mg g^{-1}_{DW} (**Figure 4.21**A), and pinocembrin concentration in *E. haemastoma* varied from 3- 28 mg g^{-1}_{DW}. Large variations were observed in *E. haemastoma* individuals where one plant had 12 mg g^{-1}_{DW} dimethylpinocembrin and some others accumulated only pinostrobin (4 mg g^{-1}_{DW}) (**Figure 4.21**B). These were not categorised as chemotypes as the individuals with these changes were less than 25% of the total plants analysed. Several unidentified peaks (retention time; 13.8 min, 14.9 min, 15.6 min) dominated the *E. stellulata* profile, identification of the compounds corresponding to these peaks was not done in this study (**Figure 4.20**). Trace amounts of 5-*O*-methylated alpinetin were found in *E. stellulata*, *E. haemastoma* and several individuals of *E. agglomerata*. Other than in *E. diversifolia*, the *C*-methylated flavanones cryptostrobin and DMM were found in *E. preissiana and E. haemastoma*, however in very low concentration (0.9- 1.5 mg g^{-1}_{DW}).

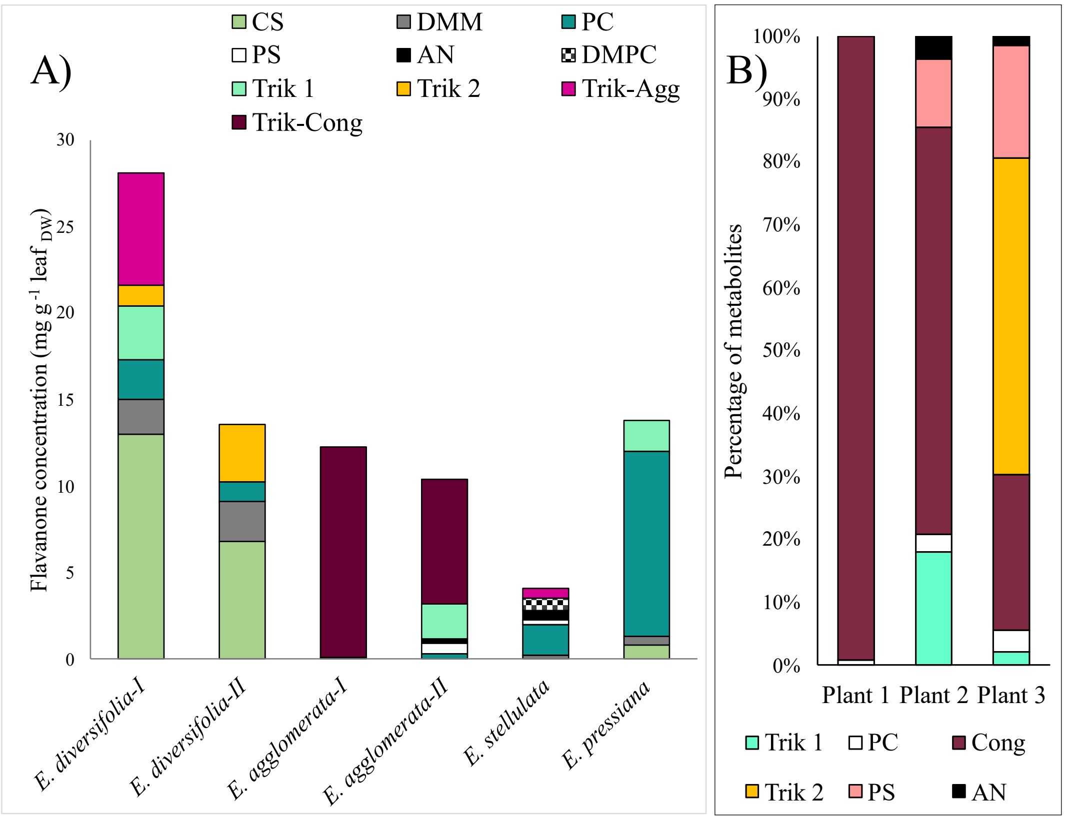


Figure 4.19:Foliar flavanones and β-triketones in candidate species. (A) Metabolic profile of species that showed triketones. Chemotypes are denoted as species name followed by I and II. The number of plants used for the data calculation is as follows: *E. diversifolia* chemotypes (six individual plants); *E. agglomerata-I* (eight plants*); E. agglomerata-II* (six plants). Concentrations of CS, DMM and β-triketones are in pinocembrin equivalents. CS- cryptostrobin; DMM- desmethoxymatteucinol; PC- pinocembrin; PS- pinostrobin; AN-alpinetin; DMPC- dimethylpinocembrin; Trik- triketone; Agg- agglomerone; Cong- conglomerone. B) Foliar gland contents of three individual *E. agglomerata* plants. Plant 1 represent *E. agglomerata-I* chemotype, plant 2 represent *E. agglomerata-II* and plant 3 represent the profile of a single *E. agglomerata* plant that showed triketone2 and PS.

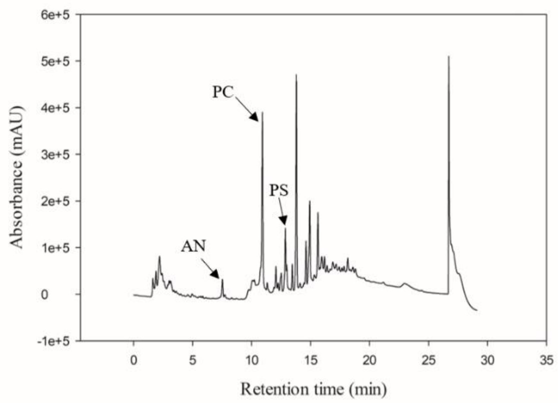


Figure 4.20: Representative RP-HPLC chromatogram of an individual *E. stellulata* leaf extract showing flavanones. AN- alpinetin; PC- pinocembrin; PS- pinostrobin; and peaks corresponding to unknown compounds with retention time later than PS.

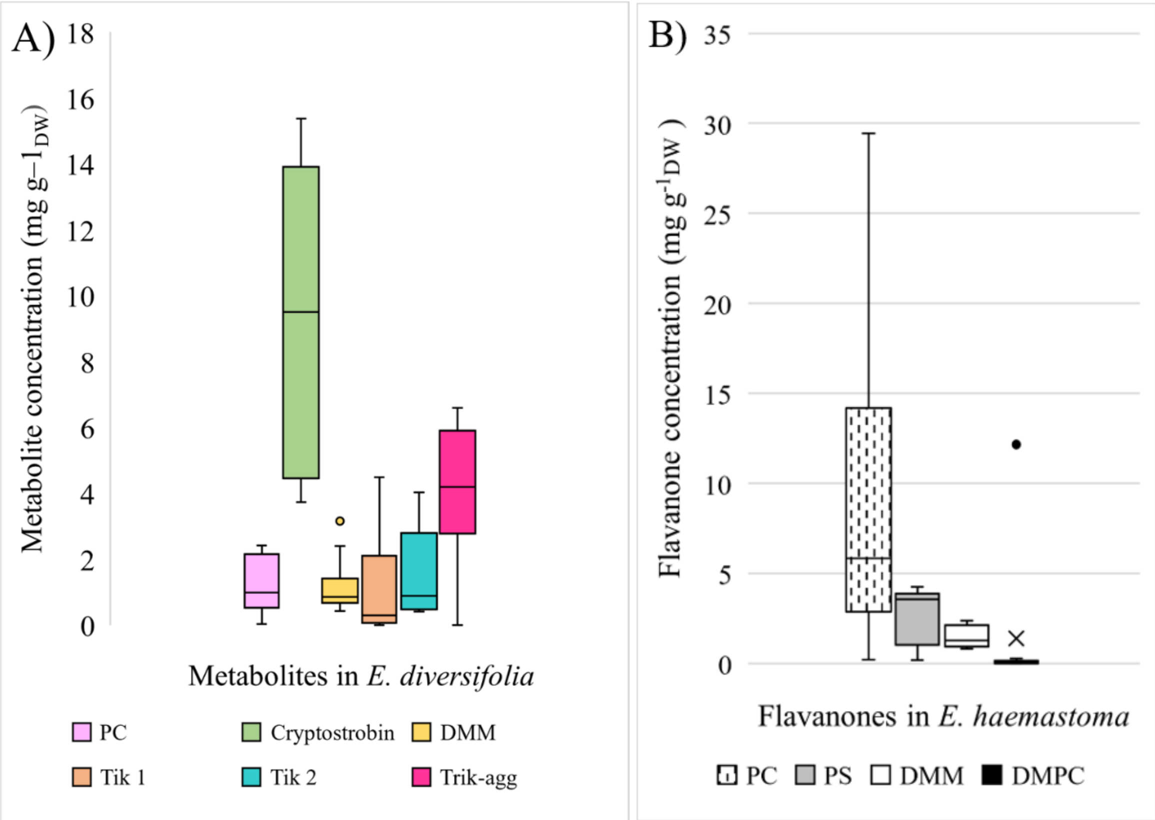


Figure 4.21: Variation in foliar metabolite concentration between individuals of in *E diversifolia* (A) and *E. haemastoma* (B). The data are of 12 individual trees.

4.3.8 Ecological aspects of methylation

Plants employ different strategies to counter the detrimental effects of the external stressful environment. Effects of long term UV-A (320–400 nm) light exposure on the flavanone profile of several *Eucalyptus* species, namely, *Non-methylated, O-methylated and C-methylated* species, were examined to identify whether flavanones are involved in the photoprotection strategy of *Eucalyptus* plants. The UV-A and control chambers were successfully set up inside the laboratory and chamber conditions were monitored continuously using various sensors. The temperature, percentage relative humidity and light intensity were found to be relatively steady over time. The initial study was focused on the assessment of total flavanone contents in the UV-A treated leaves of *Non-methylated species* seedlings. Flavanone content in ten individual control and UV-A treated plants leaf samples collected after four weeks of UV-A treatment, showed no increase in flavanones. A longer UV-A exposure (16 weeks) also did not affect flavanone concentration. The UV-A treated plants showed a decreased pinocembrin concentration per leaf dry weight compared to the UV-untreated control plants (**Figure 4.22**); however, no methylated flavanones were detected in any of the UV treated plants. The *Non-*

methylated species plants grown in the chamber were stunted in growth. Fully developed leaves were small compared to a glasshouse-grown or field grown plant. As the plants were not showing growth characteristics relatable to that of the field-grown plants, the reliability of this study was in doubt and thus the follow-up investigation to replicate the same study on leaf samples of *O- and C-methylated* plants were cancelled. Instead, I decided to continue UV exposure analysis on callus/ cell suspension cultures of *Eucalyptus*.

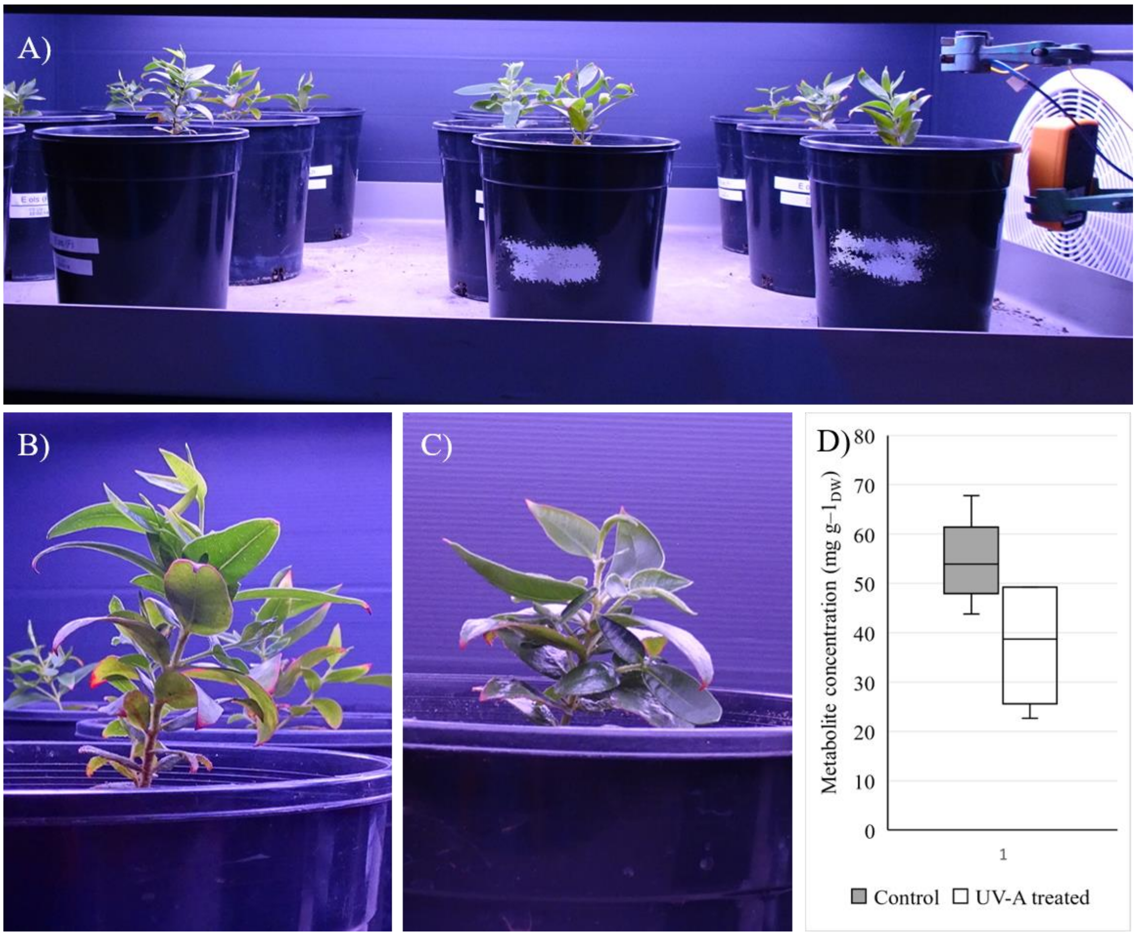


Figure 4.22: UV exposure analysis on *Eucalyptus* seedlings. A) Growth chamber set up with seedlings; B) Actively growing control plants; C) UV-A treated plant showing stunted growth. D) The average amount of pinocembrin in ten individual control and UV-A treated plants. There is no statistically significant difference between the control and UV-A treated group (P=0.115)

Cell suspension cultures of *E. croajingolensis* and *E. nitida* were successfully established, using leaf and internodal explants from tissue culture-germinated seedlings, following the organogenesis protocol developed for regeneration of *Eucalyptus polybractea* (Fernando et al., 2016). Callusing started along the cut ends of leaf explants and gradually the leaves curled up and the whole leaf explant was covered with callus (**Figure 4.23**B). In the case of internodes, callusing initiated at the cut end, and callus production was at a higher rate compared to the leaf explants (**Figure 4.23**C). *E. croajingolensis* initial callus was lightly compact, pale yellow with a red tinge on few areas (**Figure 4.23**D). The callus became more friable with subsequent subculturing. On the other hand, *E. nitida* initial callus was pale yellow, loose and more friable than the *E. croajingolensis* callus (**Figure 4.23**E). The rapidly growing areas of callus were friable compared to the non-growing or slow-growing area. Callus initiation from older leaf explants resulted in bright red colour which eventually faded away with multiple subculture (**Figure 4.23**F).

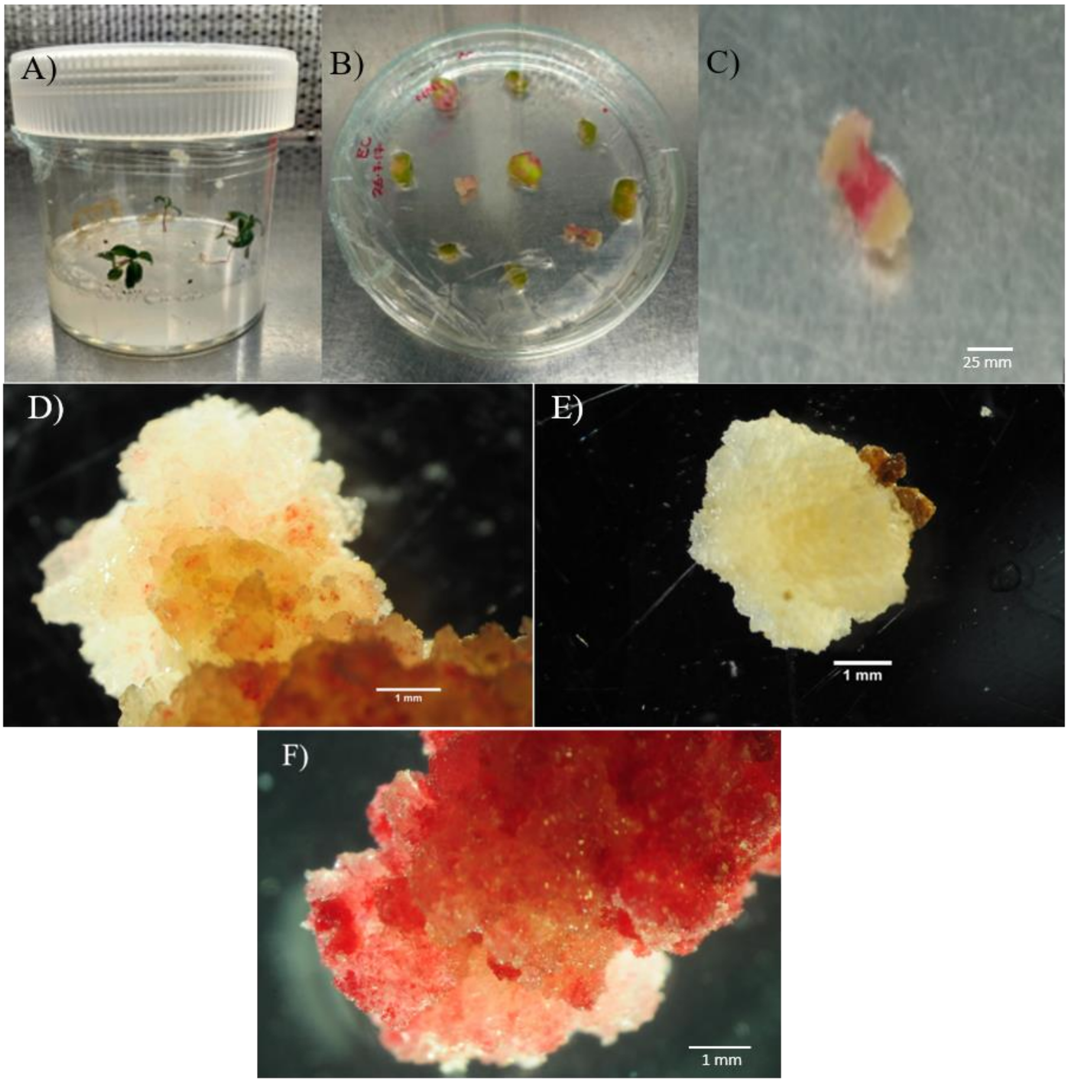


Figure 4.23: Callogenesis from various explants of *Eucalyptus*. (A) *In vitro* germinated *Eucalyptus* seedlings. (B) Callus initiation from explants. (C) Internode with callus initiated from cut ends. (D) *E. croajingolensis* callus from internodal explant. (E) *E. nitida* callus from cotyledonary leaf explant (F) *E. croajingolensis* callus from mature leaf explant.

Cell suspension culture of *Eucalyptus croajingolensis* was initiated first. Cells in the suspension culture formed clumps initially. These were removed during each subculture by sieving and after 4 months of initiation, the final culture looked brownish yellow with fewer cell clumps (**Figure 4.25**A).

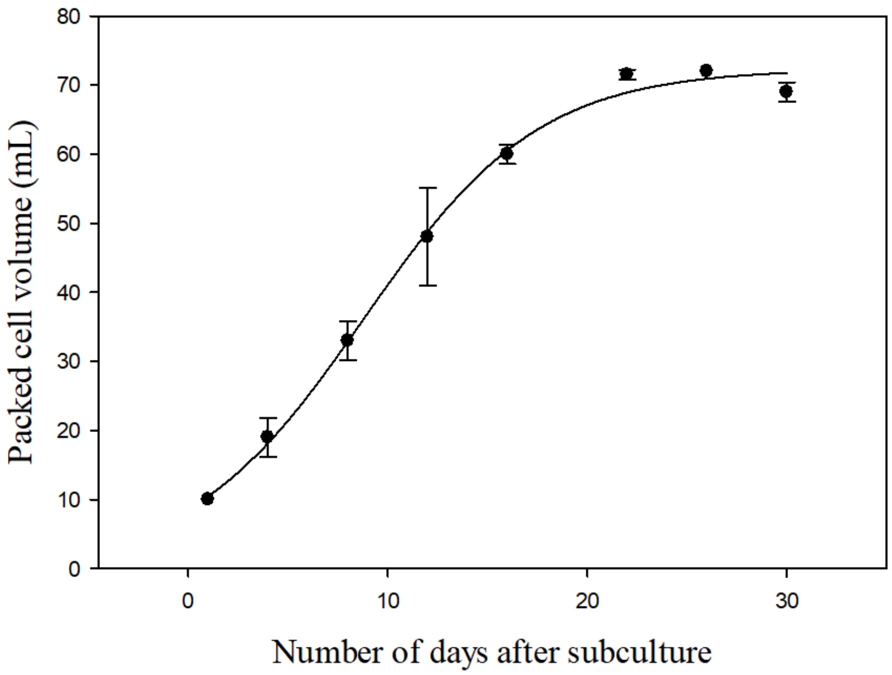


Figure 4.24. *E. croajingolensis* cell growth curve in suspension culture.

The growth pattern of cells in suspension culture showed a sigmoidal curve. Cells had a lag phase of four days followed by a linear increase and after twenty days the growth ceased and possibly declined (**Figure 4.24**). Depletion of nutrients, accumulation of toxic and other limiting factors might have led to a decline in growth. From the growth pattern of cells in suspension, it was inferred that the subculturing should be done between 15-22 days of the culture growth. Also, precursor feeding can be done during the log phase of cell growth which is a week after the day of subculture.

Unlike *E. croajingolensis*, *E. nitida* cell suspension cultures did not form any cell aggregates because the callus produced by *E. nitida* cotyledon leaves was friable. The final cell suspension had the appearance of a cloudy yellow solution. A single drop of cell suspension was observed under a stereomicroscope and showed that cells were not attached, and cells looked longer unlike typical plant cells (**Figure 4.25**-B). Flavanone content analysis of phytochemicals produced by the cells in suspension showed no presence of pinocembrin or any methylated flavanones (**Figure 4.26**).

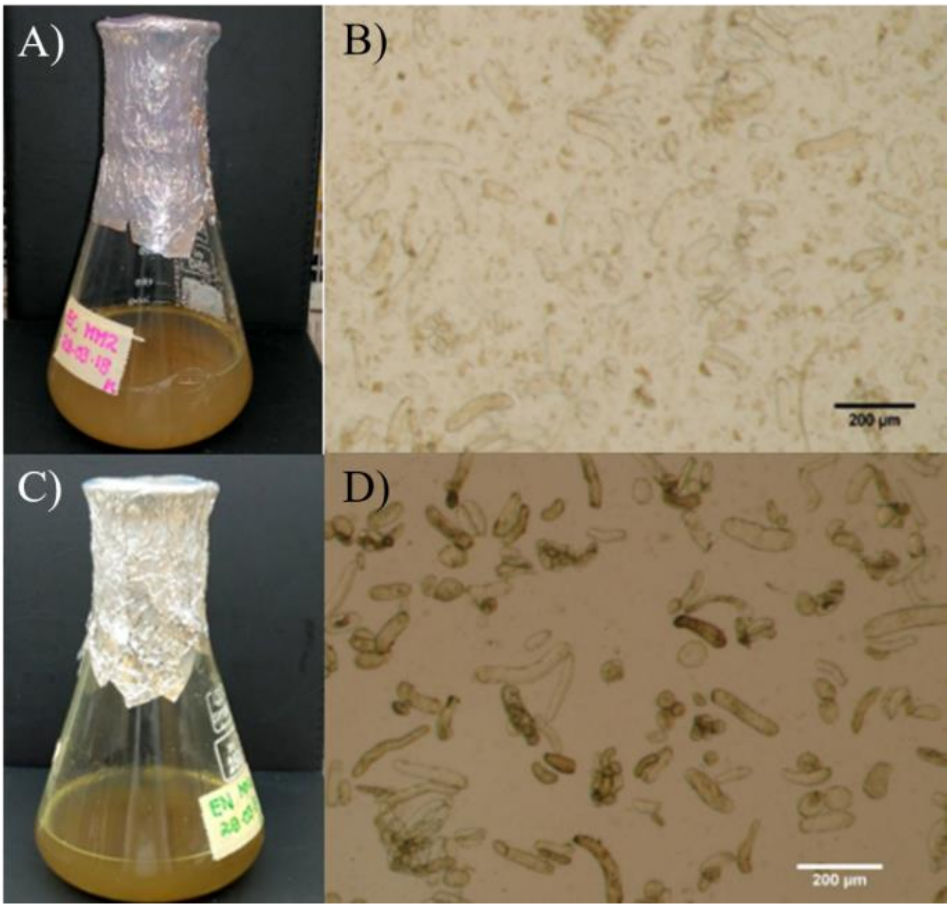


Figure 4.25. Cell suspension cultures of *Eucalyptus* showing elongated cells under light microscope. A) *E. croajingolensis* cell suspension culture. B) *E. croajingolensis* cell suspension under light microscope. C) *E. nitida* cell suspension culture. D) cell suspension under light microscope *E. nitida* cell suspension culture

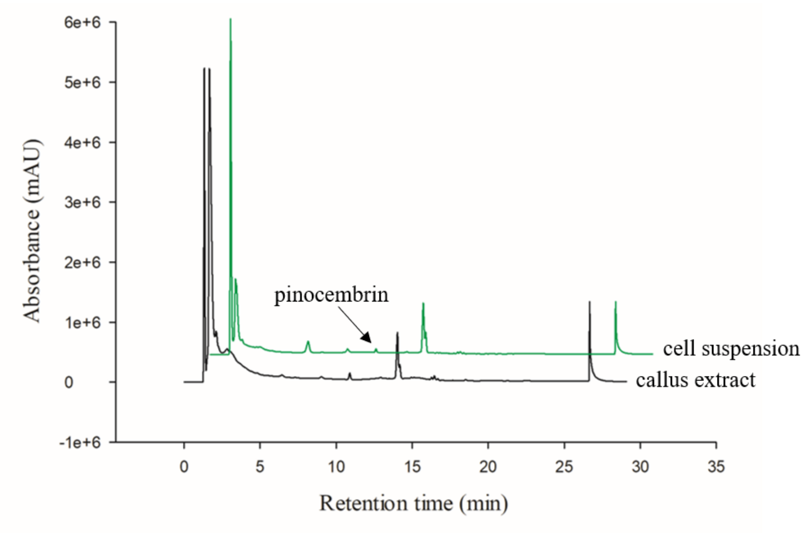


Figure 4.26: RP-HPLC chromatogram of extracts of *E.croagingolensis* callus (black) and cells of suspension culture (green).

To determine if the age of the explant had any effect on the flavanone production, I re-initiated new callus culture from fully grown leaf explants. Nevertheless, the cells did not produce any flavanones. Initial precursor feeding studies failed as the precursor pinocembrin precipitated out in the culture media and was unavailable for the cells to absorb. For ease of handling, instead of cell suspension, callus cultures of *O*- and *C*-methylated species were subjected to UV-A treatment in a closed chamber. No detectable change was observed as the flavanone content in the original callus was also trace (**Figure 4.26**). Later due to the COVID related university closure and limited access to the lab, maintenance of multiple calli and cell suspension cultures were not possible, and further experiments were not continued.

4.4 Discussion

4.4.1 *C*-methylated flavanones in *Eucalyptus*

The initial objective of this study was to look for a species that contains high amounts of *C*-methylated flavanones. Species selection was primarily based on the reported occurrence of flavanones in the subgenus *Eucalyptus* (Senaratne, 2020). Screening of individuals of subgenus *Eucalyptus* revealed that *E. diversifolia* contains two *C*-methylated flavanones, cryptostrobin (CS) and desmethoxymatteucinol (DMM). Recently, both CS and DMM along with strobopinin were reported to be confined to species belonging to *Amentum* and *Capillulus* sections of the subgenus *Eucalyptus* (Marsh et al., 2019). Nonetheless, the results of this study along with a recent publication from our lab (Goodger et al., 2019), shows a species from *Longistylus* section (*E. diversifolia*) harbouring *C*-methylated flavanones in relatively high concentration. The level of desmethoxymatteucinol observed in *E. diversifolia* in this study is comparable to those observed by Marsh et al. (2019) and Goodger et al. (2019) whereas the cryptostrobin level observed in this study is much higher. Other than in *Eucalyptus*, these two compounds have previously been reported in other genera of the family *Myrtaceae*. For example, *Myrceugenia euosma* dried branches and leaves contained both cryptostrobin and desmethoxymatteucinol (Oikawa et al., 2016); cryptostrobin along with various *C*-methylated unsubstituted B-ring flavanones were detected in *Agonis flexuosa* (Labib et al., 2020); cryptostrobin along with pinostrobin were detected in *Eugenia mattosii* leaf extract and *Corymbia torelliana* fruit resin (Massaro et al., 2014; Vechi et al., 2019). The 6-*C*-methylated flavanone strobopinin is often found along with cryptostrobin and desmethoxymatteucinol (Coutinho et al., 2008; Kuo et al., 2004; Labib et al., 2020; Marsh et al., 2019; Massaro et al., 2014); however, it was not detected in any extracts analysed in this study. While Marsh et al.

(2019) detected flavanones with both *O*- and *C*-methylations (5-hydroxy-7-methoxy-8-methylflavanone; 6-40 mg g^{-1}_{DW}) in *E. subgenus Eucalyptus* species including in *E haemastoma* and *E agglomerata*, no evidence of this compound in such quantity was detected in the samples analysed here. It must be noted that unknown peaks of lower abundance were not explored in this study, and these might be flavanone compounds containing both *O*- and *C*-methylations. Traces of *C*-methylated DMM along with *O*-methylated flavanones detected in *E. haemastoma* shows that both *C*- and *O*-methylation mechanisms are active in that species, therefore, this makes the presence of flavanones with both *O*- and *C*-methylations in low concentration feasible.

4.4.2 Co-occurrence of methylated flavanones and β-triketone in Eucalyptus

Conglomerone and agglomerone are the major β-triketones identified in this study. Agglomerone is a *C*-methylated triketone whereas conglomerone is *O*-methylated. Conglomerone concentration ranged in *E. agglomerata* from 6 to 15 6 mg g^{-1}_{DW} with an average of 12 6 mg g^{-1}_{DW} whereas *E. diversifolia* had agglomerone up to 6 mg g^{-1}_{DW} along with some other triketones. Agglomerone was first isolated from *E. agglomerata* Maiden and *E. mckieana* Blakely (Hellyer, 1964) while this is the first report showing *E. agglomerata* accumulating conglomerone.

In this study β-triketone conglomerone and *O*-methylated flavanones were localised to foliar secretory glands of *E. agglomerata*. The presence of both compounds in glands observed in this study, along with the previous studies, suggests this is likely a genus wide phenomenon (Goodger et al., 2018; Goodger et al., 2016). *Leptospermum* species (family *Myrtaceae*) have also been reported to compartmentalize β-triketones and *C*-methylated flavanones in their oil glands (Killeen et al., 2015). This sequestration is likely due to their potential for autotoxicity (Dayan et al., 2011; Killeen et al., 2015; Sirikantaramas et al., 2008). Whether triketones and flavanones are housed in the same or separate glands remains unclear. Recently, conglomerone was reported to be exclusively stored in visually distinct golden-brown foliar glands of *Eucalyptus brevistylis* separated from other sesquiterpenes (Goodger et al., 2018). In this study, isolated glands of *E. agglomerata* were translucent in appearance and showed no variation in morphology. Further studies on *E. agglomerata* and *E. diversifolia* individual gland isolation and analysis can give more information on whether triketones are co-housed with flavanones or have separate glands.

This study provides the first evidence showing the co-occurrence of β-triketones with flavanones in *Eucalyptus*. Previous studies that reported the presence of β-triketones and flavanones in *Eucalyptus* demonstrated an inverse relationship in their concentration. For example, Senaratne (2020) analysed gland contents of thirty *Eucalyptus* species (17 belong to subgenus *Symphyomyrtus*; 13 belong to subgenus *Eucalyptus*) and found that those accumulating β-triketones (*E. brevistylis* and *E. suberea*) lack flavanones. Similarly, in oil gland extracts of *L. scoparium species*, the β-triketone to flavanone ratio was >100:1 in *L. morrisonii* in favour of β-triketone, on the other hand, in *L. scoparium* (ornamental variety Rarawa) it was 1:6 in favour of flavanones (Killeen et al., 2015). In this study, this inverse relationship is true only for the *E. agglomerata-I* chemotypes that lacked methylated flavanones; instead, it predominantly contained conglomerone. *E. agglomerata* was previously reported to contain *C*-methylated flavanones (Goodger et al., 2019; Marsh et al., 2019). Contrary to expectations, this study did not find any *C*-methyl flavanones in *E. agglomerata*. Comprehensive analysis of the volatile fractions of *O*- and *C*-methylated flavanone accumulating species is warranted to validate this co-occurrence as well as to check any correlation between flavanone-triketone concentration.

4.4.3 Intra-and interspecific variation of flavanones and β-triketone in subgenus *Eucalyptus*

Metabolic profile variation between trees of different species is a well-known occurrence and the results presented in this chapter shows that there exists a significant difference in flavanone concentration between species as well as individual trees within a particular species. This observation reflects those of Marsh et al, (2019) that describe large variation in the total unsubstituted B-ring flavanone concentration between and within species. The flavanone profile of individual trees shown in this study is comparable with the findings of recent studies (Goodger et al., 2019; Goodger et al., 2016; Marsh et al., 2019). Dominant unidentified peaks with later retention times in the *E. stellulata* extract could be β-triketones (other than the ones detected in this study) or β-triketone-heterodimers or flavanone-triketone conjugates with retention times later than PS elucidated in Senaratne (2020).

Notable differences in metabolic profile observed within *E. diversifolia* and *E. agglomerata* individuals demonstrate different chemotypes. Chemotypes or chemical polymorphism is a well-known phenomenon in plants (Curado et al., 2006; Medina-Holguín et al., 2008). For example, two triketone-rich chemotypes were detected from geographically distinct populations of *Leptospermum scoparium* (Douglas et al., 2004). Similarly, four unique chemical profiles were seen in geographically distinct populations of *Anemopsis californica,* and the changes were suggested to be the result of genetic variability rather than environmental effects (Medina-Holguín et al., 2008). *Eucalyptus* has also been shown to display a variation in terpene profile within individuals of a species (Bustos-Segura et al., 2017; Goodger and Woodrow, 2012; Keszei et al., 2008). A difference in unsubstituted B-ring flavanone profile was also observed for *E. obliqua, E. globoidea* and *E. acmenoides* in a recent study (Marsh et al., 2019). Notably, in *E. obliqua* collected from different locations, one chemotype contained pinostrobin while the other was rich in dimethylpinocembrin. Interestingly, chemotypic variations in the above-mentioned studies were observed in trees from different locations. In contrast, all of the plants in this study, except three individual *E. agglomerata* plants, were grown in the same location, thus it is unlikely that the observed differences are due to environmental effects. For the plants in this study, information is only known about their respective mother trees from which the seeds were collected and details about the pollen source are unavailable. *Eucalyptus* trees are capable of producing interspecific hybrids between closely related species or between species with similar flower size (Delaporte et al., 2001; Griffin et al., 1988). Therefore, the observed chemotypic variation in individual trees within a species may be due to the genotypic differences arising from such hybridisation. This question deserves future in-depth studies.

4.4.4 *C*-methylation: a complex mechanism

For exploring flavanone *C*-methylation, the *E. diversifolia* transcriptome was successfully reconstructed using RNA-seq and *de novo* assembled using Trinity software. Overall, the BLASTx similarity analysis demonstrated a very low E-value distribution which emphasises significant homology and a high annotation accuracy. Average unigene length >1000bp is also a good sign of quality. Although cloning and expression were successful, we have been unsuccessful in finding the gene for CMT from *E. diversifolia.* Despite my attempts, no *C*-methylation was achieved on the tested flavanone substrates. Limited information on plant CMT sequence makes it difficult to identify more putative CMTs from *E. diversifolia* transcriptomic data. Also, flavanone CMTs may likely require specific *in vitro* conditions for their activity which needs to be investigated further. There is a chance, as pointed out in the introduction to this chapter, that *C*-methylation may be happening at an earlier stage in the biosynthetic pathway (Schröder et al., 1998). Failure of all six putative CMTs to catalyse flavanone *C*-methylation demonstrated in this chapter compels one to think about the likelihood of such an alternate biosynthetic route where pinocembrin formation is bypassed, and *C*-methylated flavanones are directly formed by a chalcone synthase variant. Such a mechanism will not result in the formation of pinostrobin or alpinetin. A lack of *O*-methylated flavanones in *E. diversifolia* shown in this study supports the alternate theory; however, *E. diversifolia* does accumulate pinocembrin at a concentration of up to 2 mg g^{-1}_{DW}. On the other hand, traces of DMM are also seen in *O*-methyl flavanone accumulating species *E. stellulata* and *E. haemastoma*. In addition to that, several *Eucalyptus* species have been identified to co-accumulate *O*-methylated compounds with *C*-methylated compounds (Marsh et al., 2019). In that case, there still exists the possibility of a normal chalcone synthase that could produce pinocembrin along with *C*-methylated flavonoids. It will be interesting to see a detailed flavanone profile of other *C*-methyl rich candidate species to see the variations and co-occurrence.

Transcriptomic data analysis of *E. diversifolia* resulted in the identification of secondary metabolite pathway genes putatively involved in *C*-methylation. Such an analysis of *C. deltoidea* has successfully identified transcripts encoding many enzymes involved in BIA (S)-reticuline biosynthesis (Zhong et al., 2020). Similarly, Priyavathi et al. (2018) has successfully identified lignin, flavonoid and terpene biosynthetic genes from the *Abelmoschus esculentus* leaf transcriptome data. Based on the speculated mechanism by Schroder et al. (1998), putative genes in methylmalonyl-CoA synthesis were also explored. The candidates suggested to have a close similarity to these enzymes might be a good target for further exploration of *C*-methylation in *Eucalyptus*. It should be noted that the pathway for methylmalonyl-CoA synthesis in plants is unclear and KEGG annotation of the putative genes involved is based on the elucidated pathways and enzymes in organisms other than plants (KEGG pathway ID: ko00640; ko00280). An interesting observation based on these identified KEGG pathways is that, if indeed *C*-methyl biosynthesis proceeds via methylmalonyl-CoA, then a CMT is unnecessary for incorporation of a methyl group. Instead, carboxylation of propionyl-CoA can form methylmalonyl-CoA. KEGG pathway analysis also identified two genes as putative malonyl-CoA/methylmalonyl-CoA synthetase which catalyse methylmalonate production (An and Kim, 1998; Crosby et al., 2012). Another enzyme, methylmalonyl-CoA mutase, can catalyse succinyl-CoA conversion to (R)-methylmalonyl-CoA (Zhang et al., 2010). Transcriptome data analysed in this study did not identify a candidate for this enzyme in *E.*

diversifolia. A more likely biosynthetic route in plants could be via this mutase as *Solanum tuberosum* (potato) fresh tuber extracts have been shown to possess Vitamin B_{12} dependent methylmalonyl-CoA mutase activity (Poston, 1978). Although our understanding of the role of vitamin B_{12} in plants remains ambiguous (Halarnkar and Blomquist, 1989), future research could consider this possibility.

4.5 Conclusion

The work presented in this chapter provides interesting insights into the occurrence and composition of flavanones and β-triketones in subgenus *Eucalyptus*. Two known *C*-methyl flavanones were identified in *E. diversifolia* in relatively high abundance. Notably, two species, *E. diversifolia* and *E. agglomerata*, accumulated β-triketone and chemotypic variations were observed in both species. The β-triketone conglomerone dominated the extract of one chemotype of *E. agglomerata*. Differences in the accumulation of these compounds are interesting and their ecological significance warrants further investigation.

In this study, six putative CMTs were successfully isolated and expressed; however, they did not show *in vitro* activity with the selected flavanones as substrates. Analysis of the sequenced transcriptome enabled the identification of additional putative candidates that could be analysed in future investigations.

CHAPTER 5
GENERAL DISCUSSION AND CONCLUSION

5.1 Introduction

Trees in the genus *Eucalyptus* are characterised by the presence of *O*- and *C*-methylated forms of unsubstituted B-ring flavanones. These minor flavonoid compounds have attracted attention recently as food supplements due to their therapeutic actions in humans. The primary aim of this study was to study flavanone methylation by isolating the enzymes involved. Chapter 1 details the current understanding of plant methyltransferase acting on shikimate-derived compounds. Chapters 2 and 3 describe flavanone *O*-methylation whereas Chapter 4 describes flavanone *C*-methylation. The contents of this chapter highlight the key findings discussed in previous chapters and explore their potential interconnectedness. Research questions that need further investigation in the future are also described.

5.2 Flavanone *O*-methylation

5.2.1 Two position-specific OMTs catalyse flavanone *O*-methylation in *Eucalyptus*

Eucalyptus subgenus *Eucalyptus* is known to contain at least three abundant *O*-methylated flavanone compounds – pinostrobin, dimethylpinocembrin and alpinetin. Kinetic properties of *7-OH* specific OMT, EnOMT1, from *O-methylated species*, defined the routes of pinostrobin and dimethylpinocembrin synthesis in *Eucalyptus* (**Figure 5.1**). Failure of EnOMT1 to catalyse *5-OH* methylation suggest that there exists another enzyme that has *5-OH* methylation activity. The high catalytic rate constant of EnOMT1 compared to other flavonoid OMTs might be the reason why there is a substantially lower abundance of alpinetin in most *Eucalyptus* trees surveyed in Chapter 4, as the alpinetin formed might immediately get converted to dimethylpinocembrin by highly active EnOMT1.

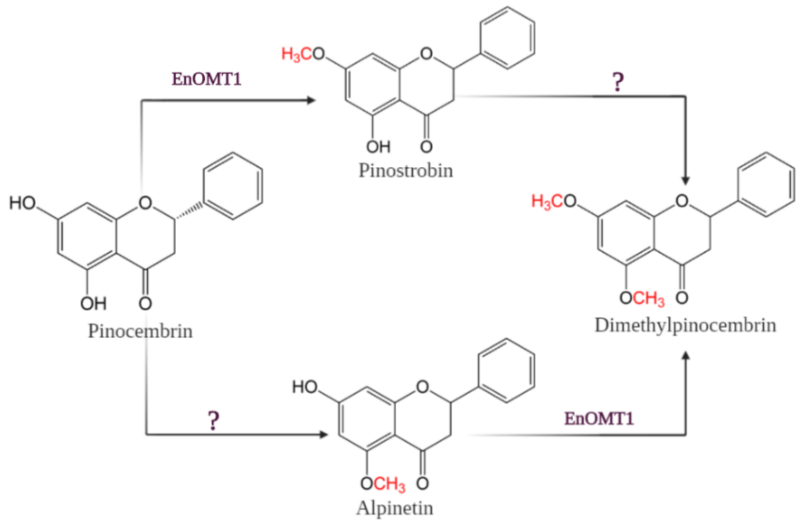


Figure 5.1: Flavanone *O*-methylation in *Eucalyptus*. EnOMT1 catalyzes 7-*O*-methylation of pinocembrin as well as alpinetin. Another 5-*OH* specific OMT that utilises pinocembrin to form alpinetin and pinostrobin to form dimethylpinocembrin is proposed.

5.2.2 Genetic basis of lack of *O*-methylated flavanones in some *Eucalyptus* species

Homologous genes of *EnOMT1* isolated using the sequence information and primer pairs of EnOMT1 suggests the reproducibility of this research in isolating these genes from *Eucalyptus*. Flavanone profiling showed that both *non-methylating species* and *E. diversifolia* (*C-methylating species*) does not accumulate the EnOMT1 catalysis product pinostrobin. Interestingly, the coding sequence of EnOMT1 homologous enzymes ExOMT1 and EdOMT1, isolated from *non-methylating species* and *C-methylating species*, respectively, share high sequence identity (six amino acid differences). Their alignment to C-terminal amino acid sequences of EnOMT1 also show differences at the same positions **(Figure 4.16)**. These amino acid changes might therefore be responsible for the difference in activity. Along with the site-directed mutation studies, this observation strengthens the argument that these non-conserved residues might therefore be responsible for the inactivity of the homologues leading to pinocembrin accumulation. Gene expression quantification using qPCR showed lower expression for *ExOMT1* and *EdOMT1*compared to that of the *O-methylated species EnOMT1*. Even with this low level of expression, both *non-methylated* and *C-methylated species* failed to show any detectable level of the EnOMT1 catalysis product pinostrobin in the individual chosen for flavanone profiling. Thus, the lack of *O*-methylation in these species might not be due to the lack of gene expression, but due to genetic changes that lead to an ineffective binding of substrates.

5.2.3 Structure-function relationship

The regiospecific transfer of a methyl group to the *7-OH* of flavanones and other flavonoid subclasses observed with EnOMT1 is supported by the results of a recent study. Park et al. (2020) found a flavonoid OMT from *Perilla frutescens* (PfOMT3) that selectively methylated *7-OH* of flavonoids with the highest catalytic efficiency for flavanone naringenin followed by the flavones apigenin and chrysin. Furthermore, substrate specificity analysis of PfOMT3 also supports the observation that flavonoid C-ring *3-OH*, as seen with quercetin in this study, is not favourable for activity. The C-ring *3-OH* may be interfering with the positioning of the substrate. The basis of the substrate discrimination is an interesting question that could be investigated through further structural studies.

Site-directed mutagenesis of the enzyme EnOMT1 along with the combined use of homology modelling and amino acid alignments of the EnOMT1 to homologous genes — ExOMT1, EcOMT1, and EdOMT1 — identified residues important for activity. As mentioned in the literature review (**section 1.6.3**; **section 1.7**), a C-terminal histidine residue is shown to be the catalytic residue in flavonoid OMTs (Zubieta et al., 2001), as well as other plant OMTs (Liscombe et al., 2012; Morris and Facchini, 2019) and it appears to be conserved in all flavonoid 7-OMTs characterised to date (Tang et al., 2020). Sequence alignment of EnOMT1 with homologous genes from subgenus *Eucalyptus* also shows a well-conserved histidine residue at an equivalent position 256 **(Figure 3.9**; **Figure 3.12**; **Figure 4.16)**. Therefore, it is reasonable to think that His255 is the catalytic residue in EnOMT1. A further study could assess the impact of His255 through site-directed mutation. In addition to the conserved catalytic His residue the adjacent residues, namely, Trp252, Val253, and Asn256, play critical associative roles, possibly forming a channel for methyl passage (Cabry et al., 2019). Most of the amino acid differences toward the C-terminal of plant MTs — as seen between EnOMT1 and ExOMT1 — are associated with the positioning of SAM (Liscombe et al., 2012; Morris and Facchini, 2019). Substrate binding is often associated with the N-terminal region forming the dimer (Morris and Facchini, 2019; Zubieta et al., 2001). The findings of this study are insufficient to clarify the specific reason behind the lack or reduced activity in the mutant forms of EnOMT1. It is uncertain whether the identified residues interact with SAM or pinocembrin. The lack of crystal structure of a flavonoid OMT that can accept flavanones makes it difficult to pinpoint the specific pinocembrin binding residues by homology modelling. Structure elucidation of EnOMT1 is therefore warranted and would make a substantial contribution to the comparative protein structure modelling field.

Knowledge of structural impacts of active-site residues can enable the production of enzymes with novel activities. For example, a recent study on sequence and crystal structure comparison between two isoflavonoid and phenylpropanoid OMTs (PDB code 1FP2 and 3REO respectively) identified a single mutation of T133M in *C. breweri* isoeugenol OMT which widened its substrate range to flavonoids and demonstrated a shift in regiospecificity from B-ring hydroxyls (3'-*OH*, 4'-*OH*) to A-ring 7-*OH* when naringenin and genistein, flavonoids that lack 3'-*OH*, were used as substrate (Tang et al., 2020). Similarly, the crystal structure of EnOMT1 could enable the identification of specific substrate binding residues, and site-directed mutagenesis of these residues would be a suitable tool to engineer new substrate preferences, possibly performing *5-OH* methylation. It is noteworthy that non-catalytic site residues also contribute to the overall enzyme structure and catalytic site topology. Such complexities should also be considered during engineering new substrate preferences.

5.2.4 EnOMT1: a tool for the biotechnological production of pinostrobin

Pinostrobin has recently attracted attention as a food supplement due to its therapeutic actions in humans (Patel et al., 2016). Both mono-*O*-methylated flavanones, pinostrobin and alpinetin, have previously been produced using *Boesenbergia rotunda* callus and cell suspension as well as adventitious root cultures (Rahmat and Kang, 2019; Yusuf et al., 2013). Precursor feeding and elicitation techniques have been effectively employed for the induction as well as increased production of desired secondary metabolites in plant cell suspension cultures (Yue et al., 2016). This study explored cell suspension and callus cultures as alternative sources of pinostrobin production. Whilst the results did not produce detectable amounts of flavanones in cell cultures, it demonstrated the establishment of *in vitro* cultures which in future could act as a platform for bioactive compound production through chemical elicitation studies. Further research should focus on using modified growth media, precursor feeding of hydrophilic compounds like phenylalanine, and fungal or jasmonic acid elicitation to further explore pinostrobin production. *Ex planta* production of pharmaceutically important compounds through bioengineered bacterial cells is also a sustainable approach to their commercialisation. Recently, *de novo* production of pinocembrin (up to 198 mg L^{-1}), naringenin, eriodictyol and homoeriodictyol from glycerol was achieved through optimised microbial production strains that did not need precursor feeding (Dunstan et al., 2020). Isolation of EnOMT1 opens up the possibility of pinostrobin production using such a platform.

5.3 Flavanone *C*-methylation

5.3.1 Possible biosynthetic routes of *C*-methylated flavanones in *Eucalyptus*

Six of the putative CMT clones isolated in chapter 4 failed to methylate flavanones *in vitro*. Clones EdCMT 1 through 4 were annotated as caffeic acid OMT-like, whereas the remainder were annotated as sterol methyltransferases. Their sequence alignment showed conserved methyltransferase motifs along with the presence of dimerization and methylation-specific domains (from PFAM based functional annotation), thereby reinforcing that they are in fact methyltransferases. This study did not evaluate the activity of these clones with other classes of possible substrates and further research could assess the activity of these isolated clones with phenylpropanoids and sterols.

The failure to identify CMTs that directly methylate flavanones is suggestive, but not conclusive, evidence of an alternate methylation route to *C*-methyl flavanones. The flavanone A-ring is described to be derived from the malonyl-CoA mediated polyketide moiety whereas the B-ring purportedly arises from cinnamate derivatives (Ayabe and Furuya, 1982; Ohkatsu et al., 2010). A recent study employing engineered *E. coli* harbouring the biosynthetic enzymes required for flavanone production showed that they synthesised only the unsubstituted B-ring flavanone pinocembrin in the presence of phenylalanine, and produced B-ring substituted flavanones in the presence of tyrosine and tyrosine ammonia lyase, again showing that the B-ring originates from caffeate/cinnamate derivatives (Dunstan et al., 2020). KEGG pathway analysis of both transcriptomes studied here did not identify any candidate for tyrosine ammonia-lyase (EC 4.2.1.25; **Figure 4.15**) suggesting a lack of this enzyme activity which may be one of the reasons why subgenus *Eucalyptus* accumulate only B-ring unsubstituted flavanones. Based on the flavonoid compounds identified so far, as mentioned in section **1.3.2.iv**, *C*-methylation is seen only on the flavonoid A-ring. This observation further implies the incorporation of *C*-methyl through the di- or triketide entity at some point in chalcone synthesis.

Incorporation of *C*-methyl through a polyketide synthase (PKS) is a possible route to *C*-methyl flavanones in *Eucalyptus*. Type-III PKS are homodimeric and use free CoA thioesters as substrates without the involvement of acyl carrier protein to catalyse polyketide formation. In plants, type-III PKSs include chalcone synthase, stilbene synthase, styrylpyrone synthase, benzalacetone synthase, and 4-coumaroyl triacetic acid synthase. Members of the plant type-

III PKS family are shown to be functionally divergent as they exhibit broad substrate specificities as well as accept a variety of physiological and non-physiological substrates to produce a vast array of chemically and structurally distinct unnatural polyketides (Abe et al., 2006a). Several plant enzymes have been demonstrated to incorporate *C*-methyl groups on secondary metabolites through methylmalonyl-CoA catalysis by plant type-III PKS (Abe et al., 2006a; Abe and Morita, 2010; Abe et al., 2006b; Schröder et al., 1998). For example, in *P. strobus*, a chalcone synthase-like type III PKS (PStrCHS2), inactive with its normal substrates (phenylpropanoid CoA esters) utilised cinnamoyl diketide-N-acetylcysteamine (NAC) thioester as a substrate and preferred methylmalonyl-CoA only for the diketide to triketide conversion step, thereby explaining the mono *C*-methylated product formation (Schröder et al., 1998). [NAC is a synthetically tractable CoA-mimic used to activate intermediate analogues or unnatural starters for entry into PKS systems (Austin and Noel, 2003; Oguro et al., 2004)]. Schroder and his colleagues (1998) proposed that *in vivo* PStrCHS2 acts as a modular PKS in association with another regular CHS and the reaction occurs through substrate channelling. Their data are consistent with such an interpretation, however, neither the cooperative activity nor the existence of such a complex has yet been demonstrated *in vitro*. While PStrCHS2 strictly carried out a one-step condensation with methylmalonyl-CoA, a PKS from *Aloe arborescens, Rheum palmatum* and a CHS from *Scutellaria baicalensis* efficiently utilised methylmalonyl-CoA as a starter as well as extender unit to produce unnatural polyketides (Abe et al., 2000). *In vitro* synthesis of *C*-methylated quinolone alkaloids was achieved by the type III PKS, benzalacetone synthase from *Rheum palmatum* (Chinese rhubarb), utilizing *N*-methylanthraniloyl-CoA and methylmalonyl-CoA for condensation (Abe et al., 2006a). Notably, both CoA esters are non-physiological substrates for *R. palmatum*. Even though *in vitro* utilization of methylmalonyl-CoA by plant PKS is demonstrated, the presence of methylmalonyl-CoA, as well as their biosynthesis in plants, is also questionable as a majority of the enzymes involved depend on Vitamin B_{12} as a cofactor (Takahashi-Iñiguez et al., 2012) and plants are unlikely to synthesise Vitamin B_{12} or contain enzymes that depend on it (Ji et al., 2021; Paizs et al., 2008). Potato tuber extract was reported to contain a Vitamin B_{12} dependent enzyme methylmalonyl-CoA mutase that synthesises methylmalonyl-CoA (Poston, 1978); however, it was later found to be a phosphatase methylmalonyl-3′-dephospho-CoA that does not need coenzyme B_{12} as a cofactor (Paizs et al., 2008). More research into methylmalonyl-CoA biosynthesis in plants is needed to substantiate its involvement in *C*-methylation.

Another possible route of flavanone *C*-methylation is via a PKS variant that directly transfers the *C*-methyl group to the polyketide entity (Shi et al., 1996a). In bacterial terpenoid biosynthesis, a SAM-dependant CMT (TleD) acts as a methyltransferase as well as a terpene cyclase (Yu et al., 2016). Briefly, methylation catalysed by TleD results in subsequent carbocation (carbonate ion) formation which in turn promoted a cationic reaction cascade to form a cyclic monoterpenoid. The likelihood of having terpenoid CMT homologues in *Eucalyptus* is high as the species are rich in various mono- and sesquiterpenoid compounds (Goodger et al., 2013; Maciel et al., 2010). Evolutionary variations might have enabled the terpenoid multifunctional methyl-cyclase enzyme to function in both flavonoid and terpenoid biosynthesis. In support of this, the crystal structure of TleD (bacterial terpenoid CMT; PDB ID: 5GM2; Yu et al., 2016) has been elucidated and its tertiary structure is strikingly similar to plant phenolic methyltransferases. One such feature is the homodimeric form in which the N-terminal helix of the monomer chain acting as the backwall of the neighbouring monomer's substrate binding site; however, these interacting homodimers formed a hexamer in solution (Yu et al., 2016). Interestingly, chalcone synthase in *M sativa* (PDB ID:1CGK) was also shown to be homodimeric with functionally dependent monomers (Abe and Morita, 2010; Ferrer et al., 1999). The tertiary structure of a large multidomain fungal PKS with a *C*-methyltransferase domain catalysing multiple *C*-methylation in polyketide citrinin synthesis (PDB ID:5MPT) demonstrated the strong interdependence of *C*-methyltransferase and ketone synthase domains. The CMT domain organised as two subdomains contained a novel N-terminal subdomain responsible for substrate binding (Storm et al., 2017). An implication of this is the possibility of a programmed association of a PKS and the isolated CMTs in this study which would be a fruitful area for further work.

5.4 β-triketone biosynthesis in *Eucalyptus*

The occurrence of *O*-methylated flavanones and the *O*-methylated triketone conglomerone with a complete lack of *C*-methylated metabolites in *E. agglomerata*, together with *C*-methylated flavanone accumulation and the *C*-methylated triketone agglomerone in *E. diversifolia*, indicate that their biosynthesis might be linked. Further evidence of this theory arises from the seemingly inverse relationship between *C*-methylated flavanones and the β-triketone grandiflorone seen in *Leptospermum* species. Due to certain structural similarities between these two types of compounds, their biosynthesis was postulated to be linked via the action of polyketide synthases (Killeen et al., 2015). Triketones are considered to be phloroglucinol derivatives (Ghisalberti, 1996) and phloroglucinol is synthesised by type III PKS in bacteria (Achkar et al., 2005). Therefore, it is likely that in *Eucalyptu*s a polyketide synthase with a methyltransferase domain could catalyse synthesis of β-triketones and flavanone compounds. In the *E. diversifolia* transcriptome, twenty-five unigenes were annotated using the Pfam database as type-III PKS-like (Appendix C) and one as polyketide cyclase (Cluster-17119.1; Appendix C); however, none of them contained a methyltransferase domain.

Purification and sequencing of crude CMT from leaf extracts could be another possible way of exploring the enzyme responsible for flavanone *C*-methylations. Many novel enzymes which do not have any homologous genes were previously identified using this technique. Cell-free extracts without small molecule filtration can be used as a crude enzyme assayed with substrates like phenylalanine or cinnamoyl-CoA to see if they make methylated compounds. If the crude extract exhibits activity with a particular substrate, the CMT responsible can be concentrated, purified and sequenced to get more information regarding the sequence. For example, the majority of the flavonoid OMTs involved in *Catharanthus roseus* were isolated from protein extracts of cell suspension cultures through several purification steps, including affinity chromatography on adenosine agarose (Cacace et al., 2003) and anti-OMT antibody chromatography (Huang et al., 2004) followed by peptide sequencing. Lack of information on CMT proteins makes it difficult to purify to homogeneity; however, combined use of the activity assays and selection based on the methyltransferase domain might enable CMT enrichment in the extract. Once purified, the enzyme could be sequenced to get peptide sequences that would facilitate the identification/isolation of the full sequence. An already available full transcriptome of *E diversifolia* will enable easy identification of the partial peptide sequence from such an experiment.

5.5 Conclusion

This study set out to investigate flavanone methylation in *Eucalyptus.* The results reported in this study gives a clear picture of flavanone *O*-methylation in *Eucalyptus* and provide a solid platform of knowledge for future research on flavanone *C*-methylation. *Eucalyptus* subgenus *Eucalyptus* is known to contain at least three major *O*-methylated flavanone compounds — pinostrobin, dimethylpinocembrin and alpinetin in lower abundance. During this investigation to uncover the OMTs involved in the methylation of the flavanone aglycone pinocembrin, a flavanone 7-OMT, EnOMT1, was successfully isolated and characterized. The biochemical characterization showed that EnOMT1 is an efficient methyltransferase enzyme. EnOMT1 is a regiospecific methyltransferase with strict positional specificity to the 7-hydroxyl of flavonoids. The reduction/absence of activity of EnOMT1 with B-ring substituted flavonoids such as luteolin, apigenin, and quercetin demonstrates the drastic effect B-ring hydroxyl configuration has on the activity of the enzyme. The result suggests that the C-ring double bond between carbon 2 and carbon 3 (the only difference between chrysin and pinocembrin 2D-structure) does not affect the *7-O*-methylation of flavonoids by EnOMT1. Therefore, in *Eucalyptus* species accumulating flavones, it may be possible that EnOMT1 is responsible for the synthesis of their methylated forms.

An EnOMT1 homologous gene (ExOMT1) was isolated and characterised from a species that does not accumulate any methylated flavanones. The observed difference in activity is likely due to the amino acid difference of the EnOMT1 sequence to ExOMT1. The fact that ExOMT1 showed minor activity at a higher pinocembrin concentration and longer incubation time and produced low levels of pinostrobin, shows that the enzyme ExOMT1, although relatively inefficient, can still bind to pinocembrin. Hence, it is possible that ExOMT1 lost the ability to efficiently utilise the substrates, so that even after having access to a very high concentration of pinocembrin *in vivo*, ExOMT1 fails to methylate it. Another possibility is that the substrate pinocembrin might be getting sequestered before the enzyme can access it. Site-directed mutagenesis further identified residues important for activity. In addition to the catalytic His residue, the adjacent residues — namely Trp252, Val253, and Asn256 — play critical associative roles for activity. The phylogenetic analysis of the OMTs isolated in this study with fifty-six biochemically characterized plant OMTs accepting a broad variety of substrates showed that EnOMT1 is closely related to class II OMTs and clustered with other plant flavonoid 7-OMTs. Nevertheless, xanthohumol 4-*O*-methyltransferase (X4OMT; Uniprot ID-B0ZB56.1) from *Humulus lupulus* acting on prenylchalcone is shown as the closest evolutionary relative. Tertiary structure prediction through homology modelling showed EnOMT1 to be a homodimer and SAM binding residues — namely, Trp148, Phe161, Met165, Asn169, Gln194, Phe216, Asp217, Arg218, Asp237, Met238, Phe239, Lys251, Trp252, Val253, and Trp257 — were predicted with a higher degree of confidence by comparing with the similar templates. A natural progression of this work is to analyse the tertiary structure of EnOMT1 which will contribute greatly to our understanding of flavanone OMTs.

The work presented in chapter 4 provides interesting insights into the co-occurrence and composition of flavanones and β-triketones in subgenus *Eucalyptus*. Screening of multiple individuals of ten different species belonging to subgenus *Eucalyptus* showed that there exists intra- as well as inter-specific variation of metabolite profiles. Quantitative variations in flavanone concentrations were also evident between individual trees within species. Two known *C*-methyl flavanones, cryptostrobin (monomethylated at C6) and desmethoxymatteucinol (dimethylated at C6 and C8), were identified in *E diversifolia* in abundance. This is the first evidence of a species from section *Longistylus* of subgenus *Eucalyptus* harbouring *C*-methylated flavanones in high concentration. Neither 6-C-methylated flavanone strobopinin nor flavanones with combinations of *O*- and *C*-methylation were detected in any extracts analysed in this study. Chemotypic variation was evident in two species, notably, *E. diversifolia* and *E. agglomerata*, which accumulated β-triketone along with flavanones. The β-triketone conglomerone dominated the extract of one chemotype of *E. agglomerata* whereas the other chemotype contained *O*-methylated flavanones with conglomerone in comparable quantities. On the other hand, the *E. diversifolia I* chemotype had higher levels of β-triketone agglomerone in them, and chemotypic variations were due to the different complements of accumulated triketones. Since all individuals analysed were from the same geographic location, the observed chemotypic variation might be due to the genotypic differences arising from hybridisation with related species. This question deserves future in-depth studies. The localisation of β-triketones and *C*- and *O*-methylated flavanones to foliar secretory glands of *E agglomerata* is the first evidence showing the co-occurrence of β-triketones with flavanones in *Eucalyptus*. Whether these two types of compounds are housed together in individual or separate foliar glands is still unclear. Individual gland isolation and analysis of triketone abundant species might provide further information regarding their storage location and ecological significance.

Transcriptomic sequencing of *E. diversifolia* was conducted to hunt for the CMTs involved in cryptostrobin and desmethoxymatteucinol biosynthesis. The *E. diversifolia* transcriptome was successfully reconstructed using RNA-seq and *de novo* assembled. The transcriptome data were analysed for the identification of secondary metabolite pathway genes putatively involved in *C*-methylation. Six putative CMT genes were cloned successfully and analysed for activity. Despite our attempts, no *in vitro* *C*-methylation was achieved on the flavanone substrates tested. The limited published information on plant CMT sequences makes it difficult to identify more putative CMTs from the *E. diversifolia* transcriptome data. Also, flavanone CMTs may likely require specific *in vitro* conditions for their activity which needs to be investigated further.

Whilst the precise mechanism of *C*-methylation in plants remains to be elucidated, it is also possible that the *C*-methyl group is incorporated earlier in the biosynthetic pathway before pinocembrin formation. Analysis of the sequenced transcriptome enabled the identification of additional putative candidates that could be examined in future investigations. This study presents alternate biosynthetic routes as an option with the support of existing literature and provides insights for future research. These alternate routes might include the involvement of a plant type III polyketide synthase, either utilising the bulky intermediate methylmalonyl-CoA or catalysing direct *C*-methyl transfer to the polyketide entity during chalcone synthesis. Co-occurrence of *C*-methylated triketones with *C*-methylated flavanones also suggests a possible polyketide synthase-mediated connection in their biosynthesis. Comprehensive analysis of the volatile fractions of *O*- and *C*-methylated flavanone accumulating species is warranted to validate their co-occurrence with triketones as well as to investigate any correlation between flavanone and triketone concentrations.

REFERENCES

Abe, I., Abe, T., Wanibuchi, K. and Noguchi, H. (2006a) Enzymatic formation of quinolone alkaloids by a plant type III polyketide synthase. *Organic letters* **8**, 6063-6065.

Abe, I. and Morita, H. (2010) Structure and function of the chalcone synthase superfamily of plant type III polyketide synthases. *Natural product reports* **27**, 809-838.

Abe, I., Morita, H., Nomura, A. and Noguchi, H. (2000) Substrate specificity of chalcone synthase: enzymatic formation of unnatural polyketides from synthetic cinnamoyl-CoA analogues. *Journal of the American Chemical Society* **122**, 11242-11243.

Abe, T., Noma, H., Noguchi, H. and Abe, I. (2006b) Enzymatic formation of an unnatural methylated triketide by plant type III polyketide synthases. *Tetrahedron letters* **47**, 8727-8730.

Achkar, J., Xian, M., Zhao, H. and Frost, J. (2005) Biosynthesis of phloroglucinol. *Journal of the American chemical society* **127**, 5332-5333.

Adams, R.P. (2017) *Identification of essential oil components by gas chromatography/mass spectrometry*:Allured publishing corporation Carol Stream, IL.

Akashi, T., VanEtten, H.D., Sawada, Y., Wasmann, C.C., Uchiyama, H. and Ayabe, S.-i. (2006) Catalytic specificity of pea O-methyltransferases suggests gene duplication for (+)-pisatin biosynthesis. *Phytochemistry* **67**, 2525-2530.

An, J.H. and Kim, Y.S. (1998) A gene cluster encoding malonyl-CoA decarboxylase (MatA), malonyl-CoA synthetase (MatB) and a putative dicarboxylate carrier protein (MatC) in Rhizobium trifolii: Cloning, sequencing, and expression of the enzymes in Escherichia coli. *European journal of biochemistry* **257**, 395-402.

Annadurai, R.S., Neethiraj, R., Jayakumar, V., Damodaran, A.C., Rao, S.N., Katta, M.A., Gopinathan, S., Sarma, S.P., Senthilkumar, V. and Niranjan, V. (2013) De Novo transcriptome assembly (NGS) of Curcuma longa L. rhizome reveals novel transcripts related to anticancer and antimalarial terpenoids. *PloS one* **8**, e56217.

Atkinson, R.G. (2018) Phenylpropenes: Occurrence, Distribution, and Biosynthesis in Fruit. *Journal of Agricultural and Food Chemistry* **66**, 2259-2272.

Atun, S., Arianingrum, R., Sulistyowati, E. and Aznam, N. (2013) Isolation and antimutagenic activity of some flavanone compounds from Kaempferia rotunda. *International Journal of Chemical and Analytical Science* **4**, 3-8.

Austin, M.B. and Noel, J.P. (2003) The chalcone synthase superfamily of type III polyketide synthases. *Natural product reports* **20**, 79-110.

Awakawa, T., Zhang, L., Wakimoto, T., Hoshino, S., Mori, T., Ito, T., Ishikawa, J., Tanner, M.E. and Abe, I. (2014) A methyltransferase initiates terpene cyclization in teleocidin B biosynthesis. *Journal of the American Chemical Society* **136**, 9910-9913.

Ayabe, S.-i. and Furuya, T. (1982) Studies on plant tissue cultures. Part 36. Biosynthesis of a retrochalcone, echinatin, and other flavonoids in the cultured cells of Glycyrrhiza echinata. A new route to a chalcone with transposed A-and B-rings. *Journal of the Chemical Society, Perkin Transactions 1*, 2725-2734.

Barron, D., Laflamme, P. and De Luca, V. (2020) Journey in the Polyphenol Research World with Ragai Ibrahim. *Journal of agricultural and food chemistry* **68**, 2849-2860.

Bauerle, M.R., Schwalm, E.L. and Booker, S.J. (2015) Mechanistic diversity of radical S-adenosylmethionine (SAM)-dependent methylation. *Journal of Biological Chemistry* **290**, 3995-4002.

Bayer, T.S., Widmaier, D.M., Temme, K., Mirsky, E.A., Santi, D.V. and Voigt, C.A. (2009) Synthesis of methyl halides from biomass using engineered microbes. *Journal of the American Chemical Society* **131**, 6508-6515.

Bedane, K.G., Kusari, S., Masesane, I.B., Spiteller, M. and Majinda, R.R. (2016) Flavanones of Erythrina livingstoniana with antioxidant properties. *Fitoterapia* **108**, 48-54.

Bennett, M.R., Thompson, M.L., Shepherd, S.A., Dunstan, M.S., Herbert, A.J., Smith, D.R., Cronin, V.A., Menon, B.R., Levy, C. and Micklefield, J. (2018) Structure and Biocatalytic Scope of Coclaurine N-Methyltransferase. *Angewandte Chemie International Edition* **57**, 10600-10604.

Berim, A., Hyatt, D.C. and Gang, D.R. (2012) A set of regioselective O-methyltransferases gives rise to the complex pattern of methoxylated flavones in sweet basil. *Plant physiology*

160, 1052-1069.

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

Brunet, G. and Ibrahim, R.K. (1980) O-Methylation of flavonoids by cell-free extracts of calamondin orange. *Phytochemistry* **19**, 741-746.

Bustos-Segura, C., Dillon, S., Keszei, A., Foley, W.J. and Külheim, C. (2017) Intraspecific diversity of terpenes of Eucalyptus camaldulensis (Myrtaceae) at a continental scale. *Australian Journal of Botany* **65**, 257-269.

Cabry, M.P., Offen, W.A., Saleh, P., Li, Y., Winzer, T., Graham, I.A. and Davies, G.J. (2019) Structure of Papaver somniferum O-methyltransferase 1 reveals initiation of noscapine biosynthesis with implications for plant natural product methylation. *ACS Catalysis* **9**, 3840-3848.

Cacace, S., Schröder, G., Wehinger, E., Strack, D., Schmidt, J. and Schröder, J. (2003) A flavonol O-methyltransferase from Catharanthus roseus performing two sequential methylations. *Phytochemistry* **62**, 127-137.

Cao, H., Jing, X., Wu, D. and Shi, Y. (2013) Methylation of genistein and kaempferol improves their affinities for proteins. *Int J Food Sci Nutr* **64**, 437-443.

Catanzariti, A.M., Soboleva, T.A., Jans, D.A., Board, P.G. and Baker, R.T. (2004) An efficient system for high-level expression and easy purification of authentic recombinant proteins. *Protein Science* **13**, 1331-1339.

Cázares-Flores, P., Levac, D. and De Luca, V. (2016) Rauvolfia serpentina N-methyltransferases involved in ajmaline and Nβ-methylajmaline biosynthesis belong to a gene family derived from γ-tocopherol C-methyltransferase. *The Plant Journal* **87**, 335-342.

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.

Chen, H., Deng, C., Nie, H., Fan, G. and He, Y. (2017) Transcriptome analyses provide insights into the difference of alkaloids biosynthesis in the Chinese goldthread (Coptis chinensis Franch.) from different biotopes. *PeerJ* **5**, e3303.

Chen, L., Teng, H., Xie, Z., Cao, H., Cheang, W.S., Skalicka-Woniak, K., Georgiev, M.I. and Xiao, J. (2018) Modifications of dietary flavonoids towards improved bioactivity: An update on structure–activity relationship. *Critical reviews in food science and nutrition* **58**, 513-527.

Cheng, X., Kumar, S., Posfai, J., Pflugrath, J.W. and Roberts, R.J. (1993) Crystal structure of the Hhal DNA methyltransferase complexed with S-adenosyl-L-methionine. *Cell* **74**, 299-307.

Chiron, H., Drouet, A., Claudot, A.-C., Eckerskorn, C., Trost, M., Heller, W., Ernst, D. and Sandermann, H. (2000) Molecular cloning and functional expression of a stress-induced multifunctional O-methyltransferase with pinosylvin methyltransferase activity from Scots pine (Pinus sylvestris L.). *Plant molecular biology* **44**, 733-745.

Christensen, A.B., Gregersen, P.L., Olsen, C.E. and Collinge, D.B. (1998) A flavonoid 7-O-methyltransferase is expressed in barley leaves in response to pathogen attack. *Plant molecular biology* **36**, 219-227.

Coiner, H., Schröder, G., Wehinger, E., Liu, C.J., Noel, J.P., Schwab, W. and Schröder, J. (2006) Methylation of sulfhydryl groups: a new function for a family of small molecule plant O-methyltransferases. *The Plant Journal* **46**, 193-205.

Conde, E., Cadahia, E. and Garcia-Vallejo, M. (1992) Optimization of TLC for research on the flavonoids in wood and bark of species of the genusEucalyptus L'Héritier. *Chromatographia* **33**, 418-426.

Consortium, G.O. (2008) The gene ontology project in 2008. *Nucleic acids research* **36**, D440-D444.

Cooke, H.A., Guenther, E.L., Luo, Y., Shen, B. and Bruner, S.D. (2009) Molecular basis of substrate promiscuity for the SAM-dependent O-methyltransferase NcsB1, involved in the biosynthesis of the enediyne antitumor antibiotic neocarzinostatin. *Biochemistry* **48**, 9590-9598.

Courtney, J.L., Lassak, E.V. and Speirs, G.B. (1983) Leaf wax constituents of some myrtaceous species. *Phytochemistry* **22**, 947-949.

Coutinho, I.D., Coelho, R.G., Kataoka, V., Honda, N.K., Silva, J., Vilegas, W. and Cardoso, C.A.L. (2008) Determination of phenolic compounds and evaluation of antioxidant capacity of Campomanesia adamantium leaves. *Ecletica química* **33**, 53-60.

Crosby, H.A., Rank, K.C., Rayment, I. and Escalante-Semerena, J.C. (2012) Structure-guided expansion of the substrate range of methylmalonyl coenzyme A synthetase (MatB) of Rhodopseudomonas palustris. *Applied and environmental microbiology* **78**, 6619-6629.

Cserzö, M., Wallin, E., Simon, I., Von Heijne, G. and Elofsson, A. (1997) Prediction of transmembrane alpha-helices in prokaryotic membrane proteins: the dense alignment surface method. *Protein engineering* **10**, 673-676.

Curado, M.A., Oliveira, C.B., Jesus, J.G., Santos, S.C., Seraphin, J.C. and Ferri, P.H. (2006) Environmental factors influence on chemical polymorphism of the essential oils of Lychnophora ericoides. *Phytochemistry* **67**, 2363-2369.

D'Andrea, G. (2015) Quercetin: a flavonol with multifaceted therapeutic applications? *Fitoterapia* **106**, 256-271.

D Auria, J.C., Chen, F. and Pichersky, E. (2003) The SABATH family of MTs in Arabidopsis thaliana and other plant species. *Recent advances in phytochemistry* **37**, 253-284.

Dang, T.-T.T. and Facchini, P.J. (2012) Characterization of three O-methyltransferases involved in noscapine biosynthesis in opium poppy. *Plant physiology* **159**, 618-631.

Dao, T.-T., Tung, B.-T., Nguyen, P.-H., Thuong, P.-T., Yoo, S.-S., Kim, E.-H., Kim, S.-K. and Oh, W.-K. (2010) C-methylated flavonoids from Cleistocalyx operculatus and their inhibitory effects on novel influenza A (H1N1) neuraminidase. *Journal of natural products* **73**, 1636-1642.

Davidson, N.M. and Oshlack, A. (2014) Corset: enabling differential gene expression analysis for de novo assembled transcriptomes. *Genome biology* **15**, 410.

Davin, L.B. and Lewis, N.G. (1992) Phenylpropanoid metabolism: biosynthesis of monolignols, lignans and neolignans, lignins and suberins. In: *Phenolic metabolism in plants* pp. 325-375. Springer.

Dayan, F.E., Howell, J.L., Marais, J.P., Ferreira, D. and Koivunen, M. (2011) Manuka oil, a natural herbicide with preemergence activity. *Weed science* **59**, 464-469.

De Luca, V. and Ibrahim, R.K. (1985) Enzymatic synthesis of polymethylated flavonols in Chrysosplenium americanum. I. Partial purification and some properties of S-adenosyl-L-methionine:flavonol 3-, 6-, 7-, and 4'-O-methyltransferases. *Arch Biochem Biophys* **238**, 596-605.

De Luca, V., Salim, V., Levac, D., Atsumi, S.M. and Yu, F. (2012) Discovery and functional analysis of monoterpenoid indole alkaloid pathways in plants. In: *Methods in enzymology* pp. 207-229. Elsevier.

de Oliveira, L.A., Breton, M.C., Bastolla, F.M., Camargo, S.d.S., Margis, R., Frazzon, J. and Pasquali, G. (2011) Reference genes for the normalization of gene expression in Eucalyptus species. *Plant and Cell Physiology* **53**, 405-422.

Delaporte, K., Conran, J. and Sedgley, M. (2001) Interspecific hybridization within Eucalyptus (myrtaceae): subgenus symphyomyrtus, sections Bisectae and Adnataria. *International Journal of Plant Sciences* **162**, 1317-1326.

Desgagné-Penix, I. and Facchini, P.J. (2012) Systematic silencing of benzylisoquinoline alkaloid biosynthetic genes reveals the major route to papaverine in opium poppy. *The Plant Journal* **72**, 331-344.

Diamond, A. and Desgagné-Penix, I. (2016) Metabolic engineering for the production of plant isoquinoline alkaloids. *Plant biotechnology journal* **14**, 1319-1328.

Dixon, M.M., Huang, S., Matthews, R.G. and Ludwig, M. (1996) The structure of the C-terminal domain of methionine synthase: presenting S-adenosylmethionine for reductive methylation of B12. *Structure* **4**, 1263-1275.

Dixon, R., Dey, P. and Lamb, C. (1983) Phytoalexins: enzymology and molecular biology. *Advances in enzymology and related areas of molecular biology* **55**.

Douglas, M.H., van Klink, J.W., Smallfield, B.M., Perry, N.B., Anderson, R.E., Johnstone, P.

and Weavers, R.T. (2004) Essential oils from New Zealand manuka: triketone and other chemotypes of Leptospermum scoparium. *Phytochemistry* **65**, 1255-1264.

Dudareva, N. and Pichersky, E. (2000) Biochemical and molecular genetic aspects of floral scents. *Plant physiology* **122**, 627-634.

Dunstan, M.S., Robinson, C.J., Jervis, A.J., Yan, C., Carbonell, P., Hollywood, K.A., Currin, A., Swainston, N., Feuvre, R.L., Micklefield, J., Faulon, J.-L., Breitling, R., Turner, N., Takano, E. and Scrutton, N.S. (2020) Engineering Escherichia coli towards de novo production of gatekeeper (2S)-flavanones: naringenin, pinocembrin, eriodictyol and homoeriodictyol. *Synthetic Biology* **5**.

Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research* **32**, 1792-1797.

Farrell, J.D., Byrne, S., Paina, C. and Asp, T. (2014) De novo assembly of the perennial ryegrass transcriptome using an RNA-Seq strategy. *PloS one* **9**, e103567.

Farrow, S.C., Kamileen, M.O., Meades, J., Ameyaw, B., Xiao, Y. and O'Connor, S.E. (2018) Cytochrome P450 and O-methyltransferase catalyze the final steps in the biosynthesis of the anti-addictive alkaloid ibogaine from Tabernanthe iboga. *Journal of Biological Chemistry* **293**, 13821-13833.

Fauman, E.B., Blumenthal, R.M. and Cheng, X. (1999) Structure and evolution of AdoMet-dependent methyltransferases. *S-Adenosylmethionine-dependent methyltransferases: structures and functions*, 1-38.

Fernández, M., Troncoso, V. and Valenzuela, S. (2015) Transcriptome profile in response to frost tolerance in Eucalyptus globulus. *Plant molecular biology reporter* **33**, 1472-1485.

Fernando, S.C., Goodger, J.Q.D., Gutierrez, S.S., Johnson, A.A.T. and Woodrow, I.E. (2016) Plant regeneration through indirect organogenesis and genetic transformation of Eucalyptus polybractea R.T. Baker. *Industrial Crops and Products* **86**, 73-78.

Ferrer, J.-L., Jez, J.M., Bowman, M.E., Dixon, R.A. and Noel, J.P. (1999) Structure of chalcone synthase and the molecular basis of plant polyketide biosynthesis. *Nature structural biology* **6**, 775-784.

Ferrer, J.-L., Zubieta, C., Dixon, R.A. and Noel, J.P. (2005) Crystal structures of alfalfa caffeoyl coenzyme A 3-O-methyltransferase. *Plant physiology* **137**, 1009-1017.

Frick, S. and Kutchan, T.M. (1999) Molecular cloning and functional expression of O-methyltransferases common to isoquinoline alkaloid and phenylpropanoid biosynthesis. *The Plant Journal* **17**, 329-339.

Gang, D.R. (2005) Evolution of flavors and scents. *Annu. Rev. Plant Biol.* **56**, 301-325.

Gang, D.R., Lavid, N., Zubieta, C., Chen, F., Beuerle, T., Lewinsohn, E., Noel, J.P. and 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.

Gao, T.Y., Jin, X., Tang, W.Z., Wang, X.J. and Zhao, Y.X. (2015) New geranylated flavanones from the fruits of Paulownia catalpifolia Gong Tong with their anti-proliferative activity on lung cancer cells A549. *Bioorg Med Chem Lett* **25**, 3686-3689.

Gasteiger, E., Hoogland, C., Gattiker, A., Wilkins, M.R., Appel, R.D. and Bairoch, A. (2005) Protein identification and analysis tools on the ExPASy server. In: *The proteomics protocols handbook* pp. 571-607. Springer.

Gauthier, A., Gulick, P.J. and Ibrahim, R.K. (1996) cDNA cloning and characterization of a 3′/5′-O-methyltransferase for partially methylated flavonols from Chrysosplenium americanum. *Plant molecular biology* **32**, 1163-1169.

Gauthier, A., Gulick, P.J. and Ibrahim, R.K. (1998) Characterization of two cDNA clones which encode O-methyltransferases for the methylation of both flavonoid and phenylpropanoid compounds. *Archives of biochemistry and biophysics* **351**, 243-249.

Ghisalberti, E.L. (1996) Bioactive acylphloroglucinol derivatives from Eucalyptus species. *Phytochemistry* **41**, 7-22.

Gion, J.-M., Rech, P., Grima-Pettenati, J., Verhaegen, D. and Plomion, C. (2000) Mapping candidate genes in Eucalyptus with emphasis on lignification genes. *Molecular Breeding* **6**, 441-449.

Gomez Roldan, M.V., Outchkourov, N., van Houwelingen, A., Lammers, M., Romero de la Fuente, I., Ziklo, N., Aharoni, A., Hall, R.D. and Beekwilder, J. (2014) An O-methyltransferase modifies accumulation of methylated anthocyanins in seedlings of tomato. *The Plant Journal* **80**, 695-708.

Goodger, J.Q., Heskes, A.M., Mitchell, M.C., King, D.J., Neilson, E.H. and Woodrow, I.E. (2010) Isolation of intact sub-dermal secretory cavities from Eucalyptus. *Plant Methods* **6**, 20.

Goodger, J.Q., Heskes, A.M. and Woodrow, I.E. (2013) Contrasting ontogenetic trajectories for phenolic and terpenoid defences in Eucalyptus froggattii. *Annals of Botany* **112**, 651-659.

Goodger, J.Q., Senaratne, S.L., Nicolle, D. and Woodrow, I.E. (2018) Differential metabolic specialization of foliar oil glands in Eucalyptus brevistylis Brooker (Myrtaceae). *Tree physiology* **38**, 1451-1460.

Goodger, J.Q., Senaratne, S.L., van der Peet, P., Browning, R., Williams, S.J., Nicolle, D. and Woodrow, I.E. (2019) Eucalyptus subgenus Eucalyptus (Myrtaceae) trees are abundant sources of medicinal pinocembrin and related methylated flavanones. *Industrial Crops and Products* **131**, 166-172.

Goodger, J.Q., Seneratne, S.L., Nicolle, D. and Woodrow, I.E. (2016) Foliar Essential Oil Glands of Eucalyptus Subgenus Eucalyptus (Myrtaceae) Are a Rich Source of Flavonoids and Related Non-Volatile Constituents. *PloS one* **11**, e0151432.

Goodger, J.Q. and Woodrow, I.E. (2012) Genetic determinants of oil yield in Eucalyptus polybractea RT Baker. *Trees* **26**, 1951-1956.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R. and Zeng, Q. (2011) Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nature biotechnology* **29**, 644.

Grebenok, R.J., Galbraith, D.W. and Della Penna, D. (1997) Characterization of Zea mays endosperm C-24 sterol methyltransferase: one of two types of sterol methyltransferase in higher plants. *Plant molecular biology* **34**, 891-896.

Green, A.R., Lewis, K.M., Barr, J.T., Jones, J.P., Lu, F., Ralph, J., Vermerris, W., Sattler, S.E. and Kang, C. (2014) Determination of the structure and catalytic mechanism of Sorghum bicolor caffeic acid O-methyltransferase and the structural impact of three brown midrib12 mutations. *Plant physiology* **165**, 1440-1456.

Griffin, A., Burgess, I. and Wolf, L. (1988) Patterns of Natural and Manipulated Hybridisation in the Genus Eucalyptus L'hérit.-1 A Review. *Australian Journal of Botany* **36**, 41-66.

Grosdidier, A., Zoete, V. and Michielin, O. (2007) EADock: docking of small molecules into protein active sites with a multiobjective evolutionary optimization. *Proteins: Structure, Function, and Bioinformatics* **67**, 1010-1025.

Grosdidier, A., Zoete, V. and Michielin, O. (2011) SwissDock, a protein-small molecule docking web service based on EADock DSS. *Nucleic acids research* **39**, W270-W277.

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W. and Gascuel, O. (2010) New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Systematic Biology* **59**, 307-321.

Hagel, J.M. and Facchini, P.J. (2013) Benzylisoquinoline alkaloid metabolism: a century of discovery and a brave new world. *Plant and Cell Physiology* **54**, 647-672.

Hahlbrock, K. and Grisebach, H. (1979) Enzymic controls in the biosynthesis of lignin and flavonoids. *Annual Review of Plant Physiology* **30**, 105-130.

Halarnkar, P.P. and Blomquist, G.J. (1989) Comparative aspects of propionate metabolism. *Comparative biochemistry and physiology. B, Comparative biochemistry* **92**, 227-231.

Han, X., Lamshöft, M., Grobe, N., Ren, X., Fist, A.J., Kutchan, T.M., Spiteller, M. and Zenk, M.H. (2010) The biosynthesis of papaverine proceeds via (S)-reticuline. *Phytochemistry* **71**, 1305-1312.

Harakava, R. (2005) Genes encoding enzymes of the lignin biosynthesis pathway in Eucalyptus. *Genetics and Molecular Biology* **28**, 601-607.

Harborne, J.B. (2001) Twenty-five years of chemical ecology. *Natural product reports* **18**, 361-379.

Hauffe, K.D., Hahlbrock, K. and Scheel, D. (1986) Elicitor-Stimulated Furanocoumarin Biosynthesis in Cultured Parsley Cells: S-Adenosyl-L-Methionine: Bergaptol and S-Adenosyl-

L-Methionine: Xanthotoxol O-Methyltransferases. *Zeitschrift für Naturforschung C* **41**, 228-239.

Hawkins, S., Boudet, A. and Grima-Pettenati, J. (2003) Characterisation of caffeic acid O-methyltransferase and cinnamyl alcohol dehydrogenase gene expression patterns by in situ hybridisation in Eucalyptus gunnii Hook. plantlets. *Plant Science* **164**, 165-173.

He, S.-M., Liang, Y.-L., Cong, K., Chen, G., Zhao, X., Zhao, Q.-M., Zhang, J.-J., Wang, X., Dong, Y. and Yang, J.-L. (2018) Identification and characterization of genes involved in benzylisoquinoline alkaloid biosynthesis in Coptis species. *Frontiers in plant science* **9**, 731.

He, X.-Z. and Dixon, R.A. (1996) Affinity Chromatography, Substrate/Product Specificity, and Amino Acid Sequence Analysis of an IsoflavoneO-Methyltransferase from Alfalfa (Medicago sativaL.). *Archives of biochemistry and biophysics* **336**, 121-129.

He, X.-Z. and Dixon, R.A. (2000) Genetic manipulation of isoflavone 7-O-methyltransferase enhances biosynthesis of 4′-O-methylated isoflavonoid phytoalexins and disease resistance in alfalfa. *The Plant Cell* **12**, 1689-1702.

He, X.-Z., Reddy, J. and Dixon, R.A. (1998) Stress responses in alfalfa (Medicago sativa L). XXII. cDNA cloning and characterization of an elicitor-inducible isoflavone 7-O-methyltransferase. *Plant molecular biology* **36**, 43-54.

Hecht, M., Bromberg, Y. and Rost, B. (2015) Better prediction of functional effects for sequence variants. *BMC genomics* **16**, S1.

Hehmann, M., Lukačin, R., Ekiert, H. and Matern, U. (2004) Furanocoumarin biosynthesis in Ammi majus L. Cloning of bergaptol O-methyltransferase. *European journal of biochemistry* **271**, 932-940.

Heinrich, M. and Teoh, H.L. (2004) Galanthamine from snowdrop—the development of a modern drug against Alzheimer’s disease from local Caucasian knowledge. *Journal of ethnopharmacology* **92**, 147-162.

Hellyer, R. (1964) The structure of agglomerone, a new β-triketone. *Australian Journal of Chemistry* **17**, 1418-1422.

Hellyer, R. (1968) The occurrence of β-triketones in the steam-volatile oils of some myrtaceous Australian plants. *Australian Journal of Chemistry* **21**, 2825-2828.

Heskes, A.M., Goodger, J.Q.D., Tsegay, S., Tim, Q., Williams, S.J. and Woodrow, I.E. (2012) Localization of Oleuropeyl Glucose Esters and a Flavanone to Secretory Cavities of Myrtaceae. *PLoS ONE* **7**, 1-9.

Hirsch, A., Bhuvaneswari, T., Torrey, J. and Bisseling, T. (1989) Early nodulin genes are induced in alfalfa root outgrowths elicited by auxin transport inhibitors. *Proceedings of the National Academy of Sciences* **86**, 1244-1248.

Hotchandani, T., de Villers, J. and Desgagné-Penix, I. (2019) Developmental Regulation of the Expression of Amaryllidaceae Alkaloid Biosynthetic Genes in Narcissus papyraceus. *Genes* **10**, 594.

Huang, T.-S., Anzellotti, D., Dedaldechamp, F. and Ibrahim, R.K. (2004) Partial purification, kinetic analysis, and amino acid sequence information of a flavonol 3-O-methyltransferase from Serratula tinctoria. *Plant physiology* **134**, 1366-1376.

Huang, Y., Komoto, J., Konishi, K., Takata, Y., Ogawa, H., Gomi, T., Fujioka, M. and Takusagawa, F. (2000) Mechanisms for auto-inhibition and forced product release in glycine N-methyltransferase: crystal structures of wild-type, mutant R175K and S-adenosylhomocysteine-bound R175K enzymes. *Journal of molecular biology* **298**, 149-162.

Hugueney, P., Provenzano, S., Verriès, C., Ferrandino, A., Meudec, E., Batelli, G., Merdinoglu, D., Cheynier, V., Schubert, A. and Ageorges, A. (2009) A novel cation-dependent O-methyltransferase involved in anthocyanin methylation in grapevine. *Plant physiology* **150**, 2057-2070.

Husselstein, T., Gachotte, D., Desprez, T., Bard, M. and Benveniste, P. (1996) Transformation of Saccharomyces cerevisiae with a cDNA encoding a sterol C-methyltransferase from Arabidopsis thaliana results in the synthesis of 24-ethyl sterols. *FEBS letters* **381**, 87-92.

Ibdah, M., Zhang, X.H., Schmidt, J. and Vogt, T. (2003) A novel Mg(2+)-dependent O-methyltransferase in the phenylpropanoid metabolism of Mesembryanthemum crystallinum. *J Biol Chem* **278**, 43961-43972.

Ibrahim, R., De Luca, V., Khouri, H., Latchinian, L., Brisson, L. and Charest, P. (1987) Enzymology and compartmentation of polymethylated flavonol glucosides in Chrysosplenium americanum. *Phytochemistry* **26**, 1237-1245.

Ibrahim, R.K., Bruneau, A. and Bantignies, B. (1998) Plant O-methyltransferases: molecular analysis, common signature and classification. *Plant molecular biology* **36**, 1-10.

Ibrahim, R.K. and Muzac, I. (2000) Chapter Eleven - The Methyltransferase Gene Superfamily: A Tree with Multiple Branches. In: *Recent Advances in Phytochemistry* (Romeo, J.T., Ibrahim, R., Varin, L. and De Luca, V. eds), pp. 349-384. Elsevier.

Igarashi, K. and Ohmuma, M. (1995) Effects of isorhamnetin, rhamnetin, and quercetin on the concentrations of cholesterol and lipoperoxide in the serum and liver and on the blood and liver antioxidative enzyme activities of rats. *Bioscience, biotechnology, and biochemistry* **59**, 595-601.

Itoh, N., Iwata, C. and Toda, H. (2016) Molecular cloning and characterization of a flavonoid-O-methyltransferase with broad substrate specificity and regioselectivity from Citrus depressa. *BMC Plant Biol* **16**, 180.

Jacobs, S.A., Harp, J.M., Devarakonda, S., Kim, Y., Rastinejad, F. and Khorasanizadeh, S. (2002) The active site of the SET domain is constructed on a knot. *Nature Structural & Molecular Biology* **9**, 833.

Jarrett, J.T., Huang, S. and Matthews, R.G. (1998) Methionine synthase exists in two distinct conformations that differ in reactivity toward methyltetrahydrofolate, adenosylmethionine, and flavodoxin. *Biochemistry* **37**, 5372-5382.

Ji, X., Luo, X., Zhang, J. and Huang, D. (2021) Effects of exogenous vitamin B 12 on nutrient removal and protein expression of algal-bacterial consortium. *Environmental Science and Pollution Research* **28**, 15954-15965.

Jones, T.G.H. and Lahey, F.N. (1939) The constitution and synthesis of conglomerone.

Joshi, C.P. and Chiang, V.L. (1998) Conserved sequence motifs in plant S-adenosyl-L-methionine-dependent methyltransferases. *Plant molecular biology* **37**, 663-674.

Ka, S., Masi, M., Merindol, N., Di Lecce, R., Plourde, M.B., Seck, M., Górecki, M., Pescitelli, G., Desgagne-Penix, I. and Evidente, A. (2020) Gigantelline, gigantellinine and gigancrinine, cherylline-and crinine-type alkaloids isolated from Crinum jagus with anti-acetylcholinesterase activity. *Phytochemistry* **175**, 112390.

Kanehisa, M. and Goto, S. (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic acids research* **28**, 27-30.

Kang, J., Kim, E., Kim, W., Seong, K.M., Youn, H., Kim, J.W., Kim, J. and Youn, B. (2013) Rhamnetin and cirsiliol induce radiosensitization and inhibition of epithelial-mesenchymal transition (EMT) by miR-34a-mediated suppression of Notch-1 expression in non-small cell lung cancer cell lines. *Journal of Biological Chemistry* **288**, 27343-27357.

Kapoor, S. (2013) Comment on isolation and identification of compounds from Penthorum chinense Pursh with antioxidant and antihepatocarcinoma properties: pinocembrin and its rapidly emerging neuroprotective effects. *Journal Of Agricultural And Food Chemistry* **61**, 1416-1416.

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

Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N. and Sternberg, M.J.E. (2015) The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols* **10**, 845-858.

Keszei, A., Brubaker, C.L. and Foley, W.J. (2008) A molecular perspective on terpene variation in Australian Myrtaceae. *Australian Journal of Botany* **56**, 197-213.

Khouri, H.E. and Ibrahim, R.K. (1987) Resolution of five position-specific flavonoid o-methyl-transferases by fast protein liquid chromatofocusing. *Journal of Chromatography A* **407**, 291-297.

Kilgore, M.B., Augustin, M.M., Starks, C.M., O'Neil-Johnson, M., May, G.D., Crow, J.A. and Kutchan, T.M. (2014) Cloning and characterization of a norbelladine 4′-O-methyltransferase involved in the biosynthesis of the Alzheimer's drug galanthamine in Narcissus sp. aff. pseudonarcissus. *PloS one* **9**.

Killeen, D.P., van Klink, J.W., Smallfield, B.M., Gordon, K.C. and Perry, N.B. (2015)

Herbicidal β-triketones are compartmentalized in leaves of L eptospermum species: localization by R aman microscopy and rapid screening. *New Phytologist* **205**, 339-349.

Kim, B.-G., Lee, Y.J., Lee, S., Lim, Y., Cheong, Y. and Ahn, J.-H. (2008) Altered regioselectivity of a poplar O-methyltransferase, POMT-7. *Journal of biotechnology* **138**, 107-111.

Kim, B.-G., Sung, S.H., Chong, Y., Lim, Y. and Ahn, J.-H. (2010) Plant flavonoid O-methyltransferases: substrate specificity and application. *Journal of Plant Biology* **53**, 321-329.

Kim, B.G., Lee, Y., Hur, H.G., Lim, Y. and Ahn, J.H. (2006) Flavonoid 3'-O-methyltransferase from rice: cDNA cloning, characterization and functional expression. *Phytochemistry* **67**, 387-394.

Kim, D.H., Kim, B.G., Lee, Y., Ryu, J.Y., Lim, Y., Hur, H.G. and Ahn, J.H. (2005) Regiospecific methylation of naringenin to ponciretin by soybean O-methyltransferase expressed in Escherichia coli. *J Biotechnol* **119**, 155-162.

Klimasauskas, S. and Lukinavicius, G. (2007) AdoMet-Dependent Methyltransferases, Chemistry of. *Wiley Encyclopedia of Chemical Biology*, 1-10.

Knogge, W. and Weissenbock, G. (1984) Purification, characterization, and kinetic mechanism of S-adenosyl-l-methionine: vitexin 2 ″-O-rhamnoside 7-O-methyltransferase of Avena sativa L. *European journal of biochemistry* **140**, 113-118.

Koch, M., Lemke, R., Heise, K.P. and Mock, H.P. (2003) Characterization of γ-tocopherol methyltransferases from Capsicum annuum L and Arabidopsis thaliana. *European journal of biochemistry* **270**, 84-92.

Koeduka, T. (2014) The phenylpropene synthase pathway and its applications in the engineering of volatile phenylpropanoids in plants. *Plant Biotechnology*, 14.0801 a.

Koeduka, T., Baiga, T.J., Noel, J.P. and Pichersky, E. (2009) Biosynthesis of t-anethole in anise: characterization of t-anol/isoeugenol synthase and an O-methyltransferase specific for a C7-C8 propenyl side chain. *Plant physiology* **149**, 384-394.

Koirala, N., Pandey, R.P., Parajuli, P., Jung, H.J. and Sohng, J.K. (2014) Methylation and subsequent glycosylation of 7, 8-dihydroxyflavone. *Journal of biotechnology* **184**, 128-137.

Koirala, N., Thuan, N.H., Ghimire, G.P., Thang, D.V. and Sohng, J.K. (2016) Methylation of flavonoids: Chemical structures, bioactivities, progress and perspectives for biotechnological production. *Enzyme Microb Technol* **86**, 103-116.

Kopycki, J.G., Rauh, D., Chumanevich, A.A., Neumann, P., Vogt, T. and Stubbs, M.T. (2008) Biochemical and structural analysis of substrate promiscuity in plant Mg2+-dependent O-methyltransferases. *Journal of molecular biology* **378**, 154-164.

Kozbial, P.Z. and Mushegian, A.R. (2005) Natural history of S-adenosylmethionine-binding proteins. *BMC structural biology* **5**, 19.

Krogh, A., Larsson, B., Von Heijne, G. and Sonnhammer, E.L. (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *Journal of molecular biology* **305**, 567-580.

Kumar, S. and Pandey, A.K. (2013) Chemistry and biological activities of flavonoids: an overview. *ScientificWorldJournal* **2013**, 162750.

Kuo, Y.-C., Yang, L.-M. and Lin, L.-C. (2004) Isolation and immunomodulatory effect of flavonoids from Syzygium samarangense. *Planta medica* **70**, 1237-1239.

Labib, R.M., Malak, L.G., Youssef, F.S. and Ross, S.A. (2020) A new stilbene from Agonis flexuosa leaves and verification of its histamine release inhibitory activity using in silico and in vitro studies. *South African Journal of Botany* **135**, 384-390.

Lam, K.C., Ibrahim, R.K., Behdad, B. and Dayanandan, S. (2007) Structure, function, and evolution of plant O-methyltransferases. *Genome* **50**, 1001-1013.

Lang, D.E., Morris, J.S., Rowley, M., Torres, M.A., Maksimovich, V.A., Facchini, P.J. and Ng, K.K. (2019) Structure–function studies of tetrahydroprotoberberine N-methyltransferase reveal the molecular basis of stereoselective substrate recognition. *Journal of Biological Chemistry* **294**, 14482-14498.

Lavid, N., Wang, J., Shalit, M., Guterman, I., Bar, E., Beuerle, T., Menda, N., Shafir, S., Zamir, D. and Adam, Z. (2002) O-methyltransferases involved in the biosynthesis of volatile phenolic derivatives in rose petals. *Plant physiology* **129**, 1899-1907.

Lee, Y.J., Kim, B.G., Chong, Y., Lim, Y. and Ahn, J.-H. (2008) Cation dependent O-methyltransferases from rice. *Planta* **227**, 641-647.

Levac, D., Murata, J., Kim, W.S. and De Luca, V. (2008) Application of carborundum abrasion for investigating the leaf epidermis: molecular cloning of Catharanthus roseus 16-hydroxytabersonine-16-O-methyltransferase. *The Plant Journal* **53**, 225-236.

Li, B. and Dewey, C.N. (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC bioinformatics* **12**, 323.

Li, B., Fillmore, N., Bai, Y., Collins, M., Thomson, J.A., Stewart, R. and Dewey, C.N. (2014) Evaluation of de novo transcriptome assemblies from RNA-Seq data. *Genome biology* **15**, 553.

Li, W., Qiao, C., Pang, J., Zhang, G. and Luo, Y. (2019) The versatile O-methyltransferase LrOMT catalyzes multiple O-methylation reactions in amaryllidaceae alkaloids biosynthesis. *International journal of biological macromolecules* **141**, 680-692.

Liscombe, D.K. and Facchini, P.J. (2007) Molecular cloning and characterization of tetrahydroprotoberberine cis-N-methyltransferase, an enzyme involved in alkaloid biosynthesis in opium poppy. *Journal of Biological Chemistry* **282**, 14741-14751.

Liscombe, D.K., Louie, G.V. and Noel, J.P. (2012) Architectures, mechanisms and molecular evolution of natural product methyltransferases. *Natural product reports* **29**, 1238-1250.

Liscombe, D.K. and O’Connor, S.E. (2011) A virus-induced gene silencing approach to understanding alkaloid metabolism in Catharanthus roseus. *Phytochemistry* **72**, 1969-1977.

Liscombe, D.K., Usera, A.R. and O’Connor, S.E. (2010) Homolog of tocopherol C methyltransferases catalyzes N methylation in anticancer alkaloid biosynthesis. *Proceedings of the National Academy of Sciences* **107**, 18793-18798.

Liu, C.-J., Deavours, B.E., Richard, S.B., Ferrer, J.-L., Blount, J.W., Huhman, D., Dixon, R.A. and Noel, J.P. (2006) Structural basis for dual functionality of isoflavonoid O-methyltransferases in the evolution of plant defense responses. *The plant cell* **18**, 3656-3669.

Liu, C.-J. and Dixon, R.A. (2001) Elicitor-induced association of isoflavone O-methyltransferase with endomembranes prevents the formation and 7-O-methylation of daidzein during isoflavonoid phytoalexin biosynthesis. *The Plant Cell* **13**, 2643-2658.

Liu, H., Xu, R.-X., Zhang, X.-S., Zhu, T.-T., Lou, H.-X. and Cheng, A.-X. (2019) The identification and functional characterization of three liverwort class I O-methyltransferases. *Phytochemistry* **159**, 190-198.

Liu, X., Wang, Y., Chen, Y., Xu, S., Gong, Q., Zhao, C., Cao, J. and Sun, C. (2020) Characterization of a Flavonoid 3'/5'/7-O-Methyltransferase from Citrus reticulata and Evaluation of the In Vitro Cytotoxicity of Its Methylated Products. *Molecules* **25**, 858.

Lois, R. (1994) Accumulation of UV-absorbing flavonoids induced by UV-B radiation in Arabidopsis thaliana L. *Planta* **194**, 498-503.

Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P.M. and Henrissat, B. (2013) The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic acids research* **42**, D490-D495.

Louie, G.V., Bowman, M.E., Tu, Y., Mouradov, A., Spangenberg, G. and Noel, J.P. (2010) Structure-function analyses of a caffeic acid O-methyltransferase from perennial ryegrass reveal the molecular basis for substrate preference. *Plant Cell* **22**, 4114-4127.

Lovell, S.C., Davis, I.W., Arendall III, W.B., De Bakker, P.I., Word, J.M., Prisant, M.G., Richardson, J.S. and Richardson, D.C. (2003) Structure validation by Cα geometry: ϕ, ψ and Cβ deviation. *Proteins: Structure, Function, and Bioinformatics* **50**, 437-450.

Lu, S.C. (2000) S-adenosylmethionine. *The international journal of biochemistry & cell biology* **32**, 391-395.

Lücker, J., Martens, S. and Lund, S.T. (2010) Characterization of a Vitis vinifera cv. Cabernet Sauvignon 3′, 5′-O-methyltransferase showing strong preference for anthocyanins and glycosylated flavonols. *Phytochemistry* **71**, 1474-1484.

lwashina, T. (2000) structure and distribution of flavanoids in plants. *Journal of Plant Research* **113**.

Maciel, M., Morais, S., Bevilaqua, C., Silva, R., Barros, R., Sousa, R., Sousa, L., Brito, E. and

Souza-Neto, M. (2010) Chemical composition of Eucalyptus spp. essential oils and their insecticidal effects on Lutzomyia longipalpis. *Veterinary parasitology* **167**, 1-7.

Malterud, K.E., Diep, O.H. and Sund, R.B. (1996) C-Methylated Dihydrochalcones from Myrica gale L: Effects as Antioxidants and as Scavengers of 1, 1-Diphenyl-2-Picrylhydrazyl. *Basic & Clinical Pharmacology & Toxicology* **78**, 111-116.

Marsh, K.J., Saraf, I., Hocart, C.H., Youngentob, K., Singh, I.-P. and Foley, W.J. (2019) Occurrence and distribution of unsubstituted B-ring flavanones in Eucalyptus foliage. *Phytochemistry* **160**, 31-39.

Martin, J.L. and McMillan, F.M. (2002) SAM (dependent) I AM: the S-adenosylmethionine-dependent methyltransferase fold. *Current opinion in structural biology* **12**, 783-793.

Massaro, C.F., Katouli, M., Grkovic, T., Vu, H., Quinn, R.J., Heard, T.A., Carvalho, C., Manley-Harris, M., Wallace, H.M. and Brooks, P. (2014) Anti-staphylococcal activity of C-methyl flavanones from propolis of Australian stingless bees (Tetragonula carbonaria) and fruit resins of Corymbia torelliana (Myrtaceae). *Fitoterapia* **95**, 247-257.

Mathesius, U., Schlaman, H.R., Spaink, H.P., Of Sautter, C., Rolfe, B.G. and Djordjevic, M.A. (1998) Auxin transport inhibition precedes root nodule formation in white clover roots and is regulated by flavonoids and derivatives of chitin oligosaccharides. *The Plant Journal* **14**, 23-34.

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

Medina-Holguín, A.L., Holguín, F.O., Micheletto, S., Goehle, S., Simon, J.A. and O'Connell, M.A. (2008) Chemotypic variation of essential oils in the medicinal plant, Anemopsis californica. *Phytochemistry* **69**, 919-927.

Michaelis, L. and Menten, M.L. (2007) *Die kinetik der invertinwirkung*:Universitätsbibliothek Johann Christian Senckenberg.

Michel, G., Sauvé, V., Larocque, R., Li, Y., Matte, A. and Cygler, M. (2002) The structure of the RlmB 23S rRNA methyltransferase reveals a new methyltransferase fold with a unique knot. *Structure* **10**, 1303-1315.

Miler, M., Zivanovic, J., Ajdzanovic, V., Orescanin-Dusic, Z., Milenkovic, D., Konic-Ristic, A., Blagojevic, D., Milosevic, V. and Sosic-Jurjevic, B. (2016) Citrus flavanones naringenin and hesperetin improve antioxidant status and membrane lipid compositions in the liver of old-aged Wistar rats. *Exp Gerontol* **84**, 49-60.

Misako, K. and Kouichi, M. (2004) Caffeine synthase and related methyltransferases in plants. *Frontiers in bioscience: a journal and virtual library* **9**, 1833-1842.

Mizrachi, E., Hefer, C.A., Ranik, M., Joubert, F. and Myburg, A.A. (2010) De novo assembled expressed gene catalog of a fast-growing Eucalyptus tree produced by Illumina mRNA-Seq. *BMC genomics* **11**, 1-12.

Mo, Y., Nagel, C. and Taylor, L.P. (1992) Biochemical complementation of chalcone synthase mutants defines a role for flavonols in functional pollen. *Proceedings of the National Academy of Sciences* **89**, 7213-7217.

Morishige, T., Tsujita, T., Yamada, Y. and Sato, F. (2000) Molecular characterization of theS-adenosyl-L-methionine: 3′-hydroxy-N-methylcoclaurine 4′-O-methyltransferase involved in isoquinoline alkaloid biosynthesis in Coptis Japonica. *Journal of Biological Chemistry* **275**, 23398-23405.

Morris, J.S. and Facchini, P. (2019) Corrigendum: Molecular Origins of Functional Diversity in Benzylisoquinoline Alkaloid Methyltransferases. *Frontiers in Plant Science* **10**, 1436.

Murata, J. and Luca, V.D. (2005) Localization of tabersonine 16-hydroxylase and 16-OH tabersonine-16-O-methyltransferase to leaf epidermal cells defines them as a major site of precursor biosynthesis in the vindoline pathway in Catharanthus roseus. *The Plant Journal* **44**, 581-594.

Murata, J., Roepke, J., Gordon, H. and 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. 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.

Muzac, I., Wang, J., Anzellotti, D., Zhang, H. and Ibrahim, R.K. (2000) Functional expression of an Arabidopsis cDNA clone encoding a flavonol 3′-O-methyltransferase and characterization of the gene product. *Archives of biochemistry and biophysics* **375**, 385-388.

Myburg, A.A., Grattapaglia, D., Tuskan, G.A., Hellsten, U., Hayes, R.D., Grimwood, J., Jenkins, J., Lindquist, E., Tice, H. and Bauer, D. (2014) The genome of Eucalyptus grandis. *Nature* **510**, 356-362.

Nagel, J., Culley, L.K., Lu, Y., Liu, E., Matthews, P.D., Stevens, J.F. and Page, J.E. (2008) EST analysis of hop glandular trichomes identifies an O-methyltransferase that catalyzes the biosynthesis of xanthohumol. *The Plant Cell* **20**, 186-200.

Negre, F., Kolosova, N., Knoll, J., Kish, C.M. and 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.

Nicolet, Y. (2020) Structure–function relationships of radical SAM enzymes. *Nature Catalysis* **3**, 337-350.

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

Nomura, T. and Kutchan, T.M. (2010) Three new O-methyltransferases are sufficient for all O-methylation reactions of ipecac alkaloid biosynthesis in root culture of Psychotria ipecacuanha. *Journal of Biological Chemistry* **285**, 7722-7738.

Nugent, T. and Jones, D.T. (2012) Detecting pore-lining regions in transmembrane protein sequences. *BMC bioinformatics* **13**, 169.

Oates, C.N., Külheim, C., Myburg, A.A., Slippers, B. and Naidoo, S. (2015) The transcriptome and terpene profile of Eucalyptus grandis reveals mechanisms of defense against the insect pest, Leptocybe invasa. *Plant and Cell Physiology* **56**, 1418-1428.

Oguro, S., Akashi, T., Ayabe, S.-i., Noguchi, H. and Abe, I. (2004) Probing biosynthesis of plant polyketides with synthetic N-acetylcysteamine thioesters. *Biochemical and biophysical* *research communications* **325**, 561-567.

Ohkatsu, Y., Sakurai, T. and Sato, T. (2010) Relationship between Chemical Structure and Antioxidant Function of Flavonoids. *Journal of the Japan Petroleum Institute* **53**, 213-221.

Ohno, S. (1999) Gene duplication and the uniqueness of vertebrate genomes circa 1970–1999. In: *Seminars in cell & developmental biology* pp. 517-522. Elsevier.

Oikawa, N., Nobushi, Y., Wada, T., Sonoda, K., Okazaki, Y., Tsutsumi, S., Park, Y.K., Kurokawa, M., Shimba, S. and Yasukawa, K. (2016) Inhibitory effects of compounds isolated from the dried branches and leaves of murta (Myrceugenia euosma) on lipid accumulation in 3T3-L1 cells. *Journal of Natural Medicines* **70**, 502-509.

Oliveira, M. and Pais, M. (1990) Glandular trichomes of Humulus lupulus var. Brewer's gold (hops): Ultrastructural aspects of peltate trichomes. *Journal of submicroscopic cytology and pathology* **22**, 241-248.

Otake, Y., Hsieh, F. and Walle, T. (2002) Glucoronidation versus oxidation of the flavonoid galangin by human liver microsomes and hepatocytes. *The American Society for Pharmacology and Experimental Therapeutics* **30**.

Ouchemoukh, S., Amessis-Ouchemoukh, N., Gómez-Romero, M., Aboud, F., Giuseppe, A., Fernández-Gutiérrez, A. and Segura-Carretero, A. (2016) Characterisation of phenolic compounds in Algerian honeys by RP-HPLC coupled to electrospray time-of-flight mass spectrometry. *LWT - Food Science and Technology*.

Ounaroon, A., Decker, G., Schmidt, J., Lottspeich, F. and Kutchan, T.M. (2003) (R, S)-Reticuline 7-O-methyltransferase and (R, S)-norcoclaurine 6-O-methyltransferase of Papaver somniferum–cDNA cloning and characterization of methyl transfer enzymes of alkaloid biosynthesis in opium poppy. *The plant journal* **36**, 808-819.

Paizs, C., Diemer, T. and Rétey, J. (2008) The putative coenzyme B12-dependent methylmalonyl-CoA mutase from potatoes is a phosphatase. *Bioorganic chemistry* **36**, 261-264.

Pak, F., Gropper, S., Dai, W., Havkin-Frenkel, D. and Belanger, F. (2004) Characterization of a multifunctional methyltransferase from the orchid Vanilla planifolia. *Plant cell reports* **22**, 959-966.

Panche, A., Diwan, A. and Chandra, S. (2016) Flavonoids: an overview. *Journal of nutritional science* **5**.

Park, M.R., Chen, X., Lang, D.E., Ng, K.K. and Facchini, P.J. (2018) Heterodimeric O-methyltransferases involved in the biosynthesis of noscapine in opium poppy. *The Plant Journal* **95**, 252-267.

Patel, N.K., Jaiswal, G. and Bhutani, K.K. (2016) A review on biological sources, chemistry and pharmacological activities of pinostrobin. *Nat Prod Res* **30**, 2017-2027.

Paux, E., Tamasloukht, M.B., Ladouce, N., Sivadon, P. and Grima-Pettenati, J. (2004) Identification of genes preferentially expressed during wood formation in Eucalyptus. *Plant molecular biology* **55**, 263-280.

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

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C. and Ferrin, T.E. (2004) UCSF Chimera—a visualization system for exploratory research and analysis. *Journal of computational chemistry* **25**, 1605-1612.

Phillipson, J.D., Roberts, M.F. and Zenk, M. (2012) *The chemistry and biology of isoquinoline alkaloids*:Springer Science & Business Media.

Pichersky, E. and Gang, D.R. (2000) Genetics and biochemistry of secondary metabolites in plants: an evolutionary perspective. *Trends in Plant Science* **5**, 439-445.

Piechulla, B., Magnus, N., Lemfack, M.C. and von Reuss, S. (2020) Terpenoid Cyclization by SAM-Dependent C-Methyl Transferase. *Trends in Chemistry* **2**, 585-586.

Poeydomenge, O., Boudet, A.M. and Grima-Pettenati, J. (1994) A cDNA encoding S-adenosyl-L-methionine: caffeic acid 3-O-methyltransferase from Eucalyptus. *Plant physiology* **105**, 749.

Poston, J.M. (1978) Coenzyme B12-dependent enzymes in potatoes: leucine 2, 3-aminomutase and methylmalonyl-coa mutase. *Phytochemistry* **17**, 401-402.

Pott, M.B., Hippauf, F., Saschenbrecker, S., Chen, F., Ross, J., Kiefer, I., Slusarenko, A., Noel, J.P., Pichersky, E. and 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.

Preisig, C.L., Matthews, D.E. and VanEtten, H.D. (1989) Purification and characterization of S-adenosyl-L-methionine: 6a-hydroxymaackiain 3-O-methyltransferase from Pisum sativum. *Plant physiology* **91**, 559-566.

Priyavathi, P., Nagesh, S., Kavitha, V., Johnson, C. and Gopal, P. (2018) Comprehensive Leaf Transcriptome of a Non-model Plant, Abelmoschus esculentus for the Functional Genomics Studies. *J Genet Genome Res* **5**, 036.

Qiao, X., Ye, G., Liu, C.-F., Zhang, Z.-X., Tu, Q., Dong, J., Li, Y.-Q., Guo, D.-A. and Ye, M. (2012) Chemical analysis of Eriocaulon buergerianum and adulterating species by high-performance liquid chromatography with diode array detection and electrospray ionization tandem mass spectrometry. *Journal of pharmaceutical and biomedical analysis* **57**, 133-142.

Rahmat, E. and Kang, Y. (2019) Adventitious root culture for secondary metabolite production in medicinal plants: A Review. *Journal of Plant Biotechnology* **46**, 143-157.

Rakwal, R., Agrawal, G.K., Yonekura, M. and Kodama, O. (2000) Naringenin 7-O-methyltransferase involved in the biosynthesis of the flavanone phytoalexin sakuranetin from rice (Oryza sativa L.). *Plant Science* **155**, 213-221.

Ramachandran, G.N., Ramakrishnan, C. and Sasisekharan, V. (1963) Stereochemistry of polypeptide chain configurations. *Journal of Molecular Biology* **7**, 95.

Rao, S.T. and Rossmann, M.G. (1973) Comparison of super-secondary structures in proteins. *Journal of molecular biology* **76**, 241-256.

Reikofski, J. and Tao, B.Y. (1992) Polymerase chain reaction (PCR) techniques for site-directed mutagenesis. *Biotechnology advances* **10**, 535-547.

Reis, A., Magne, K., Massot, S., Tallini, L.R., Scopel, M., Bastida, J., Ratet, P. and Zuanazzi, J.A. (2019) Amaryllidaceae alkaloids: identification and partial characterization of montanine production in Rhodophiala bifida plant. *Scientific reports* **9**, 1-11.

Robertson, G., Schein, J., Chiu, R., Corbett, R., Field, M., Jackman, S.D., Mungall, K., Lee, S., Okada, H.M. and Qian, J.Q. (2010) De novo assembly and analysis of RNA-seq data. *Nature methods* **7**, 909-912.

Robin, A.Y., Giustini, C., Graindorge, M., Matringe, M. and Dumas, R. (2016) Crystal structure of norcoclaurine-6-O-methyltransferase, a key rate-limiting step in the synthesis of benzylisoquinoline alkaloids. *The Plant Journal* **87**, 641-653.

Ross, J.R., Nam, K.H., D'Auria, J.C. and 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.

Roy, A., Kucukural, A. and Zhang, Y. (2010) I-TASSER: a unified platform for automated protein structure and function prediction. *Nature protocols* **5**, 725.

Rueffer, M. and Zenk, M. (1985) Berberine synthase, the methylenedioxy group forming enzyme inberberine synthesis. *Tetrahedron letters* **26**, 201-202.

Salim, V., Jones, A.D. and DellaPenna, D. (2018) Camptotheca acuminata 10-hydroxycamptothecin O-methyltransferase: an alkaloid biosynthetic enzyme co-opted from flavonoid metabolism. *The Plant Journal* **95**, 112-125.

Salzman, R., Fujita, T., Zhu-Salzman, K., Hasegawa, P. and Bressan, R. (1999) An improved RNA isolation method for plant tissues containing high levels of phenolic compounds or carbohydrates. *Plant Molecular Biology Reporter* **17**, 11-17.

Santos, S.A., Vidigal, P.M., Guimarães, L.M., Mafia, R.G., Templeton, M.D. and Alfenas, A.C. (2020) Transcriptome analysis of Eucalyptus grandis genotypes reveals constitutive overexpression of genes related to rust (Austropuccinia psidii) resistance. *Plant Molecular Biology*, 1-19.

Saslowsky, D.E., Warek, U. and Winkel, B.S. (2005) Nuclear localization of flavonoid enzymes in Arabidopsis. *Journal of Biological Chemistry* **280**, 23735-23740.

Sato, F., Takeshita, N., Fitchen, J.H., Fujiwara, H. and Yamada, Y. (1993) S-Adenosyl-L-methionine: scoulerine-9-O-methyltransferase from cultured Coptis japonica cells. *Phytochemistry* **32**, 659-664.

Scheid, F., Rohmer, M. and Benveniste, P. (1982) Biosynthesis of Δ5. 23 sterols in etiolated coleoptiles from Zea mays. *Phytochemistry* **21**, 1959-1967.

Schmidt, A., Li, C., Jones, A.D. and Pichersky, E. (2012) Characterization of a flavonol 3-O-methyltransferase in the trichomes of the wild tomato species Solanum habrochaites. *Planta* **236**, 839-849.

Schröder, G., Wehinger, E., Lukačin, R., Wellmann, F., Seefelder, W., Schwab, W. and Schröder, J. (2004) Flavonoid methylation: a novel 4′-O-methyltransferase from Catharanthus roseus, and evidence that partially methylated flavanones are substrates of four different flavonoid dioxygenases. *Phytochemistry* **65**, 1085-1094.

Schröder, J., Raiber, S., Berger, T., Schmidt, A., Schmidt, J., Soares-Sello, A.M., Bardshiri, E., Strack, D., Simpson, T.J. and Veit, M. (1998) Plant polyketide synthases: a chalcone synthase-type enzyme which performs a condensation reaction with methylmalonyl-CoA in the biosynthesis of C-methylated chalcones. *Biochemistry* **37**, 8417-8425.

Schubert, H.L., Blumenthal, R.M. and Cheng, X. (2003) Many paths to methyltransfer: a chronicle of convergence. *Trends in biochemical sciences* **28**, 329-335.

Schubert, H.L., Wilson, K.S., Raux, E., Woodcock, S.C. and Warren, M.J. (1998) The X-ray structure of a cobalamin biosynthetic enzyme, cobalt-precorrin-4 methyltransferase. *Nature Structural & Molecular Biology* **5**, 585.

Senaratne, S.L. (2020) Non-volatile secondary metabolites in foliar oil glands of Eucalyptus species.

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.

Shi, J., Gonzales, R.A. and Bhattacharyya, M.K. (1996a) Identification and Characterization of an S-Adenosyl-L-methionine:-Sterol-C-methyltransferase cDNA from Soybean. *Journal of*

Biological Chemistry **271**, 9384-9389.

Shi, J., Gonzales, R.A. and Bhattacharyya, M.K. (1996b) Identification and characterization of an S-adenosyl-L-methionine: Δ24-sterol-C-methyltransferase cDNA from soybean. *Journal of Biological Chemistry* **271**, 9384-9389.

Shimizu, T., Lin, F., Hasegawa, M., Okada, K., Nojiri, H. and Yamane, H. (2012) Purification and identification of naringenin 7-O-methyltransferase, a key enzyme in biosynthesis of flavonoid phytoalexin sakuranetin in rice. *Journal of Biological Chemistry* **287**, 19315-19325.

Sievers, F. and Higgins, D.G. (2018) Clustal Omega for making accurate alignments of many protein sequences. *Protein Science* **27**, 135-145.

Simirgiotis, M.J., Benites, J., Areche, C. and Sepúlveda, B. (2015) Antioxidant capacities and analysis of phenolic compounds in three endemic Nolana species by HPLC-PDA-ESI-MS. *Molecules* **20**, 11490-11507.

Singh, A., Menéndez-Perdomo, I.M. and Facchini, P.J. (2019) Benzylisoquinoline alkaloid biosynthesis in opium poppy: an update. *Phytochemistry Reviews* **18**, 1457-1482.

Sirikantaramas, S., Yamazaki, M. and Saito, K. (2008) Mechanisms of resistance to self-produced toxic secondary metabolites in plants. *Phytochemistry Reviews* **7**, 467.

Storm, P.A., Herbst, D.A., Maier, T. and Townsend, C.A. (2017) Functional and structural analysis of programmed C-methylation in the biosynthesis of the fungal polyketide citrinin. *Cell chemical biology* **24**, 316-325.

Struck, A.W., Thompson, M.L., Wong, L.S. and Micklefield, J. (2012) S-Adenosyl-methionine-dependent methyltransferases: highly versatile enzymes in biocatalysis, biosynthesis and other biotechnological applications. *ChemBioChem* **13**, 2642-2655.

Sudsai, T., Prabpai, S., Kongsaeree, P., Wattanapiromsakul, C. and Tewtrakul, S. (2014) Anti-inflammatory activity of compounds from Boesenbergia longiflora rhizomes. *J Ethnopharmacol* **154**, 453-461.

Sun, B., Wang, P., Wang, R., Li, Y. and Xu, S. (2018) Molecular Cloning and Characterization of a meta/para-O-Methyltransferase from Lycoris aurea. *International journal of molecular* *sciences* **19**, 1911.

Sun, X., Zhang, Y., Chen, S. and Fu, Y. (2016) Characterization and identification of the chemical constituents in the root of Lindera reflexa Hemsl. using ultra-high performance liquid chromatography coupled with linear trap quadrupole orbitrap mass spectrometry. *J Pharm Biomed Anal* **126**, 34-47.

Sutthivaiyakit, S., Seeka, C., Wetprasit, N. and Sutthivaiyakit, P. (2013) C-methylated flavonoids from Pisonia grandis roots. *Phytochemistry Letters* **6**, 407-411.

Suzuki, Y., Hibino, T., Kawazu, T., Wada, T., Kihara, T. and Koyama, H. (2003) Extraction of total RNA from leaves of Eucalyptus and other woody and herbaceous plants using sodium isoascorbate. *Biotechniques* **34**, 988-993.

Ta, N. and Walle, T. (2007) Aromatase inhibition by bioavailable methylated flavones. *The Journal of steroid biochemistry and molecular biology* **107**, 127-129.

Takahashi-Iñiguez, T., García-Hernandez, E., Arreguín-Espinosa, R. and Flores, M.E. (2012) Role of vitamin B12 on methylmalonyl-CoA mutase activity. *Journal of Zhejiang University SCIENCE B* **13**, 423-437.

Tang, Q., Vianney, Y.M., Weisz, K., Grathwol, C.W., Link, A., Bornscheuer, U.T. and Pavlidis, I.V. (2020) Influence of substrate binding residues on the substrate scope and regioselectivity of a plant O-methyltransferase against flavonoids. *ChemCatChem*.

Taylor, L. and Jorgensen, R. (1992) Conditional male fertility in chalcone synthase-deficient petunia. *Journal of Heredity* **83**, 11-17.

Tengg, M., Stecher, H., Remler, P., Eiteljörg, I., Schwab, H. and Gruber-Khadjawi, M. (2012) Molecular characterization of the C-methyltransferase NovO of Streptomyces spheroides, a valuable enzyme for performing Friedel–Crafts alkylation. *Journal of Molecular Catalysis B: Enzymatic* **84**, 2-8.

Thavamanikumar, S., Southerton, S. and Thumma, B. (2014) RNA-Seq using two populations reveals genes and alleles controlling wood traits and growth in Eucalyptus nitens. *PLoS One* **9**, e101104.

Thoison, O., Sévenet, T., Niemeyer, H.M. and Russell, G.B. (2004) Insect antifeedant compounds from Nothofagus dombeyi and N. pumilio. *Phytochemistry* **65**, 2173-2176.

Torres, M.A., Hoffarth, E., Eugenio, L., Savtchouk, J., Chen, X., Morris, J.S., Facchini, P.J. and Ng, K.K.-S. (2016) Structural and functional studies of pavine N-methyltransferase from Thalictrum flavum reveal novel insights into substrate recognition and catalytic mechanism. *Journal of Biological Chemistry* **291**, 23403-23415.

Trott, O. and Olson, A.J. (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of computational chemistry* **31**, 455-461.

Tsai, C.J., Harding, S.A., Tschaplinski, T.J., Lindroth, R.L. and Yuan, Y. (2006) Genome-wide analysis of the structural genes regulating defense phenylpropanoid metabolism in Populus. *New Phytologist* **172**, 47-62.

Tuchinda, P., Reutrakul, V., Claeson, P., Pongprayoon, U., Sematong, T., Santisuk, T. and Taylor, W.C. (2002) Anti-inflammatory cyclohexenyl chalcone derivatives in Boesenbergia pandurata. *Phytochemistry* **59**, 169-173.

Tucker, D.J., Wallis, I.R., Bolton, J.M., Marsh, K.J., Rosser, A.A., Brereton, I.M., Nicolle, D. and Foley, W.J. (2010) A metabolomic approach to identifying chemical mediators of mammal-plant interactions. *J Chem Ecol* **36**, 727-735.

Untergasser, A. (2008) RNA miniprep using CTAB. *Untergasser's Lab. Summer*.

Valentic, T.R., Payne, J.T. and Smolke, C.D. (2020) Structure-Guided Engineering of a Scoulerine 9-O-Methyltransferase Enables the Biosynthesis of Tetrahydropalmatrubine and Tetrahydropalmatine in Yeast. *ACS Catalysis* **10**, 4497-4509.

Vechi, G., Tenfen, A., Boeder, A.M., Hernandez-Gómez, L., de Córdova, C.M.M., Monache, F.D. and Filho, V.C. (2019) Chemical composition and antimycoplasmic activity of Eugenia mattosii leaves, stems and isolated compounds. *Natural Product Communications* **14**, 1934578X1901400111.

Vidgren, J., Svensson, L.A. and Liljas, A. (1994) Crystal structure of catechol O-methyltransferase. *Nature* **368**, 354.

Vogt, T. (2004) Regiospecificity and kinetic properties of a plant natural productO-methyltransferase are determined by its N-terminal domain. *FEBS Letters* **561**, 159-162.

Walker, A.M., Sattler, S.A., Regner, M., Jones, J.P., Ralph, J., Vermerris, W., Sattler, S.E. and Kang, C. (2016) The structure and catalytic mechanism of Sorghum bicolor caffeoyl-CoA O-methyltransferase. *Plant physiology* **172**, 78-92.

Walle, T., Ta, N., Kawamori, T., Wen, X., Tsuji, P.A. and Walle, U.K. (2007) Cancer chemopreventive properties of orally bioavailable flavonoids--methylated versus unmethylated flavones. *Biochem Pharmacol* **73**, 1288-1296.

Wang, J. and Pichersky, E. (1999) Identification of specific residues involved in substrate discrimination in two plant O-methyltransferases. *Archives of biochemistry and biophysics* **368**, 172-180.

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F.T., de Beer, T.A.P., Rempfer, C. and Bordoli, L. (2018) SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic acids research* **46**, W296-W303.

Wen, L., Jiang, Y., Yang, J., Zhao, Y., Tian, M. and Yang, B. (2017) Structure, bioactivity, and synthesis of methylated flavonoids. *Annals of the New York Academy of Sciences* **1398**, 120-129.

Widiez, T., Hartman, T.G., Dudai, N., Yan, Q., Lawton, M., Havkin-Frenkel, D. and Belanger, F.C. (2011) Functional characterization of two new members of the caffeoyl CoA O-methyltransferase-like gene family from Vanilla planifolia reveals a new class of plastid-localized O-methyltransferases. *Plant molecular biology* **76**, 475-488.

Wils, C.R., Brandt, W., Manke, K. and Vogt, T. (2013) A single amino acid determines position specificity of an Arabidopsis thaliana CCoAOMT-like O-methyltransferase. *FEBS letters* **587**, 683-689.

Wollenweber, E. and Kohorst, G. (1981) Epicuticular leaf flavonoids from Eucalyptus species and from Kalmia latifolia. *Zeitschrift für Naturforschung C* **36**, 913-915.

Wollenweber, E., Wehde, R., Dörr, M., Lang, G. and Stevens, J.F. (2000) C-Methyl-flavonoids from the leaf waxes of some Myrtaceae. *Phytochemistry* **55**, 965-970.

Woodard, R., Tsai, M., Floss, H., Crooks, P. and Coward, J. (1980) Stereochemical course of the transmethylation catalyzed by catechol O-methyltransferase. *Journal of Biological Chemistry* **255**, 9124-9127.

Wu, Q., Preisig, C.L. and VanEtten, H.D. (1997) Isolation of the cDNAs encoding (+) 6a-hydroxymaackiain 3-O-methyltransferase, the terminal step for the synthesis of the phytoalexin pisatin in Pisum satvium. *Plant molecular biology* **35**, 551-560.

Wu, S., Watanabe, N., Mita, S., Ueda, Y., Shibuya, M. and Ebizuka, Y. (2003) Two O-methyltransferases isolated from flower petals of Rosa chinensis var. spontanea involved in scent biosynthesis. *Journal of bioscience and bioengineering* **96**, 119-128.

Xiao, Y., Li, J., Liu, H., Zhang, Y., Zhang, X., Qin, Z. and Chen, B. (2020) The Effect of Co-Transforming Eucalyptus urophylla Catechol-O-methyltransferase and Caffeoyl-CoA O-methyltransferase on the Biosynthesis of Lignin Monomers in Transgenic Tobacco. *Russian Journal of Plant Physiology* **67**, 879-887.

Yahyaa, M., Berim, A., Nawade, B., Ibdah, M., Dudareva, N. and Ibdah, M. (2019) Biosynthesis of methyleugenol and methylisoeugenol in Daucus carota leaves: Characterization of eugenol/isoeugenol synthase and O-Methyltransferase. *Phytochemistry* **159**, 179-189.

Yauk, Y.K., Chagné, D., Tomes, S., Matich, A.J., Wang, M.Y., Chen, X., Maddumage, R., Hunt, M.B., Rowan, D.D. and Atkinson, R.G. (2015) The O-methyltransferase gene M do OMT 1 is required for biosynthesis of methylated phenylpropenes in ripe apple fruit. *The Plant Journal* **82**, 937-950.

Yoneyama, N., Morimoto, H., Ye, C.-X., Ashihara, H., Mizuno, K. and 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.

Yu, F., Li, M., Xu, C., Sun, B., Zhou, H., Wang, Z., Xu, Q., Xie, M., Zuo, G., Huang, P., Guo, H., Wang, Q. and He, J. (2016) Crystal structure and enantioselectivity of terpene cyclization in SAM-dependent methyltransferase TleD. *Biochemical Journal* **473**, 4385-4397.

Yue, W., Ming, Q.L., Lin, B., Rahman, K., Zheng, C.J., Han, T. and Qin, L.P. (2016) Medicinal plant cell suspension cultures: pharmaceutical applications and high-yielding strategies for the desired secondary metabolites. *Crit Rev Biotechnol* **36**, 215-232.

Yusuf, N.A., M Annuar, M.S. and Khalid, N. (2013) Existence of bioactive flavonoids in rhizomes and plant cell cultures of'Boesenbergia rotund'(L.) Mansf. Kulturpfl. *Australian Journal of Crop Science* **7**, 730-734.

Zhang, C., Freddolino, P.L. and Zhang, Y. (2017) COFACTOR: improved protein function prediction by combining structure, sequence and protein–protein interaction information. *Nucleic acids research* **45**, W291-W299.

Zhang, H., Boghigian, B.A. and Pfeifer, B.A. (2010) Investigating the role of native propionyl-CoA and methylmalonyl-CoA metabolism on heterologous polyketide production in Escherichia coli. *Biotechnology and bioengineering* **105**, 567-573.

Zhang, J., Subramanian, S., Stacey, G. and Yu, O. (2009) Flavones and flavonols play distinct critical roles during nodulation of Medicago truncatula by Sinorhizobium meliloti. *The Plant Journal* **57**, 171-183.

Zhang, Q., Van Der Donk, W.A. and Liu, W. (2012) Radical-mediated enzymatic methylation: a tale of two SAMS. *Accounts of Chemical Research* **45**, 555-564.

Zhang, X.-H. and Chinnappa, C. (1997) Molecular characterization of a cDNA encoding caffeoyl-coenzyme A 3-O-methyltransferase ofStellaria longipes. *Journal of biosciences* **22**, 161-175.

Zhang, Y.-H., Li, Y.-F., Wang, Y., Tan, L., Cao, Z.-Q., Xie, C., Xie, G., Gong, H.-B., Sun, W.-Y., Ouyang, S.-H., Duan, W.-J., Lu, X., Ding, K., Kurihara, H., Hu, D., Zhang, Z.-M., Abe, I. and He, R.-R. (2020) Identification and characterization of N9-methyltransferase involved in converting caffeine into non-stimulatory theacrine in tea. *Nature Communications* **11**, 1473.

Zhang, Y. and Skolnick, J. (2005) TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic acids research* **33**, 2302-2309.

Zhao, N., Ferrer, J.-L., Moon, H.S., Kapteyn, J., Zhuang, X., Hasebe, M., Neal Stewart, C., Gang, D.R. and 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. and 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, Y., Wang, N., Wu, H., Zhou, Y., Huang, C., Luo, J., Zeng, Z. and Kong, L. (2020) Structure-based tailoring of the first coumarins-specific bergaptol O-methyltransferase to synthesize bergapten for depigmentation disorder treatment. *Journal of advanced research* **21**, 57-64.

Zheng, W., Zhang, C., Bell, E.W. and Zhang, Y. (2019) I-TASSER gateway: A protein structure and function prediction server powered by XSEDE. *Future Generation Computer Systems* **99**, 73-85.

Zhong, F., Huang, L., Qi, L., Ma, Y. and Yan, Z. (2020) Full-length transcriptome analysis of Coptis deltoidea and identification of putative genes involved in benzylisoquinoline alkaloids biosynthesis based on combined sequencing platforms. *Plant Molecular Biology* **102**, 477-499.

Zhong, R., Morrison, W.H., Negrel, J. and Ye, Z.-H. (1998) Dual methylation pathways in lignin biosynthesis. *The plant cell* **10**, 2033-2045.

Zhou, J.-M., Fukushi, Y., Wollenweber, E. and Ibrahim, R.K. (2008) Characterization of Two O.-Methyltransferase-like Genes in Barley and Maize. *Pharmaceutical biology* **46**, 26-34.

Zhou, J.-M., Gold, N.D., Martin, V.J., Wollenweber, E. and Ibrahim, R.K. (2006) Sequential O-methylation of tricetin by a single gene product in wheat. *Biochimica et Biophysica Acta (BBA)-General Subjects* **1760**, 1115-1124.

Zhou, J.-M., Lee, E., Kanapathy-Sinnaiaha, F., Park, Y., Kornblatt, J.A., Lim, Y. and Ibrahim, R.K. (2010) Structure-function relationships of wheat flavone O-methyltransferase: Homology modeling and site-directed mutagenesis. *BMC plant biology* **10**, 156.

Zubieta, C., He, X.-Z., Dixon, R.A. and 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.

Zubieta, C., Kota, P., Ferrer, J.-L., Dixon, R.A. and Noel, J.P. (2002) Structural basis for the modulation of lignin monomer methylation by caffeic acid/5-hydroxyferulic acid 3/5-O-methyltransferase. *The Plant Cell* **14**, 1265-1277.

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

APPENDIX A

Consolidated list of functionally characterised MTs mentioned in this thesis.

APPENDIX B

List of functionally characterised methyltransferases used for phylogenetic analysis in Chapter 2

Category	Branch labels in phylogenetic tree	Accession number	Organism
A	Alkaloid DXNMT_C. canephora	DQ422955	*Caffea canephora*
A	Alkaloid XNMT (DQ422954)_C. canephora	DQ422954	*Caffea canephora*
F	Flavonoid OMT (CAA54616)_H. vulgare	CAA54616	*Hordeum vulgare (*Barley)
F	Anthocyanin OMT (AIS23587)_S. lycopersicum	AIS23587	*Solanum lycopersicum*
F	Chalcone Isoliquritigenin 2'OMT (AAB48059)_M. sativa	AAB48059	*Medicago sativa*
F	EcOMT1_E. croajingolensis		*Eucalyptus croajingolensis*

F	EnOMT1_E. nitida		*Eucalyptus nitida*
F	ExOMT1_Nonmethylated species		*Non-methylate species*
F	F myricetin 3'5'OMT (JF499656)_S. habrochaites.	JF499656	*Solanum habrochaites (wild tomato)*
F	F myricetin 3OMT (KC513419)_S. habrochaites.	KC513419	*Solanum habrochaites (wild tomato)*
F	F myricetin 7OMT (JF499657)_S. habrochaites.	JF499657	*Solanum habrochaites (wild tomato)*
F	Flav>CCoA OMT(JF344740)_V. planifolia	JF344740	*Vanilla planifolia*
F	Flav>CCoA OMT(JF344741)_V. planifolia	JF344741	*Vanilla planifolia*
F	Flavanone Naringenin 4'OMT(C6TAY1)_G. max	C6TAY1	*Glycine max Soyabeen*
F	Flavanone Naringenin 7OMT (AB692949)_O. sativa	AB692949	*Oryza sativa*
F	Flavone 3'OMT (U16793)_C. americanum	U16793	*Chrysosplenium americanum*
F	Flavone OMT (ABQ58826)_Z. mays	ABQ58826	*Zea mays*
F	Flavone Tricetin (AAP23942)_T. aestivum	AAP23942	*Triticum aestivum*
F	Flavonoid 4'OMT (AAR02420.1)_C. roseus	AAR02420.1	*Catharanthus roseus*
F	Flavonoid Anthocyanin OMT (ADJ57332)_V. vinifera	ADJ57332	*Vitis vinifera*

F	Flavonoid OMT (LC126059)_C. depressa	LC126059	*Citrus depressa*
F	Flavonol 3'-5' OMT (U16794)_C. americanum	U16794	*Chrysosplenium americanum*
F	Flavonol 3'OMT (AY337460)_M. piperita	AY337460	*Mentha x piperita*
F	Flavonol 4'OMT(AY337461)_M. piperita	AY337461	*Mentha x piperita*
F	Flavonol 7OMT (AY337457)_M. piperita	AY337457	*Mentha x piperita*
F	Flavonol 8OMT (AY337459)_M. piperita	AY337459	*Mentha x piperita*
F	Flavonol Myricetin OMT (AAM97497)_C. roseus	AAM97497	*Catharanthus roseus*
F	Isoflavanone HI4'OMT (BAC58013)_L. japonicus	BAC58013	*Lotus japonicus*
F	Isoflavanone OMT (AB091684)_G. echinata	AB091684	*Glycyrrhiza echinata*
F	Isoflavanone4'OMT (AAY18581)_M. truncatula	AAY18581	*Medicago truncatula*
F	Isoflavone 7OMT (AAC49927)_M. sativa	AAC49927	*Medicago sativa*
F	Isoflavone Diadzein 7OMT (BAC58012)_G. echinata	BAC58012	*Glycyrrhiza echinata*
F	Phytoalexin (+)-6a-hydroxymaackiainOMT (AAC49856)P. sativum	AAC49856	*Pisum sativum*

F	Phytoalexin (+)-6a-hydroxymaackiainOMT (P0DH60)_P. sativum	P0DH60	*Pisum sativum*
F	PP Flavone Tricetin (DQ223971)_T. aestivum	DQ223971	*Triticum aestivum*
F	Prenylchalcone Desmethylxanthohumol OMT (ABZ89565)_H. lupulus	ABZ89565	*Humulus lupulus*
F	prenylchalcone Xanthohumol 4OMT (ABZ89566)_H. lupulus	ABZ89566	*Humulus lupulus*
P	CA (X62096)_Populus tremuloides	X62096	*Populus tremuloides*
P	CA ConiferaldehydeOMT(AF139533)_L. styraciflua	AF139533	*Liquidambar styraciflua*
P	Caf 3OMT(AY028439)_C. roseus	AY028439	*Catharanthus roseus*
P	Cafeeic acid OMT (U70424)_A. thaliana	U70424	*Arabidopsis thaliana*
P	Caffeic acid 3OMT (AAB46623)_M.sativa	AAB46623	*Medicago sativa*
P	Caffeic acid OMT (AY443007)_A. majus	AY443007	*Ammi majus*
P	Caffeic acid OMT(HQ661801)_S. bicolor	HQ661801	*Sorghum bicolor*
P	Caffeoyl aldehyde (AY555144)_V. planifolia	AY555144	*Vanilla planifolia*
P	Caffeoyl CoA OMT (AAC28973)_M. sativa	AAC28973	*Medicago sativa*

P	Caffeoyl CoA OMT(AAN61072)_M. crystallinum	AAN61072	*Mesembryanthemum crystallinum*
P	Carboxyl Hormone IAA OMT (Q9FLN8)_A. thaliana	Q9FLN8	*Arabidopsis thaliana*
P	Catechol /Cafacid OMT(X71430)_N. tabacum	X71430	*Nicotiana tabacum*
P	CCoA (M69184)_P. crispum	M69184	*Petroselinum crispum*
P	CCoA/PP Tricetin OMT (AB110168)_O. sativa	AB110168	*Oryza sativa*
P	CCoA/PP Tricetin OMT (AK061757)_O. sativa	AK061757	*Oryza sativa*
P	CCoAOMT (Z54233)_V. vinifera	Z54233	*Vitis vinifera*
P	CCoAOMT(AJ224894)_P. trichocarpa	AJ224894	*Populus trichocarpa*
P	Furanocoumarin Bergaptol OMT(A8J6X1) _G. littoralis	A8J6X1	*Glehnia littoralis*
P	Hormone Jasmonic acid OMT (AAG23343)_A. thaliana	AAG23343	*Arabidopsis thaliana*
P	Isoeugenol OMT (U86760)_C. breweri	U86760	*Clarkia breweri*
P	Phenylpropanoid OMT (DQ288259)_O. sativa	DQ288259	*Oryza sativa*
Phenol	Catechol OMT (BC081850)_Rat	BC081850	Rat

APPENDIX C

Functional annotation details of Unigene clusters described in Chapter 4

E.diversifolia annotation summary

APPENDIX D

Seed source, provenance and propagation details of the species used in this study.

	Species #	**Seed source**	**Provenance**	**Propagation**
1	*E. apiculata*	Nindethana seed 945	Cultivated tree(s)	UoM
2	*E. preissiana*	Nindethana seed 1222	Woogenellup, WA	UoM
3	*E. nitida*	Nindethana seed 1170	Cultivated tree(s)	UoM
4	*E. stellulata*	Royston Petrie Seeds	ACT region	Kuranga Nursery
5	*E. mitchelliana*	Kuranga Nursery	Mt. Buffalo, Vic.	Kuranga Nursery
6	*E. haemastoma*	Royston Petrie Seeds	S coast region, NSW	Facey's Nursery
7	*E. diversifolia subsp. diversifolia*	ERA Nurseries	Portland, Vic	ERA Nurseries
8	*E. agglomerata*	CSIRO ATSC 15815	Wandera, NSW	UoM
9	*E. muelleriana 856*	Seeds of Gippsland 856	Darriman, Vic.	UoM
10	*E. croajingolensis 1615*	Seeds of Gippsland 1612	Woodside, Vic.	UoM

APPENDIX E

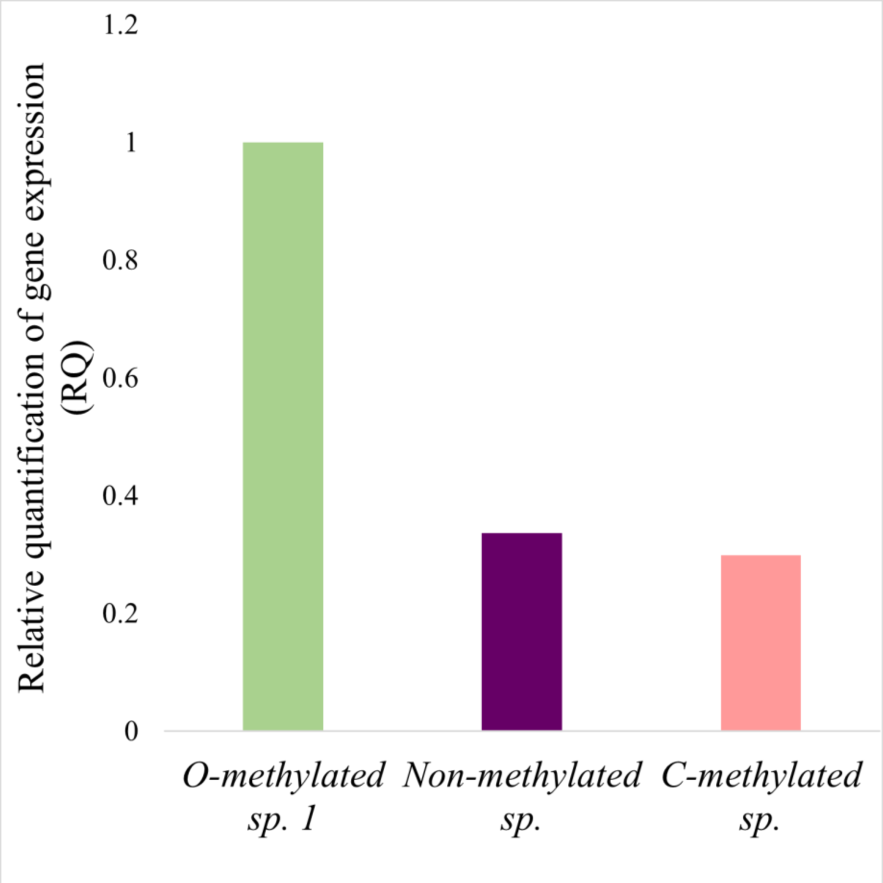


Figure E1: qPCR results in terms of relative quantification of gene expression showing low expression of OMT homologues in the *non-methylated* and *C-methylated sp.* compared to the *O-methylated species 1*. Data was normalised to the reference gene mentioned in Table 2.1