ELUCIDATION AND CHARACTERIZATION OF GENES ASSOCIATED WITH MONTBRETIN A BIOSYNTHESIS WITHIN *CROCOSMIA x CROCOSMIIFLORA* AND WHITE PINE WEEVIL DEFENSE IN *PICEA SITCHENSIS*

by

CHRISTOPHER ROBERT ROACH

B.Sc. Cellular Biology and Genetics, The University of British Columbia, 2010

A DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(Genome Sciences and Technology)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

November 2017

ABSTRACT

Plant-specialized metabolites have long been utilized as medicines, cosmetics, flavours, and industrial raw materials. To explore the biosynthesis of a specialized metabolite in a non-model system and utilize the biosynthetic genes for future application, genomics-informed research typically flows through three phases: i) development of genomic or transcriptomic resources, ii) discovery and characterization of biosynthetic genes, and iii) application of the genes and enzymes for improved production of the specialized metabolite. This thesis describes hypothesis-driven research along these three phases in two different plant species and two different metabolic systems.

My research with *Crocosmia x crocosmiiflora* focused on resource development and discovery of biosynthetic genes of a specialized metabolite of interest, montbretin A (MbA). I developed new resources for this system including metabolite-profiles and transcriptome sequences and annotations. This work resulted in insight into the spatial and temporal patterns of MbA accumulation in *C. x crocosmiiflora* and a first reference transcriptome with annotation for this species. Using these resources, I functionally characterized four UDP-xylose synthases and five UDP-rhamnose synthases. I discuss the application of these genes for possible use in an improved MbA production system and provide a proof of concept for using these genes to enable characterization of downstream MbA biosynthetic genes. I also identified 14 UDP-glycosyltransferases as candidate MbA biosynthetic genes through a guilt-by-association analysis; however, their functional characterization did not support a role in MbA biosynthesis.

In the second biological system, Sitka spruce (*Picea sitchensis*), I performed a detailed characterization of a set of monoterpene synthases involved in the biosynthesis of the (+)-3-carene. Using domain swapping and site-directed mutagenesis, I demonstrated the catalytic plasticity of monoterpene synthases across a family of (+)-3-carene synthase-like genes associated with *P. sitchensis* resistance against the white pine weevil (*Pissodes strobi)*. This work identified a single amino acid as most critical in determining both product profile and enzyme kinetics. Furthermore, I described mechanisms by which this amino acid directs product profiles through differential stabilization of the reaction intermediate. The work presented highlights the inherent plasticity and potential for evolution of alternative product profiles of these monoterpene synthases of conifer defense against pests.

LAY SUMMARY

Plant-specialized metabolites are valuable resources employed by humans. This thesis explores metabolites in two plant species, showcasing the research pipeline used to explore such systems. Research on *Crocosmia x crocosmiiflora* focused on identifying the genes involved in biosynthesis of montbretin A (MbA), a specialized metabolite of interest. This thesis established the first ever set of biological resources for *C. x crocosmiiflora*. Building off these resources, members of the UDP-xylose synthase and UDP-rhamnose synthase gene families were functionally characterized. While attempts to identify UDP-glycosyltransferases involved in biosynthesis were unsuccessful, results provided useful insight for future attempts. Research on Sitka spruce focused on exploring the plasticity of a family of (+)-3-carene synthase-like genes. Results showed the effect a single amino acid can have on altering the functioning of an enzyme. This work highlights the inherent plasticity and potential for evolution of these monoterpene synthases of conifer defense against pests.

PREFACE

Chapter 2:

Roach, C. R., Yuen, M., Madilao, L. L., Irmisch, S, Withers, S. G. and Bohlmann, J. (2017). Development of *Crocosmia* resources for the elucidation of the montbretin A biosynthetic pathway. *In preparation*.

C.R. Roach conceived, designed, and performed all experiments described herein, as well as wrote the manuscript. M. Yuen provided technical assistance and support in producing the transcriptomic assembly and analyzes of the transcriptomic data. L.L. Madilao provided input for liquid chromatography-mass spectrometry equipment operations, programs, and parameters. S. Irmisch supported the *in silico* annotation of montbretin A biosynthetic genes and reviewed the manuscript. S. Withers provided montbretin A and valuable suggestions for the design of experiments analyzing montbretin A accumulation within *Crocosmia x crocosmiiflora*. J. Bohlmann directed research and supported chapter preparation.

Chapter 3:

Roach, C. R., Madilao, L. L. and Bohlmann, J. (2017). Functional characterization of *Crocosmia x crocosmiiflora* nucleotide sugar interconversion enzymes involved in montbretin A biosynthesis. *In preparation*.

C.R. Roach conceived, designed, and performed all experiments described herein, as well as wrote the manuscript. L.L. Madilao provided input for liquid chromatography-mass spectrometry equipment operations, programs, and parameters. J. Bohlmann directed research and supported chapter preparation.

Chapter 4:

Roach, C. R., Irmisch, S., Madilao, L. L., Withers, S. G. and Bohlmann, J. Identification of Crocosmia x crocosmiiflora UDP-glycosyltransferases involved in montbretin A biosynthesis.

C.R. Roach conceived, designed and performed all experiments described herein, as well as wrote the manuscript. S. Irmisch identified the pASK-IBA37(+) vector as capable of expressing functioning UDP-glycosyltransferases for *in vitro* assays, as well as provided valuable support and suggestions for experiments testing *in vitro* activity of UDP-

glycosyltransferases. L.L. Madilao provided input for liquid chromatography-mass spectrometry equipment operations, programs, and parameters. S. Withers first proposed the use montbretin A substructures as substrates for *in vitro* assay with candidate GT1 UGTs, provided montbretin A, access to high performance liquid chromatography equipment, and technical support in the production of the potential montbretin A intermediates. J. Bohlmann directed research and supported chapter preparation.

Chapter 5:

Roach, C. R., Hall, D. E., Zerbe, P. and Bohlmann, J. (2014) Plasticity and evolution of (+)-3-carene synthase and (−)-sabinene synthase functions of a Sitka spruce monoterpene synthase gene family associated with weevil resistance. *J. Biol. Chem.*, **289**(34), 23859-23869.

C.R. Roach conceived, designed and performed all experiments described herein, as well as wrote the manuscript. D. Hall provided valuable support in the designing and planning of the experiments, as well as provided technical assistance with gas chromatography-mass spectrometry equipment, operations, and programs. P. Zerbe provided technical assistance with protein modeling and molecular docking programs, as well as supported work exploring protein-intermediate interactions. Both D. Hall and P. Zerbe supported manuscript preparation. J. Bohlmann directed research and supported manuscript preparation.

All tables and figures within this thesis are produced by C.R. Roach. Proper copyright permission was obtained where necessary.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

2-MBT	2-mercaptobenzothiazole
α	Alpha
AS	Aureusidin synthase
AtRHM2-C	*Arabidopsis thaliana* UDP-rhamnose synthase epimerase/reductase domain
AtRHM2-N	*Arabidopsis thaliana* UDP-rhamnose synthase dehydratase domain
β	Beta
BGL	Blood glucose level
bp	Base pair
BUSCO	Benchmarking Universal Simple-Copy Orthologs
CAZy	Carbohydrate Active Enzyme database
cDNA	Complementary deoxyribonucleic acid
CDS	Coding sequences
CEGMA	Core Eukaryotic Gene Mapping Approach
CHI	Chalcone isomerase
CHS	Chalcone synthase
CVC	Central vascular cylinder
DAD	Diode array detector
DFR	Dihydroflavonol 4-reductase
dH20	Distilled water
DMAPP	Dimethylallyl diphosphate
EPB	Early biosynthesis pathway
F3′5′H	Flavonoid 3′5′-hydroxylase
F3H	Flavanone 3-hydroxylase
F3′H	Flavonoid 3′-hydroxylase
$FeCL_3$	Iron chloride
FLS	Flavonol synthase
FNS	Flavone synthase
FPP	Farnesyl diphosphate
FT	Fresh tissue
Gbp	Giga base pair
GGPP	Geranylgeranyl diphosphate
GO	Gene Ontology
GPP	Geranyl diphosphate
GT	Glycosyltransferase
HCl	Hydrochloric acid
HPA	Human pancreatic amylase
IFD	2-hydroxyisoflavanone dehydratase
IPP	Isopentyl diphosphate
ITO	Indium-tin oxide
k_{cat}	Catalytic turnover constant
K_i	Inhibitory constant

K_M	Michaelis constant
KOG	EuKaryotic Orthologous Groups
LBP	Late biosynthesis pathway
LC	Liquid chromatography
LC-MS	Liquid chromatography-mass spectrometry
LCR	Lucoanthocyanidins reductase
LPP	Linalyl diphosphate
MALDI	Matrix-Assisted Laser Desorption Ionization
MbA	Montbretin A
MbA-CR′	MbA without the caffeic acid and 4′-O rhamnopyranose moieties
MbA-CXR′	MbA without the caffeic acid and 4′-O-rhamnopyranosyl-xylopyranose moieties
MbA-G′	MbA without the 3-O glucopyranose moiety
MbA-R′	MbA without the 4′-O rhamnopyranose moiety
MbA-XR′	MbA without the 4′-O-rhamnopyranosyl-xylopyranose moiety
MEP	2-C-methyl-D-erythritol phosphate 5
mono-TPS	monoterpene synthase
MSD	Mass spectrometry detector
MVA	Mevalonate pathway
MVA	Mevalonate
NDP-sugars	Nucleoside 5′-diphosphate sugars
NGS	Next-generation sequencing
NMR	Nuclear magnetic resonance spectroscopy
NSE	Nucleotide sugar interconversion enzymes
NSEs	Nucleoside diphosphate sugar interconversion enzymes
NSTs	Nucleotide sugar transporters
p	Para
PAL	Phenylalanine ammonia lyase
PATH	Phenylpropanoid biosynthesis
PCR	Polymerase chain reaction
pH	Potential hydrogen
PSPG	Plant secondary product glycosylation
RHM	Uridine 5′-diphosphate rhamnose synthase
RNA	Ribonucleic acid
RNA-seq	Transcriptome sequencing
RPKM	Reads per kilobase of transcript per million mapped reads
SN_2	Nucleophilic substitution with a rate determining step involving two components
T2D	Type 2 diabetes
TAL	Tyrosine ammonia lyase
UDP	Uridine 5′-diphosphate
UDP-4K6DG	Uridine 5′-diphosphate-4-keto-6-deoxy-glucose
UDP-4K-GlcA	Uridine 5′-diphosphate-4-keto-glucuronic acid
UDP-4KP	Uridine 5′-diphosphate-4-keto-pentose
UDP-4KPS	Uridine 5′-diphosphate-4-keto pentose synthase
UDP-Glc	Uridine 5′-diphosphate glucose

UDP-GlcA	Uridine 5′-diphosphate glucuronic acid
UDP-Rha	Uridine 5′-diphosphate rhamnose
UDP-sugars	Uridine 5′-diphosphate sugars
UDP-Xyl	Uridine 5′-diphosphate xylose
UER	Uridine 5′-diphosphate-4-6-deoxy-glucose 3,5-epimerase/uridine diphosphate-4-keto-rhamnose reductase
UGTs	Uridine diphosphate glycosyltransferases
UV	Ultraviolet
UXS	Uridine 5′-diphosphate xylose synthase
WT	Wild type

ACKNOWLEDGEMENTS

I must first thank my Ph.D. supervisor, Dr. Jörg Bohlmann, for the opportunities of conducting research as part of his group and allowing me to pursue the multiple non-research opportunities which have allowed my own unique professional development path. Similarly, I must thank Dr. Stephen Withers, Dr. Katherine Ryan, and Dr. Harry Brumer for the advice and guidance they have offered while serving as members of my academic committee.

While all members of the Bohlmann lab have contributed to my development, special thanks must be given. To Dr. Dawn Hall, Dr. Philipp Zerbe, and Dr. Sandra Irmisch, for their roles as mentors and guidance on my projects. To Mack Yuen, on his assistance and training on working with transcriptomic data. To Lina Madilao, for her expertise with GC/LC-MS analytical analysis. And to Karen Reid and Angela Chiang, for their fantastic lab management.

Outside of the Bohlmann lab, special thanks must be given to Dr. Xiaohua Zhang and Dr. Mirium Koetzler for their assistance in producing the potential montbretin A intermediates.

Thanks, must also be given to all those members of the UBC community, from student societies to university administration to the UBC Senate and Board of Governors, who have always practiced patience and the willingness to view all university activities as an opportunity to further students' education.

Finally, and most importantly, I must thank those in my personal life who have been there through the highs and the lows and have always supporting me. To my family, thank you for your unconditional love and willingness to help whenever and with whatever. To my friends, for understanding and accepting the work schedule I have kept these past years. And special thanks to the Great Simon for your love and support throughout the years. I hope one day soon I can pay back all that you have done for me throughout this process.

Oh and to coffee. Definitely coffee.

CHAPTER 1: INTRODUCTION

1.1 HUMAN USE OF SPECIALIZED METABOLITES

1.1.1 Specialized Metabolism in Plants

As sessile organisms, plants have the challenge of not only surviving but thriving under constantly changing environmental conditions over lifespans of several weeks to hundreds of years. They must maintain essential processes such as growth, development, and reproduction, while also responding to many different forms of biotic and abiotic stresses. To cope with biotic and abiotic stresses, plants have evolved complex systems of specialized metabolism traditionally referred to as secondary metabolism. Specialized metabolites take form as small molecules, which possess an enormous structural diversity based on thousands of skeletal structures and many functional modifications thereof, and have functions in the interactions of plants with their environments (Pichersky and Lewinsohn, 2011). It is currently estimated that plants produce 200,000 – 300,000 different specialized metabolites (Dixon and Strack, 2003; Lawrence, 1964). However, the hypothesis that these numbers are probably an underestimation is supported by three observations: i) the large diversity of specialized metabolism genes among plant species with available genomic data, ii) the large number of specialized metabolites detected in individual species, and iii) the vast number of species that have not had their metabolomes and genomes explored (Yonekura-Sakakibara and Saito, 2009). Within the myriad of plant-specialized metabolites, certain patterns emerge for how plant use these compounds. For example, many terpenoids serve as defense compounds through actions as antimicrobials, insecticidals, attractants to predators of herbivores, and physical barriers (Keeling and Bohlmann, 2006b; Martin *et al.*, 2003; Phillips and Croteau, 1999). Alkaloids serve as feeding deterrents due to their liver- and neuro-toxic effects on vertebrates and mutagenic effects on insects (Frei *et al.*, 1992; Mattocks, 1986; Schmeller *et al.*, 1997). Phenolics serve as protection against ultraviolet light, antioxidants, and colour and sensory characteristics (Alasalvar *et al.*, 2001; Balasundram *et al.*, 2006; Cuvelier *et al.*, 1996).

Since humans started exploring plants for use thousands of years ago, one of the biggest boons has been through harnessing their bioproducts. Currently, plant-specialized metabolites are so ubiquitous in everyday life, most individuals are not aware of all the ways we employ them; from cosmetics to flavours to health products.

In this thesis, I will use the term "plant-specialized metabolites" in place of the classical terms "plant secondary metabolites" or "plant natural products". This change in nomenclature follows a trend in the academic field which emphasizes moving away from using terms which could suggest, from human and research perspectives, that one branch of plant metabolism is more important than another (i.e. primary vs. secondary). Throughout this thesis any reference to "specialized metabolism" should be considered a reference to "plant-specialized metabolism".

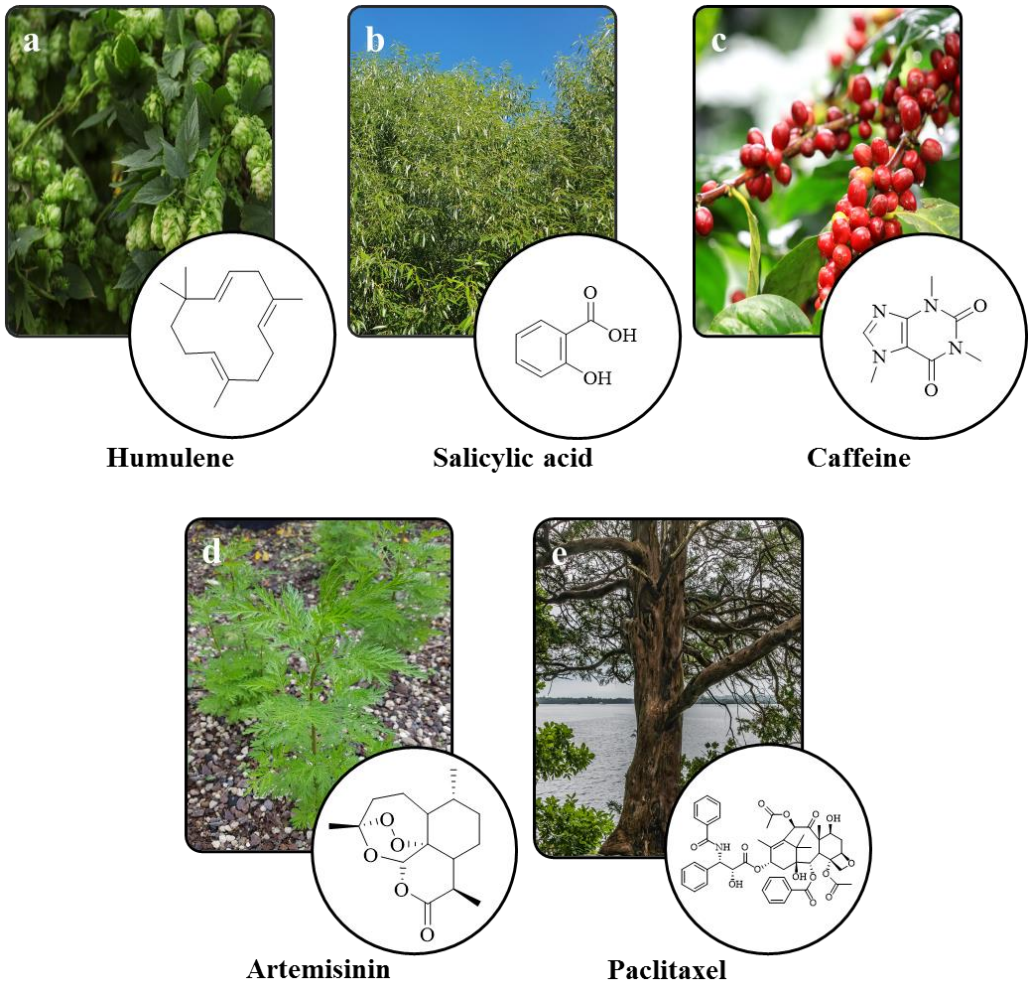


Figure 1.1: Examples of plant-specialized metabolites. (a) Humulene found in *Humulus lupulus*. (b) Salicylic acid found in *Salix alba*. (c) Caffeine found in *Coffea arabica*. (d) Artemisinin found in *Artemisia annua*. (e) Paclitaxel found in *Taxus brevifolia*.

1.1.2 Human Application of Specialized Metabolites

In addition to their role in plants, specialized metabolites have played a large role in human history. For thousands of years, humans have utilized botanical extractions as food

preservatives, medicines, pigments, and weapons. Modern uses of specialized metabolites include sources of new pharmaceuticals, chemicals and biomaterials and many of the most economically valuable molecules within a given market are specialized metabolites (Facchini *et al.*, 2012). With high-throughput genomic, transcriptomic, metabolomic, and proteomic technologies, our ability to identify and characterize both specialized metabolites with potential use for human application, and their biosynthetic pathways continue to grow (Borevitz and Ecker, 2004). With the onset of this "-omics age", we have developed the means to collect large amounts of heterogeneous biological information about an organism; from metabolome data to proteome, transcriptome, or genome sequence. These tools empower us with the ability to explore how a biological system functions. However, even with these technologies, harnessing a specialized metabolite for sustained human use is still a daunting task. With this goal in mind, research typically flows through the following three phases (Fig. 1.2):

1) Development of resources to provide a foundation for studying a specialized metabolite system.

2) Characterization of the genes involved in the specialized metabolite system

3) Utilizing the functions of the specialized metabolite system for human applications.

The body of work presented in this thesis investigates questions in each of these three areas with an overall emphasis on improving our understanding of a specialized metabolic system.

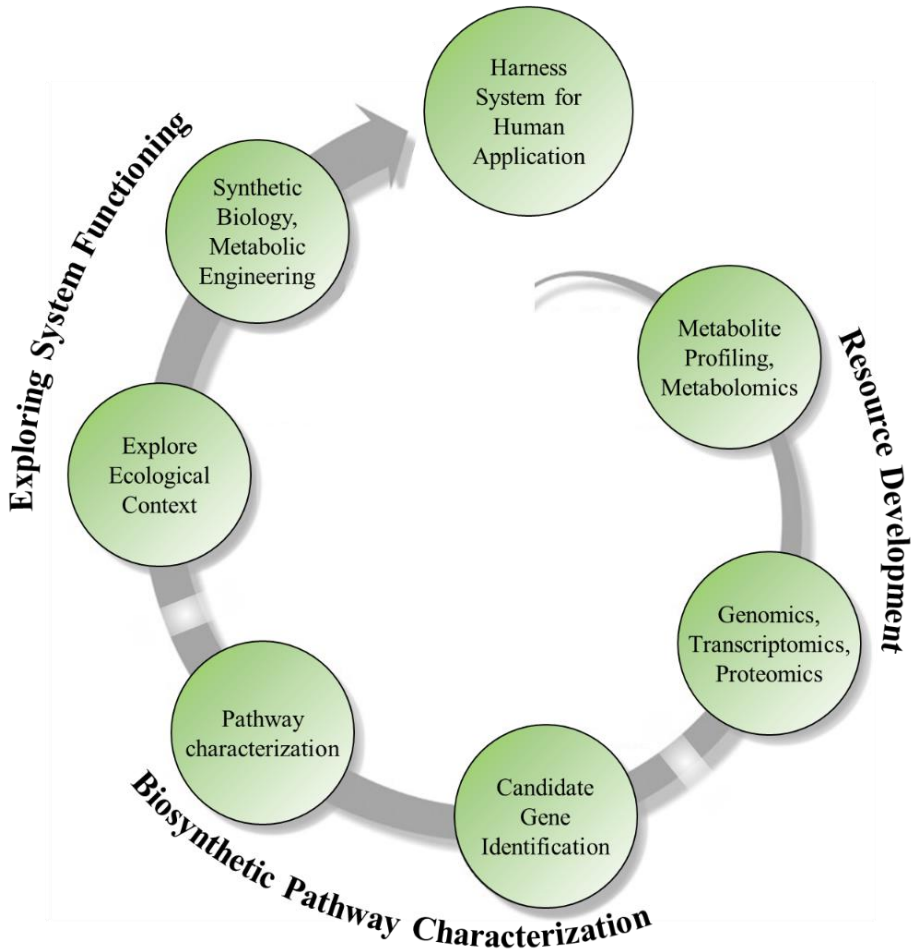


Figure 1.2: Research process for studying plant-specialized metabolism system.

1.1.3 Examples of Employing Plant-Specialized Metabolite Systems

Two well-known examples of this three-phase approach can be seen in the exploration of the specialized metabolites paclitaxel and artemisinin.

Paclitaxel is a high-value pharmaceutical originally isolated from bark of the conifer *Taxus brevifolia* used to treat breast, lung, and non-small cell cancers (Cragg, 1998). Due to the prohibitively difficult and expensive chemical synthesis and low levels obtained from natural harvest, obtaining sufficient amounts to meet demand has proved challenging (Holton *et al.*, 1994; Malik *et al.*, 2011; Nicolaou *et al.*, 1994; Wani *et al.*, 1971). To establish a sustainable production system, the paclitaxel biosynthetic pathway was explored in *T. brevifolia* and other *Taxus* species. Development of initial key resources included *T. brevifolia* cell culture lines capable of producing paclitaxel (Christen *et al.*, 1991; Stierle *et al.*, 1993) and cDNA sequence collections (Jennewein *et al.*, 2004). Using a hypothesized biosynthetic

pathway and observed putative intermediates, a combination of PCR based cloning and functional screening approaches lead to the characterization of a first set of eight genes of the biosynthetic pathway (Hefner *et al.*, 1996; Jennewein *et al.*, 2004; Long *et al.*, 2008; Menhard and Zenk, 1999; Schoendorf *et al.*, 2001; Walker *et al.*, 2000; Walker and Croteau, 2000; Wildung and Croteau, 1996). Building on this knowledge, extensive research was untaken to establish a sustainable production system using cultured *Taxus* cells and metabolically engineered microbial hosts (DeJong *et al.*, 2006; Li *et al.*, 2009; Meng *et al.*, 2011; Wei *et al.*, 2012; Zhao *et al.*, 2008). While this work has yielded industrial scale semi-synthetic production systems, despite close to 30 years of work, full pathway elucidation has yet to be completed.

Artemisinin is a high-value antimalarial pharmaceutical produced in *Artemisia annua*, (Paddon and Keasling, 2014). The *A. annua* artemisinin biosynthetic pathway was explored to establish additional and alternative production systems for artemisinin. Development of resources included metabolite profiling of *A. annua* leaf and gland secretory cell extracts for likely biosynthetic intermediates (Bertea *et al.*, 2005), and the use of *A. annua* transcriptome data in combination with genomic data from other *Asteraceae* plants to guide gene discovery (Wang *et al.*, 2009a). Using these resources, four genes of the biosynthetic pathway, which produce artemisinic acid from the sesquiterpene precursor farnesyl diphosphate (FPP) were characterized (Mercke *et al.*, 2000; Paddon *et al.*, 2013; Ro *et al.*, 2006; Teoh *et al.*, 2009). Based on this extensive research, microbial systems capable of producing artemisinin precursors were developed. Building on previous knowledge about the mevalonate pathway (MVA), initially an *Escherichia coli* system was used to optimize MVA pathway gene expression and growth conditions to improve yields of the artemisinin precursor amorphadiene (Newman *et al.*, 2006; Tsuruta *et al.*, 2009). Subsequently metabolic engineering of *Saccharomyces cerevisiae* was employed to produce artemisinic acid (Ro *et al.*, 2006; Teoh *et al.*, 2009). Artemisinic acid is then chemically converted to artemisinin (Brown, 2010).

1.1.4 Plant Specialized Metabolite Production Using Recombinant Production Platforms

Similar to artemisinin and paclitaxel, the development of sustainable production systems for high-value plant specialized metabolites (Facchini *et al.*, 2012). Sustainable production of these metabolites is often challenging due to low *in planta* availability and low extraction yields, or due to inefficient chemical synthesis resulting from complex chemical structures and inability to separate the metabolite from isomers and epimers that compromise the biological activity of the products (Mora-Pale *et al.*, 2013). While development of an *in planta* system able to produce high enough levels of the target metabolite is an ideal solution, metabolic engineering of microbial or plant cell systems has become a commonly used method to address these economic and sustainability issues. Conventional metabolic engineering strategies employ two separate phases. The first focuses on *de novo* pathway engineering of the specialized metabolite's biosynthetic pathway coupled with protein engineering and mutagenesis. The second optimizes production through overexpression of rate-limiting steps and deletion of precursor competing-pathways to redirect carbon flux, and the use of cheap precursors to promote direct synthesis (Xu and Koffas, 2010). These approaches have resulted in great success towards the development of microbial systems able to produce a range of specialized metabolites including terpenoids (Leonarda *et al.* 2010, Ajikumar *et al.*, 2010), flavonoids (Lim *et al.*, 2011, Xua *et al.*, 2011), alkaloids (Nakagawa *et al.*, 2011), polyketides (Boghigian *et al.*, 2011) and fatty acids (Zhang *et al.*, 2011). However, even with the advent of the omics-era and the plethora of biological data available, only a handful of metabolically engineered production systems have resulted in commercially viable systems. Continued work identifying new specialized metabolite

biosynthetic genes of interest, employing new tools allowing more efficient rewiring of the cell's native metabolism, and optimizing gene expression to alleviate the toxicity displayed by many of the target specialized metabolites or their intermediates will continue to improve our ability to utilize these expression systems (Mora-Pale *et al.*, 2013)

1.2 MONTBRETIN A AND *CROCOSMIA x CROCOSMIIFLORA*

1.2.1 Diabetes mellitus

Type 2 diabetes (T2D) is a chronic endocrine disease state characterized by hyperglycemia, hyperlipidemia, relative hypoinsulinemia, and increased micro- and macro-vascular disease (Fowler, 2008). The prevalence of the disease has been persistently rising the past few decades and is predicted to be the seventh leading cause of health-related death worldwide (Mathers and Loncar, 2006). Effective management of T2D focuses on improving insulin sensitivity and slowing the release of glucose from a meal or body storage. Because of the high correlation between T2D and certain dietary and lifestyle factors, most noticeably obesity and physical inactivity, the first treatment method most often prescribed is the adoption of a healthier lifestyle (Crandall *et al.*, 2008). This is often coupled with pharmaceutical treatment for proper blood glucose level (BGL) management. Currently, there are several classifications of drugs used in the treatment of T2D, including sulfonylureas, biguanides, thiazolidinedione, glinides, dipeptidyl peptidase inhibitors, and α-glucosidase inhibitors; each acting to lower BGL through a variety of mechanisms (Nathan *et al.*, 2009; van de Laar, 2008).

The common pharmaceutical treatment used to slow the release of glucose after a meal is inhibition of α-amylase or α-glucosidase, enzymes responsible for digesting starch into oligosaccharides and oligosaccharides into monosaccharides, respectively (Crandall *et al.*, 2008). While α-glucosidase inhibitors such as Acarbose©, Miglitol©, and Voglibose© are currently available, activity results in natural colon flora having access to higher than normal levels of small- to medium-size oligosaccharides for digestion through anaerobic respiration as well as osmotic changes arising from the high concentration of oligosaccharides (Aoki *et al.*, 2010). This often results in side effects that include flatulence, diarrhea, and abdominal discomfort, which can lead to patient non-compliance. Because this flora digest starch at a much slower rate than smaller oligosaccharides, a selective inhibitor targeting α-amylase

activity and not α-glucosidase would be preferable due to its ability to decrease BGLs and minimize these factors leading to patient non-compliance.

1.2.2 Montbretin A as an HPA Inhibitor

Tarling *et al.* (2008) screened 30,000 National Cancer Institute terrestrial plant extracts looking specifically for human pancreatic amylase (HPA) inhibitors. Of these, the strongest inhibitors identified were a family of glycosylated acyl flavonols, montbretin A – C, found in extracts from *Crocosmia x crocosmiiflora*. Of the three, montbretin A (MbA) showed the highest inhibitory kinetics with a K_i of 8.1 nM. MbA contains a flavonol core, myricetin, which is glycosylated on the 3 and 4′ positions (Fig. 1.3). The 3-hydroxyl carries the α-linked, linear trisaccharide D-glucopyranosyl-(β1→2)-D-glucopyranosyl-(β1→2)-L-rhamnopyranose. Attached to the central glucosyl sugar motif is a 6-O-caffeic ester. The 4′-hydroxyl carries the β-linked, linear disaccharide L-rhamnopyranosyl-(β1→4)-D-xylopyranose.

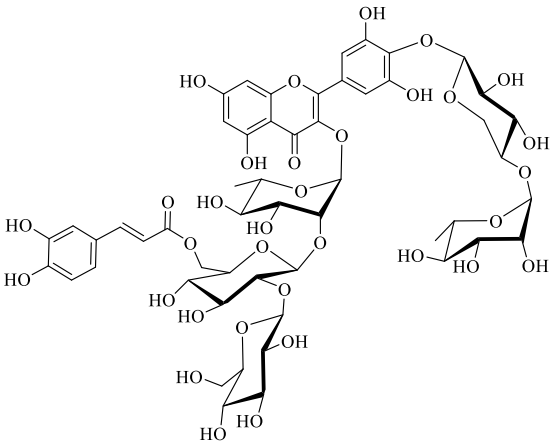


Figure 1.3: Structure of montbretin A.

Analysis of MbA's ability to inhibit sugar degradation enzymes showed activity specific for HPA and not intestinal wall α-glucosidases (Tarling *et al.*, 2008). Kinetic analysis of HPA inhibition identified the myricetin core as a competitive inhibitor with K_i of 110 μM and the caffeic acid motif as a non-competitive inhibitor with K_i of 1.3 mM (Tarling *et al.*, 2008). Structural binding studies with HPA and MbA degradation products identified the myricetin and caffeic acid moieties linked by the D-glucopyranosyl-(β1→2)-D-glucopyranosyl disaccharide component as the essential, high-affinity core structure (Williams

et al., 2015). X-ray structure binding analysis identified that MbA inhibits HPA through internal π-stacking interactions between the myricetin and caffeic acid, which organize their ring hydroxyls for optimal hydrogen bonding around HPA's catalytic residues (Williams *et al.*, 2015).

To assess MbA's potential as an oral T2D therapeutic, MbA was administered to Zucker diabetic fatty rats, an animal model of type 2 diabetes (Yuen *et al.*, 2016). When compared to animals receiving either Acarbose©, a common α-glucosidase inhibitor, or no treatment, chronic oral administration of MbA was found to be effective at decreasing BGL. Moreover, this study also showed that MbA improved the oxidative status of the fatty diabetic animals as well as lowered the levels of markers for increased risk of cardiovascular complications associated with diabetes. Overall, these results demonstrated that MbA is a strong candidate for further research as a T2D therapeutic for humans.

1.2.3 Flavonoids in Human Health

As a flavonoid, MbA is part of one of the largest and most diverse groups of specialized metabolites, estimated to contain 9,000 different structures (Ferrer *et al.*, 2008; Tohge *et al.*, 2013). With the exceptions of aurones, flavonoids have a common diphenylpropane (C6-C3-C6) backbone, which consists of two aromatic rings (A and B) connected by a central heterocyclic ring (C) (Fig. 1.4). Biosynthesized from phenylpropanoid and acetate-derived precursors, flavonoids are grouped into ten subgroups: anthocyanins, aurones, chalcones, condensed tannins, flavanones, flavones, flavonols, isoflavonoids, leucoanthocyanidins, and phlobaphenes (Ferrer *et al.*, 2008; Winkel-Shirley, 2001). While flavonoids play critical roles within plants, they also have various beneficial health activities as anti-inflammatory, anti-oncogenic, cardiovascular, and disease prevention agents in humans.

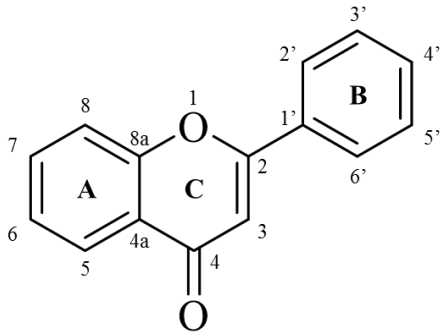


Figure 1.4: Backbone structure of flavonoids.

As anti-inflammatory agents, flavonoids reduce inflammation by preventing the ROS-based activation of transcription factors and cytokines important in triggering inflammation (Schreck *et al.*, 1991). While inflammation is a normal part of an immune response, prevention of chronic inflammation can reduce the risk for several degenerative diseases such as arthritis, atherosclerosis heart disease, or Alzheimer's disease (Brod, 2000; O'Byrne and Dalgleish, 2001). As anti-oncogenic agents, flavonoids play roles in both cancer prevention and inhibiting cancerous growth by interfering with a large number of regulatory pathways including those of growth, energy metabolism, apoptosis, cell division, transcription, and stress response (Gu *et al.*, 2005; Sarkar and Li, 2004). Through these actions, flavonoids have been able to affect signalling transduction pathways to prevent expression of tumor promoting factors (Atalay *et al.*, 2003) and exert cytostatic effects by activating proteins involved in programmed cell death (Richter *et al.*, 1999). As a cardiovascular agent, flavonoids reduce atherosclerosis and prevent arterial plaque build up (Frankel *et al.*, 1993; Tikkanen *et al.*, 1998). These activities are achieved, respectively, by promoting arterial muscle relaxation through stimulating release of muscle relaxants and inhibiting intracellular Ca^{2+} release needed for contraction (Ajay *et al.*, 2003; Carrón *et al.*, 2010), and preventing platelets and other lipids from sticking to lipoproteins through lipoprotein oxidation (Diaz *et al.*, 1997). As disease prevention agents, flavonoids play roles as immune modulators and anti-microbials. As immune system modulators, flavonoids have been observed to activate immune system cells, such as lymphocytes and macrophages, by stimulating their signalling cascades (Middleton Jr, 1998). As antimicrobials, flavonoids have been observed to act against both bacterial and viral organisms. While there is a current lack of understanding of the mechanism of action, it is believed that flavonoids play a role in inhibition of microbial polymerases or binding and inhibiting proper function of nucleic acid, membrane proteins, and capsid proteins (Cushnie and Lamb, 2005; Selway, 1986).

1.2.4 Flavonoid Backbone Biosynthesis

Flavonoids are derived from the aromatic amino acids phenylalanine and tyrosine, which come from the shikimate pathway (Knaggs, 2003). To convert either of these amino acids into flavonoids, they must first flow through the early steps of the phenylpropanoid

pathway (Fig. 1.4) (Tohge *et al.*, 2013; Winkel-Shirley, 2001). The first step of this is the elimination of ammonia in phenylalanine by phenylalanine ammonia lyase (PAL; EC 4.3.1.24) or tyrosine by tyrosine ammonia lyase (TAL; EC 4.3.1.23) to produce cinnamic acid and *p*-coumaric acid, respectively (Young and Neish, 1966). *p*-Coumaric acid may also be derived from the oxidation of cinnamic acid by cinnamic 4-hydroxylase (C4H; EC 1.14.13.11) (Russell and Conn, 1967). An additional oxidation of cinnamic acid by *p*-coumarate 3-hydroxylase (C3H; EC 1.14.13.-) yields caffeic acid (Kojima and Takeuchi, 1989). These three acids can then become "activated" by 4-coumarate-CoA ligase (4CL; EC 6.2.1.12) through the attachment of the coenzyme A (CoA) group (Gross and Zenk, 1974). These activated compounds can then be used to form a variety of phenolic-based metabolites such as lignins, lignans, coumarins, and stilbenoids, or continue in the biosynthesis of flavonoids.

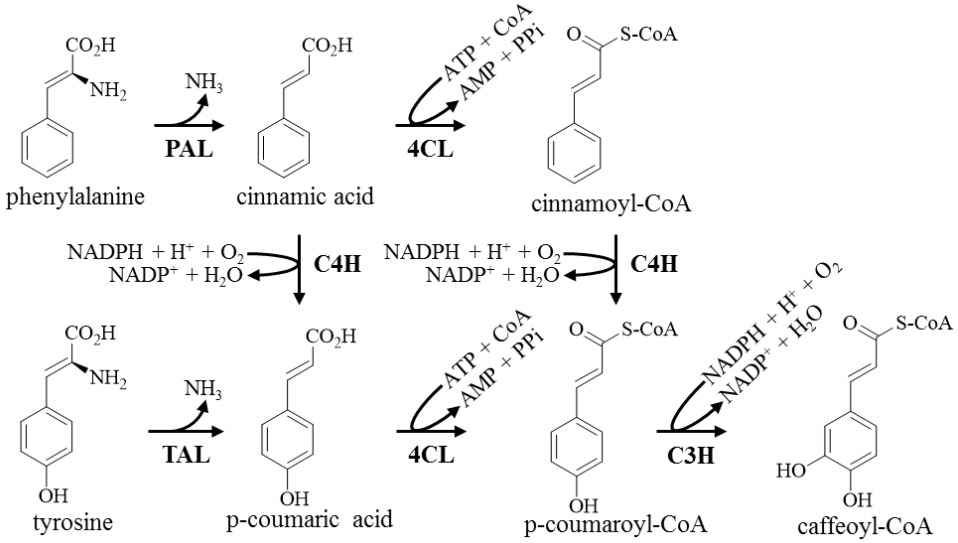


Figure 1.5: Biosynthesis of cinnamoyl-CoA, *p*-coumaroyl-CoA, or caffeoyl-CoA from phenylalanine and tyrosine. Abbreviations: PAL, phenylalanine ammonia lyase; TAL, tyrosine ammonia lyase; 4CL, 4-coumarate-CoA ligase; C4H, cinnamate 4-hydroxylase; C3H, coumarate 3-hydroxylase.

The polyketide chain extension of *p*-coumaroyl-CoA with three units of malonyl-CoA by chalcone synthase (CHS; EC 2.3.1.74) is the first dedicated step of the flavonoid biosynthesis pathway (Fig. 1.5) (Heller and Hahlbrock, 1980). The result is a chalcone, a polyketide that can be folded to generate the different flavonoids (Tohge *et al.*, 2013; Winkel-Shirley, 2001). This chalcone can then be converted into an aurone by aureusidin synthase (AS; EC 1.21.3.6) (Nakayama *et al.*, 2000) or undergo an isomerization catalyzed by chalcone

isomerase (CHI; EC 5.5.1.6), the next step in the biosynthetic pathway shared by the remaining flavonoids (Moustafa and Wong, 1967). This isomerization reaction involves a stereospecific ring closure of chalcones into their corresponding flavanones through an intermolecular nucleophilic attack of one of the phenolic hydroxyl groups onto the unsaturated ketone. This links the two aromatic rings through the formation of the C-ring which produces the flavanone naringenin (Jez and Noel, 2002). The flavanones represent one of the most important branching points in flavonoid metabolism. At this point, flavanones can go into one of five different branches of flavonoid biosynthesis to form the seven remaining types of flavonoids.

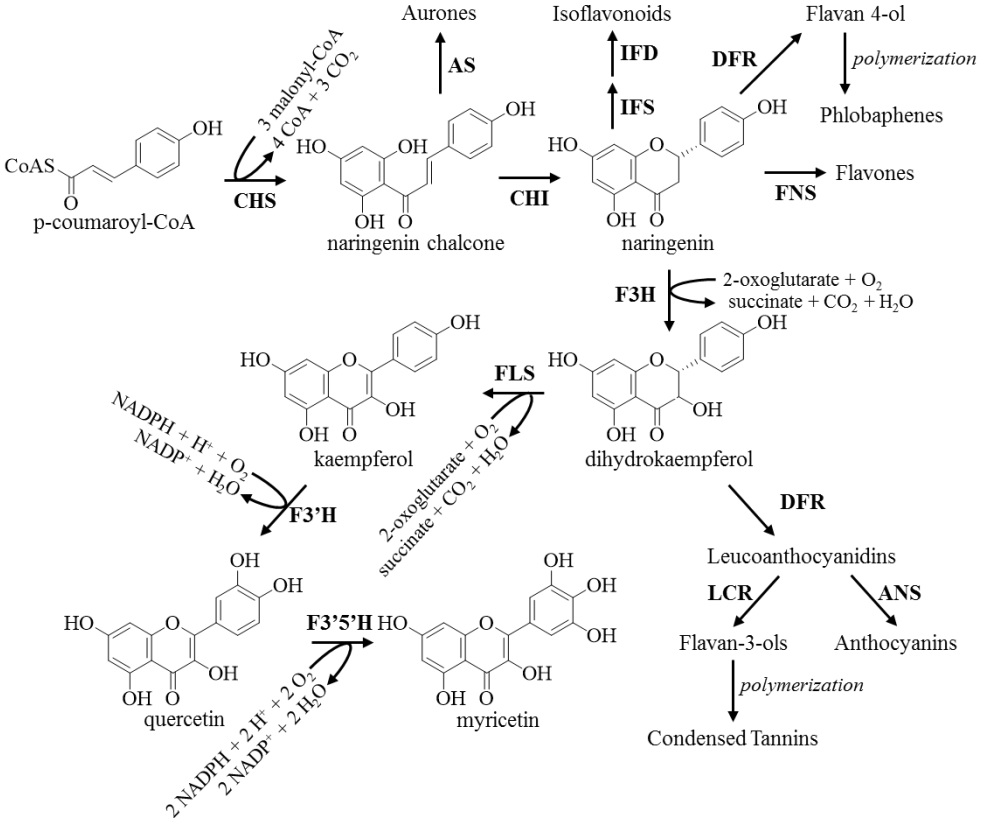


Figure 1.6: Biosynthesis of flavonoids from of *p*-coumaroyl-CoA and malonyl-CoA. Abbreviations: CHS, chalcone synthase; AS, Aureusidin synthase; CHI, chalcone isomerase; IFS, isoflavone synthase; IFD, 2-hydroxyisoflavanone dehydratase; DFR, dihydroflavonol 4-reductase; FNS, flavone synthase; F3H, flavanone 3-hydroxylase; ANS, anthocyanin reductase; LCR, leucoanthocyanidins reductase; FLS, flavonol synthase; F3′H, flavonoid 3′-hydroxylase; F3′5′H, flavonoid 3′5′-hydroxylase.

In the first branch, isoflavonoids are produced by a C2-C3 aryl migration and hydroxylation reaction on the flavanones catalyzed by isoflavone synthase (IFS; EC 1.14.13.136) (Steele *et al.*, 1999). This reaction is followed by a dehydration reaction of the

2-hydroxyisoflavanones catalyzed by 2-hydroxyisoflavanone dehydratase (IFD; EC 4.2.1.105) to form isoflavonoids (Hakamatsuka *et al.*, 1998). At this point, multiple additional enzymes can act on the metabolite to form a suite of isoflavonoids. In the second branch, flavanones undergo a dehydration reaction at the C2-C3 position catalyzed by flavone synthase (FNS; EC 1.14.11.22) to yield flavones (Martens *et al.*, 2001). In the third branch, a stereospecific C3 hydroxylation of naringenin by flavanone-3-hydroxylase (F3H; EC 1.14.11.9) produces dihydroflavonols (Forkmann *et al.*, 1980). These dihydroflavonols can undergo hydroxylation at the 3′ and 5′ position of the B-ring by flavonoid 3′-hydroxylase (F3´H; EC 1.14.13.21) or flavonoid 3′, 5′-hydroxylase (F3´5´H; EC 1.14.13.88) (Forkmann *et al.*, 1980; Menting *et al.*, 1994). Here, the pathway diverges into flavonols and anthocyanins. These dihydroflavonols can then undergo a C2-C3 reduction catalyzed by flavonol synthase (FLS; EC 1.14.11.23) to produce flavonols (Lukačin *et al.*, 2003). Alternatively, the dihydroflavonols can be reduced at the C4 position by dihydroflavonol 4-reductase (DFR; EC 1.1.1.219) to produce leucoanthocyanidins (Fischer *et al.*, 1988). These can then be converted into anthocyanins by the enzyme anthocyanin reductase (ANS; EC 1.14.11.19) (Saito *et al.*, 1999). In the fourth branch, a C4 reduction of leucoanthocyanidins by leucoanthocyanidins reductase (LCR; EC 1.17.1.3) produces flavan-3-ols (Tanner and Kristiansen, 1993). These flavan-3-ols then undergo a polymerization to form condensed tannins. However, it is not clear whether polymerization occurs enzymatically or non-enzymatically (Vogt, 2010). In the fifth branch, a C4 reduction of flavanones by DFR produces flavan-4-ols. Similar to condensed tannins, the polymerization process which produces phlobaphenes from flavan-4-ols is unclear.

1.2.5 *Crocosmia x crocosmiiflora*

Crocosmia spp. is a popular ornamental plant widely found in North America and Europe. A genus of perennial plants in the family *Iridaceae*, *Crocosmia spp.* is native to the South African grasslands (Fig. 1.7). Due to its growth versatility and attractive flowers, *Crocosmia spp.* is primarily cultivated for horticultural purposes. *Crocosmia spp.* typically has long (up to one meter), erect, sword-shaped leaves with distinct parallel veining and pleating. During flowering season (June – September in the Northern hemisphere), tall arching stems display funnel-shaped flowers in bright shades of brown, yellow, orange, or red (Manning and Goldblatt, 2008). This foliage grows from swollen underground modified stem

structures called corms that serve as storage organs and helps plants survive conditions such as drought, summer heat, or winter (Dominy *et al.*, 2008). These corms grow in vertical chains with younger corms forming atop older ones. The vertical chain of corms is fragile and can be separated into individual corms, each one able to form a new plant. Corms have contractile roots, which act to drag the corm deeper into the ground in response to temperature and light until it reaches uniform conditions (Kostelijk, 1984). Within this genus, the hybrid *Crocosmia x crocosmiiflora*, also known as montbretia, is one of the most commonly found members. Cultivation and cross breeding have resulted in over 300 different cultivars (Goldblatt *et al.*, 2004).

Figure 1.7: Examples of *Crocosmia x crocosmiiflora* cultivars. (a) Emily McKenzie cultivar. (b) Lucifer cultivar. (c) Emberglow cultivar.

In the past three decades, a number of specialized metabolites have been isolated and identified from the corms of different members of *Crocosmia spp.* Nagamoto *et al.* (1988) observed metabolites from water extracts of *Crocosmia x crocosmiiflora* corms possessing strong antitumor capabilities in mice with transplanted carcinomas (Nagamoto *et al.*, 1988). Although follow-up by Asada *et al.* to identify the compounds responsible for the noted antitumor activity identified the saponins crocosmiosides A – I (Asada *et al.*, 1989; Asada *et al.*, 1990) and montbretin A and B (Asada *et al.*, 1988), as well as masonosides A – C from the species *Crocosmia masonorum* (Asada *et al.*, 1994), none of these compounds were reported to display antitumor activity. Corms of a close relative of *Crocosmia*, *Tritonia crocosmaeflora*, were found to contain a napthazarin derivative, tricrozarin A, possessing broad-spectrum antimicrobial activity against gram-positive bacteria, yeast, and fungi (Masuda *et al.*, 1987). The same extraction also yielded another napthazarin derivative, tricrozarin B,

which showed antitumor capabilities against human and murine-based cancer cell lines (Masuda *et al.*, 1987).

Overall, these findings demonstrate that MbA is a strong candidate for further development as a T2D therapeutic for humans. To this end, one of the most significant problems is establishing a sustainable, economical source of montbretin A. Chemical synthesis of the natural product would be challenging due to the sheer size and complexity of the glycosylation pattern. Current available data reports that corms possess approximately 800 mg of MbA per kg of fresh weight (Andersen *et al.*, 2009). Based on estimates of an average patient needing up to 180 mg MbA/kg body weight per day (Andersen *et al.*, 2009) and that the plant must be killed to extract the majority of the MbA, natural harvest of *Crocosmia* can not currently be employed as a method for sustainable production. A solution to the production problem could be to harness the biosynthetic mechanism employed by *Crocosmia* to either produce a *Crocosmia x crocosmiiflora* system with heightened levels of MbA or an engineered microorganism capable of producing MbA. With little public data available on MbA or *Crocosmia spp.*, this approach requires extensive and laborious exploration and characterization of the MbA biosynthetic pathway.

1.2.6 Hypothesized Montbretin A Biosynthesis Pathway

To develop an MbA production system, its biosynthetic pathway in *C. x crocosmiiflora* must be characterized. Based on the chemical structure of MbA, I propose that the MbA biosynthetic pathway is composed of two subpathways: the early biosynthesis pathway (EPB) and late biosynthesis pathway (LBP).

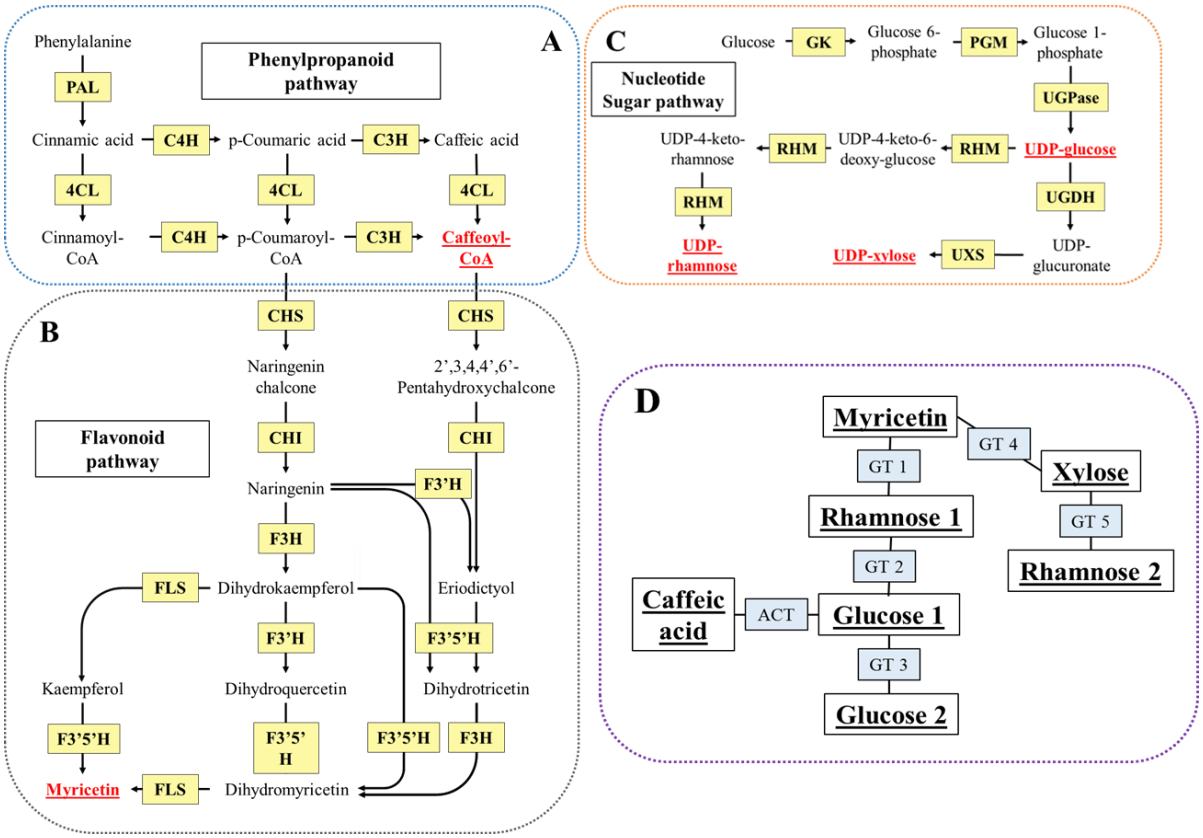


Figure 1.8: Proposed *in planta* biosynthesis of montbretin A. Within proposed pathway, sections A, B, and C correspond to the "Early Biosynthetic Pathway" while section D corresponds to the "Late Biosynthetic Pathways".

The proposed MbA EBP employs activity of the phenylpropanoid, flavonoid, and nucleotide sugar metabolism pathways towards the production of the individual components of MbA: myricetin, glucose, rhamnose, xylose, and caffeic acid (Fig. 1.8). Produced from phenylalanine or tyrosine in the phenylpropanoid pathway, caffeoyl-CoA or *p*-coumaroyl-CoA can enter the flavonoid pathway where they can be used to form the flavonol myricetin. In the nucleotide sugar pathway, glucose obtained from photosynthesis, sugar recycling, or storage is converted to uridine 5′-diphosphate-glucose (UDP-Glc). UDP-Glc can then be converted to UDP-rhamnose (UDP-Rha) or UDP-xylose (UDP-Xyl).

The hypothesized MbA LBP involves the assembly of the individual components of MbA (Fig. 1.8). In the LBP, myricetin is decorated with five monosaccharides through the activity of five glycosyltransferases (GTs), and the caffeic ester moiety is added through the activity of either an acyltransferase or a serine carboxypeptidase-like enzyme.

Of the enzymes proposed to be employed in the biosynthesis of MbA, those in the phenylpropanoid and flavonoid pathways are well characterized in several plant systems. Additionally, the EBP-specific end products, caffeoyl-CoA and myricetin, are readily accessible. While most of the nucleotide sugar pathway enzymes have been well characterized, UDP-xylose synthase (UXS) and UDP-rhamnose synthase (RHM) have only been characterized in a small number of plant systems. Additionally, due to patents on these genes (Bao *et al.*, 2014; Oka and Jigami, 2007) and complications with chemical synthesis, UDP-Xyl and UDP-Rha are prohibitively expensive. While genes performing 3-O- and 4′-O-glycosylations similar to those involved in the first steps of the LBP have been functionally characterized in other plant species, few genes have been characterized to perform reactions similar to the needed secondary glycosylation, tertiary glycosylations, and acylation reactions (D'Auria *et al.*, 2007; Hong *et al.*, 2007; Ko *et al.*, 2008; Moraga *et al.*, 2009; Trapero *et al.*, 2012; Yonekura-Sakakibara *et al.*, 2012). With the goal of studying MbA biosynthesis in *C. x crocosmiiflora* to establish a sustainable production system, emphasis will be put on identifying the UXS, RHM, and GTs involved in MbA biosynthesis.

1.2.7 UDP-Glycosyltransferases

Glycosylation reactions appear to be ubiquitous in nature (Bowles *et al.*, 2005; Bowles *et al.*, 2006). These reactions have a large range of effects on their aglycone's physicochemical properties such as altering solubility and stability, facilitating storage and compartment localization, molecular recognition, chemical defense, cellular homeostasis, and energy storage (Bowles *et al.*, 2005; Liang *et al.*, 2015). This, in turn, has important effects on a compound's functioning. Because of the importance which glycosylation modifications have on plant life and the large diversity of glycosylated compounds within species, most plants contain hundreds of glycosyltransferases (EC: 2.4.-.-), the enzymes responsible for performing these reactions. These enzymes have been classified into one of 94 families based on sequence similarity to genes in the Carbohydrate Active Enzyme database (CAZy, www.cazy.org) (Campbell *et al.*, 1997; Coutinho *et al.*, 2003; Lombard *et al.*, 2014). Within the glycosyltransferase enzyme class, family 1 glycosyltransferases (GT1) or uridine diphosphate glycosyltransferases (UGTs) (EC: 2.4.1.-) are the main players in the glycosylation of small molecules (Bowles *et al.*, 2005). While plants contain multiple glycosyltransferase families,

within a species the family 1 GT sub-class typically contains the most members (Caputi *et al.*, 2012). To date, hundreds of GT1 UGTs from various plants have been reported to glycosylate flavonoids, phenylpropanoids, terpenoids, benzoates, plant hormones, and many other metabolites (Caputi *et al.*, 2008; Lanot *et al.*, 2006; Lim *et al.*, 2002; Lim *et al.*, 2004; Poppenberger *et al.*, 2005). Current estimates from species with characterized genomes suggest that an average of 0.5% of hypothesized genes are GT1 UGTs (Caputi *et al.*, 2012).

1.2.7.1 UDP-Glycosyltransferase Structures

To date, x-ray crystal structures for eight plant GT1 UGTs have been reported (Brazier-Hicks *et al.*, 2007; Hiromoto *et al.*, 2015; Li *et al.*, 2007; Modolo *et al.*, 2009; Offen *et al.*, 2006; Shao *et al.*, 2005; Thompson *et al.*, 2017; Wetterhorn *et al.*, 2016). These structures all have the GT-B fold, which is characterized by N- and C-terminal domains that possess similar Rossmann-like folds, a unique structural motif found in many nucleotide-binding proteins. Comparison of the crystal structures shows that plant GT1 UGTs have a high degree of structural similarity, especially at the C-terminal domain which binds the sugar donor (Wang, 2009). Each Rossmann-like fold contains a central β-sheet flanked by α-helices on either side. These two domains are separated by a linker region, which is compacted to form an inter-domain cleft and the enzyme active site. When bound, the nucleotide sugar donor mainly interacts with the C-terminal domain while the acceptor mainly binds to the N-terminal domain.

In general, plant GT1 UGTs are between 400 – 550 amino acids long. A key signature characteristic of plant GT1 UGTs is the conserved 44 amino acid signature "plant secondary product glycosyltransferase" (PSPG) motif (Campbell *et al.*, 1997). The conserved amino acids of this motif are involved in hydrogen bond interactions with the nucleotide sugar donor. It has been seen that variations are more common in the N-terminal domain, particularly in the loops and helices of the active site (Wang, 2009). These are predicted to accommodate the diversity of potential acceptor substrates. During the reaction, binding of the sugar donor triggers a change from open to closed conformation, causing the acceptor and donor sugar to interact with both N- and C-terminal domains and undergo the transfer reaction.

1.2.7.2 UDP-Glycosyltransferases Catalytic Mechanism

Plant GT1 UGTs catalyze the transfer of sugar moieties from a donor by following a direct displacement S_N2-like reaction mechanism (Fig. 1.9) (Lairson *et al.*, 2008). Within the N-terminal domain, a conserved active site catalytic base (usually Asp, Glu, or His) acts on the target hydroxyl group of the acceptor molecule to remove a proton during glycoside bond formation. A nearby conserved aspartic acid interacts with the catalytic base by forming a hydrogen bond and balances its charge after deprotonating the acceptor. Concurrently, the acceptor attacks the C1 carbon atom of the sugar moiety of the UDP-sugar, resulting in the direct displacement of the UDP moiety and formation of the glycoside (Wang, 2009). Glycosylation reactions usually follow a sequential "bi-bi mechanism" in which the acceptor and donor are bound sequentially, followed by the sugar transfer, release of the newly glycosylated product, and finally release of the nucleotide moiety.

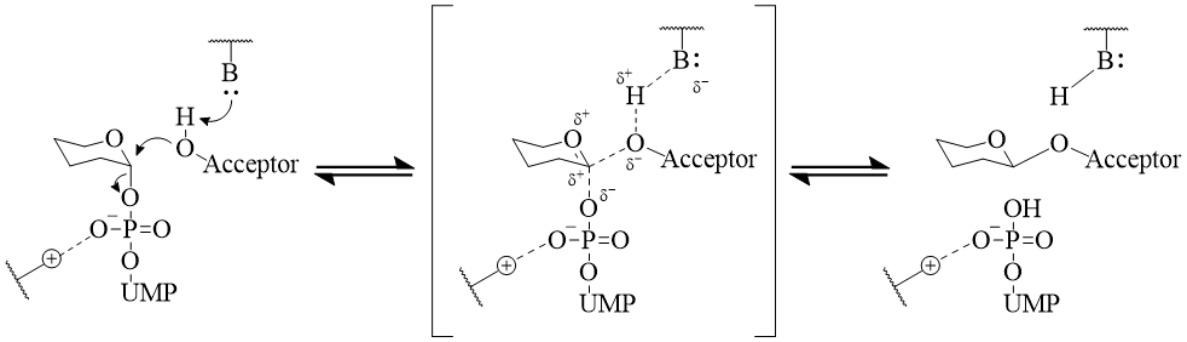


Figure 1.9: Basic UDP-glycosyltransferase catalytic mechanism for family 1 glycosyltransferases. Family 1 Glycosyltransferases (GT1 UGTs) follow a classic S_N2-like mechanism resulting in a inversion of anomeric stereochemistry upon glycosylation. An oxocarbenium-ion transition state (square brackets) forms with the help of a catalytic base usually provided by an active-site amino acid side chain. This base abstracts a proton from the hydroxyl group of the acceptor, facilitating nucleophilic attack at the sugar anomeric C1 carbon, forming a glyosidic bond between the sugar donor and the acceptor. The resulting negative charge on the phosphate group is typically stabilized by a positive amino acid side chain or helix dipole. "B" represents the catalytic base within the GT1 UGT while the "+" represents the positive amino acid side chain or helix dipole.

1.2.8 Nucleotide Sugar Interconversion Enzymes

Nucleotide sugar interconversion enzymes (NSE) are responsible for producing the majority of activated NDP-sugars with minor levels being formed from sugar salvage pathways and the promiscuous sugar phosphorylase SLOPPY (Bar-Peled and O'Neill, 2011; Kotake *et al.*, 2004). These sugars serve as donors for glycosyltransferases and provide critical substrates

needed to form the diverse set of glycan-specialized metabolites. Of the multiple small gene families that make up the NSE family, UXS and RHM are of interest for their value in the elucidation of the MbA biosynthetic pathway.

1.2.8.1 UDP-Xylose Synthase Structure

To date, only one x-ray crystal structure for UXS has been characterized (Eixelsberger *et al.*, 2012). The structural analysis showed UXS could be split into two domains. The N-terminal NAD^+-binding domain is a modified version of the classic Rossmann fold, built of a seven-stranded parallel β-sheet in between two arrays of α-helices. The C-terminal UDP-glucuronic acid (UDP-GlcA) binding domain is composed of two, two-stranded β-sheets and a three α-helix bundle. The cavity formed in between these two domains forms the enzyme active site.

In general, plant UXS are between approximately 350 – 450 amino acids in length. They contain two conserved motifs in the C-terminus: GXXGXXG and YXXXK. The GXXGXXG motif is characteristic of dinucleotide binding proteins and helps position and stabilize the substrate in the active site (Rossmann and Argos, 1978). The YXXXK motif contains most of the serine-tyrosine-lysine catalytic triad needed for enzyme activity (Duax *et al.*, 2000). Additionally, some UXS have been predicted to contain an N-terminal transmembrane domain (Harper and Bar-Peled, 2002).

1.2.8.2 UDP-Xylose Synthase Catalytic Mechanism

UDP-Xyl is synthesized from UDP-GlcA acid by UXS (EC 4.1.1.35) (Harper and Bar-Peled, 2002) (Fig. 1.10). Once the substrate and NAD^+ cofactor are bound, a catalytic tyrosine deprotonates the C4 hydroxyl of UDP-GlcA while the NAD^+ oxidizes the C4 carbon. These actions promote the formation of a C4 keto bond and produce the UDP-4-keto-hexauronic acid intermediate. The keto bond promotes stabilization of the carbanion upon decarboxylation, which is then followed by an NADH-facilitated reduction of the C4 carbon and proton return to the C4 hydroxyl group by the catalytic tyrosine.

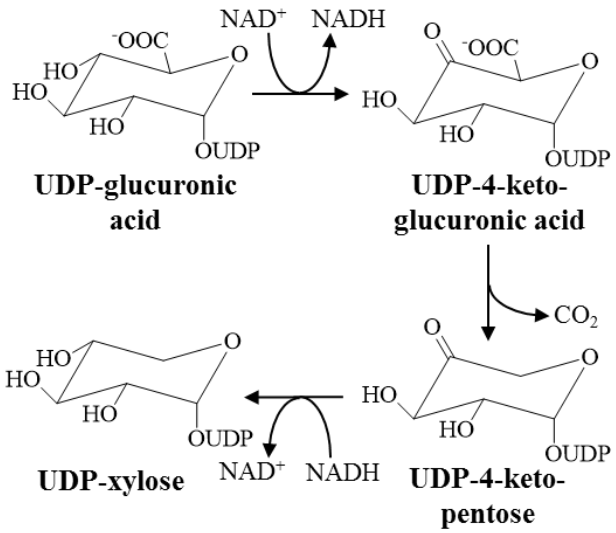


Figure 1.10: UDP-xylose synthase catalytic mechanism. UDP-xylose synthase decarboxylates UDP-GlcA in the presence of NAD^+ to produce the UDP-4-keto-pentose intermediate and NADH. UDP-4-keto-pentose in the presence of NADH is then converted to UDP-xylose and NAD^+.

1.2.8.3 UDP-Rhamnose Synthase Structure

RHM is predicted to have two main domains; an N-terminal domain which performs the dehydration reaction and a C-terminal domain which performs the epimerization and reduction reactions (Oka *et al.*, 2007; Watt *et al.*, 2004). As of writing, a single x-ray crystal structure of the C-terminal domain of RHM has been characterized (Han *et al.*, 2015). Similar to the characterized UXS, the C-terminal domain of RHM shows two distinct sub-domains. The first sub-domain, which plays a role in binding the NADPH and NADH cofactors, possesses a Rossmann-like fold. This fold possesses a central core composed of four parallel β-strands flanked by four α-helices. The second sub-domain is formed by five α-helices connected by two β-strands and several loops. A deep cleft is formed in between these two sub-domains and is surrounded by a bundle of helices and loops. The uncharacterized N-terminal domain is predicted to have a similar structure with two sub-domains and its active site in the cleft forming between them.

In general, plant RHM are between approximately 650 – 700 amino acids in length. Both the N- and C-terminus domains contain two conserved motifs: the GXXGXXG and YXXXK. The GXXGXXG motif is characteristic of dinucleotide binding proteins and helps position and stabilize the substrate in the active site (Rossmann and Argos, 1978). The

YXXXK motif contains most of the serine/threonine-tyrosine-lysine catalytic triad needed for enzyme activity (Duax *et al.*, 2000).

1.2.8.4 UDP-Rhamnose Synthase Catalytic Mechanism

In plants, UDP-Rha is synthesized from UDP-Glc by the single enzyme, RHM, through sequential dehydratase, epimerase, and reductase reactions (EC 4.2.1.76, EC 5.1.3.-, and EC 1.1.1.- respectively) (Fig. 1.11) (Kamsteeg *et al.*, 1978; Watt *et al.*, 2004). Once the substrate and NAD^+ cofactor are bound in the N-terminal active site, a catalytic tyrosine deprotonates the C4 hydroxyl of UDP-Glc while the NAD^+ oxidizes the C4 carbon. These actions promote the formation of a C4 keto bond and elimination of the hydroxyl, followed by addition of hydride to C6 to produce the UDP-4-keto-6-deoxy-α-D-glucose intermediate. This intermediate then leaves the N-terminal and enters the C-terminal active site.

Once the intermediate and the NADPH co-factor are in the C-terminus active site, epimerization at the C3 and C5 are predicted to proceed through an enediol/enolate intermediate (Liu and Thorson, 1994). In the epimerization, a general acid transiently protonates the C4 oxygen and stabilizes the enediol/enolate intermediate. A general base compliments this, removing a proton from C3 or C5 of the intermediate. Subsequently, reprotonation of the C3 and C5 occur from the opposite face of the sugar ring. Protonation of the C4 oxygen by the general base then completes the reaction. The NADPH-dependent reduction catalyzed at C4 of the 4-keto, 6-deoxy mannose ring is performed by the conserved Ser/Thr–Tyr–Lys catalytic triad. Here, these three residues play an analogous role to the dehydrogenase reaction where the lysine helps stabilize the co-factor so it can facilitate reduction of the C4 carbon and the catalytic tyrosine can facilitate proton transfer to the C4 hydroxyl group, resulting in UDP-Rha.

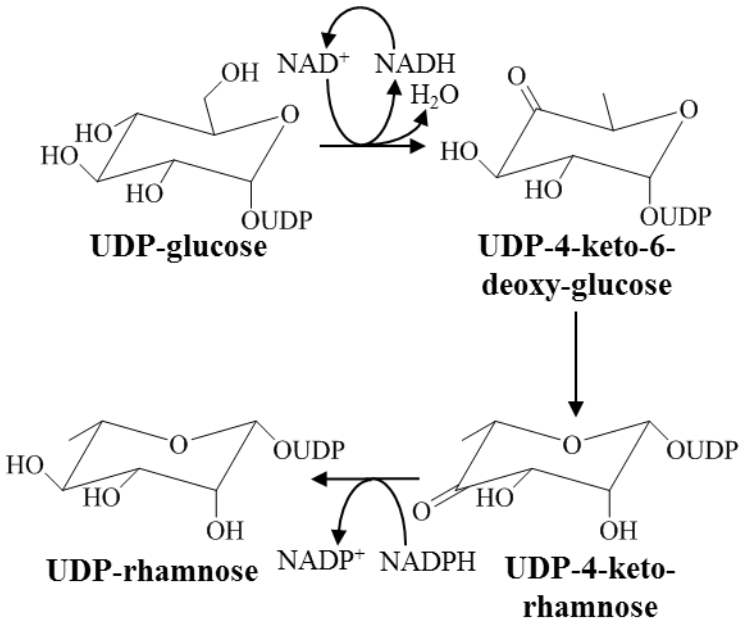


Figure 1.11: UDP-rhamnose synthase catalytic mechanism. In the N-terminus of RHM and in the presence of NAD^+, UDP-Glc has its C4 hydroxyl group oxidized and is dehydrated then reduced to form the UDP-4-keto-6-deoxy-glucose intermediate. After moving to the C-terminus of RHM, this intermediate then undergoes sequential epimerization reactions, to form the UDP-4-keto-rhamnose intermediate, and reduction by NADPH, to form UDP-rhamnose.

1.2.9 Pathway Elucidation Using Guilt-By-Association

The lack of information available for *de novo* systems such as *C. x crocosmiiflora* presents a large challenge for the elucidation of biosynthetic genes involved in a metabolic pathway. The integration of deep transcript and targeted metabolite profiles from corresponding plant tissues is a key component in establishing an integrated platform to select biosynthetic gene candidates involved in specialized metabolism (Facchini *et al.*, 2012; Saito *et al.*, 2008) The guilt-by-association principle proposes that a set of genes involved in a biological process are co-regulated and thus co-expressed under control of a shared regulatory system (Saito *et al.*, 2008). When extended to metabolite accumulation, this principle can be applied as a strategy for prioritization of genes putatively involved in metabolite biosynthesis. Although the assumed correlation between gene expression and metabolite accumulation has been shown to be acceptable (Urbanczyk-Wochniak *et al.*, 2003), their relationship can be non-linear or ambiguous (Gibon *et al.*, 2006). While the majority of successful applications of the guilt-by-association principle to elucidate target metabolite biosynthetic genes have been through pathway modulation through mutation or stress, success has also been achieved through comparison of different organs of the plant (Saito *et al.*, 2008).

1.3 (+)-3-CARENE SYNTHASE-LIKE FAMILY AND SITKA SPRUCE

1.3.1 Terpenoids in conifers

With estimates of 25,000 – 50,000 identified members, terpenoids constitute one of the most abundant and structurally diverse groups of specialized metabolite (Cheng *et al.*, 2007). In conifers, terpenes in the form of oleoresin play a critical role for defense against herbivores and pathogens. Oleoresin is a complex mixture of mostly monoterpenes (C_{10}), sesquiterpenes (C_{15}), and diterpene resin acids (C_{20}) (Keeling and Bohlmann, 2006a; Phillips and Croteau, 1999; Trapp and Croteau, 2001). Several oleoresin terpenoids have been shown to act against such diverse orgnaisms as bacteria (Himejima *et al.*, 1992), fungi (Kopper *et al.*, 2005; Paine and Hanlon, 1994), insects such as various bark beetles (Paine *et al.*, 1997) and weevils (Alfaro *et al.*, 2002; Tomlin *et al.*, 1996), or mammals (Phillips *et al.*, 1999). Terpenes act as direct as physical or chemical defenses when conifers are challenged by insects or pathogens. In addition to these direct effects, volatile terpenoids facilitate indirect tritrophic defense interactions with predatory or parasitic organisms of the aggressing pest (Keeling and Bohlmann, 2006b). Predatory and parasitic insects may use the terpenoids induced or constitutively produced by conifers to locate their herbivorous host (Pettersson, 2001; Raffa and Klepzig, 1989; Hilker et al., 2002; Mumm and Hilker, 2005; Grégoire et al., 1991; Grégoire et al., 1992).

1.3.2 Terpenoid Backbone Biosynthesis

Terpenoids are derived from the 5-carbon building blocks isopentyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP), which are produced in plants by two different pathways: the mevalonate (MVA) pathway and the 2-C-methyl-D-erythritol phosphate 5 (MEP) pathway (Fig. 1.12) (Gershenzon and Kreis, 1999). In plants, the MVA pathway is primarily responsible for the production of precursors for cytosolic terpenoids or isoprenoids, while the MEP pathway is primarily responsible for plastidial terpenoid or isoprenoid biosynthesis (Lichtenthaler, 1998).

The different prenyl diphosphate substrates that are used by terpene synthases to produce the array of terpenoids are produced by condensation reactions of DMADP with one, two or three units of IPP carried out by prenyltransferases (Ramos-Valdivia *et al.*, 1997).

These condensation reactions produce the acyclic prenyl diphosphates geranyl diphosphate (GPP), FPP, and geranylgeranyl diphosphate, (GGPP) (Ogura and Koyama, 1998; Ramos-Valdivia *et al.*, 1997) (Fig. 1.13). GPP is the substrate for plastidial monoterpene biosynthesis in conifers and other plants, where GPP is converted by monoterpene synthases (mono-TPS) into various acyclic and cyclic monoterpenes olefins or alcohols.

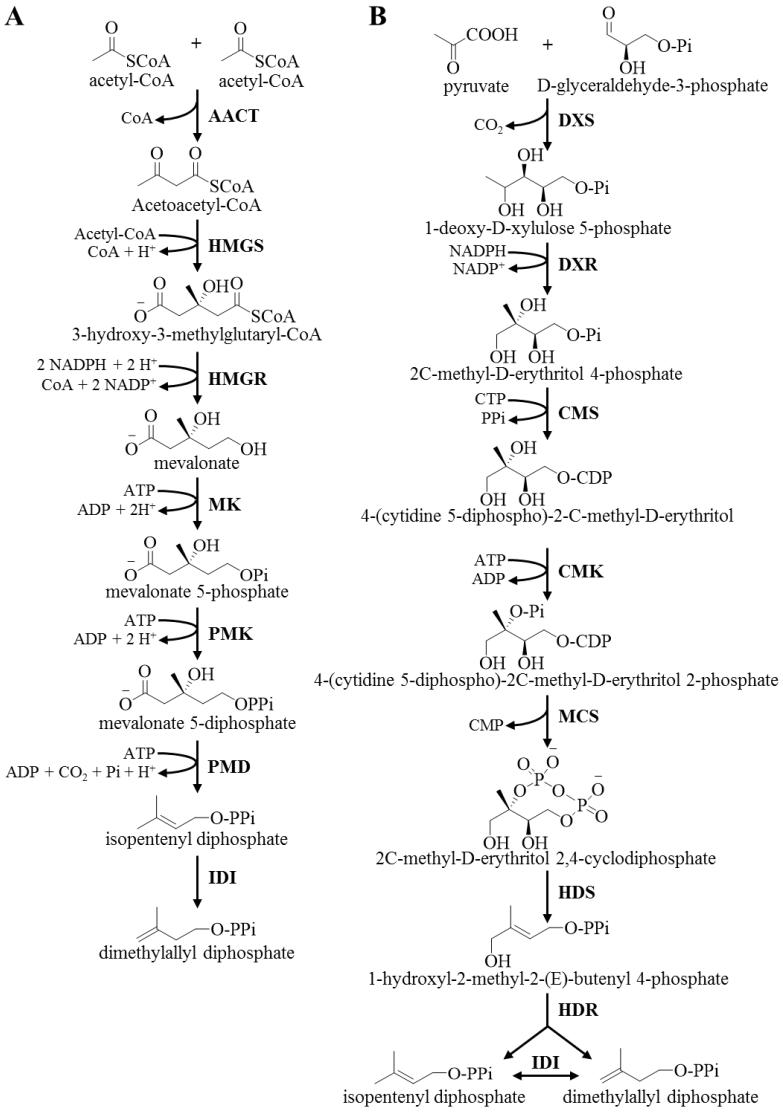


Figure 1.12: The MVA and MEP pathways. (A) The MVA Pathway. Abbreviations: AACT, acetyl-CoA acetyltransferase; HMGS, hydroxymethylglutaryl-CoA synthase; HMGR, hydroxymethylglutaryl-CoA reductase; MK, mevalonate kinase; PMK, phosphomevalonate kinase; PMD, phosphomevalonate decarboxylase; IDI, isopentenyl diphosphate isomerase. **(B)** The MEP Pathway. Abbreviations: DXS, 1-deoxyxylulose-5-phosphate synthase; DXR, 1-deoxy-D-xylulose-5-phosphate reductoisomerase; CMS, 4-diphosphocytidyl-2-C-methyl-D-erythritol synthase; CMK, 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase; MCS, 2-methyl-D-erythritol 2,4-cyclodiphosphate synthase; HDS, 1-hydroxyl-2-methyl-2-(E)-butenyl 4-diphosphate synthase; HDR, hydroxymethylbutenyl diphosphate reductase.

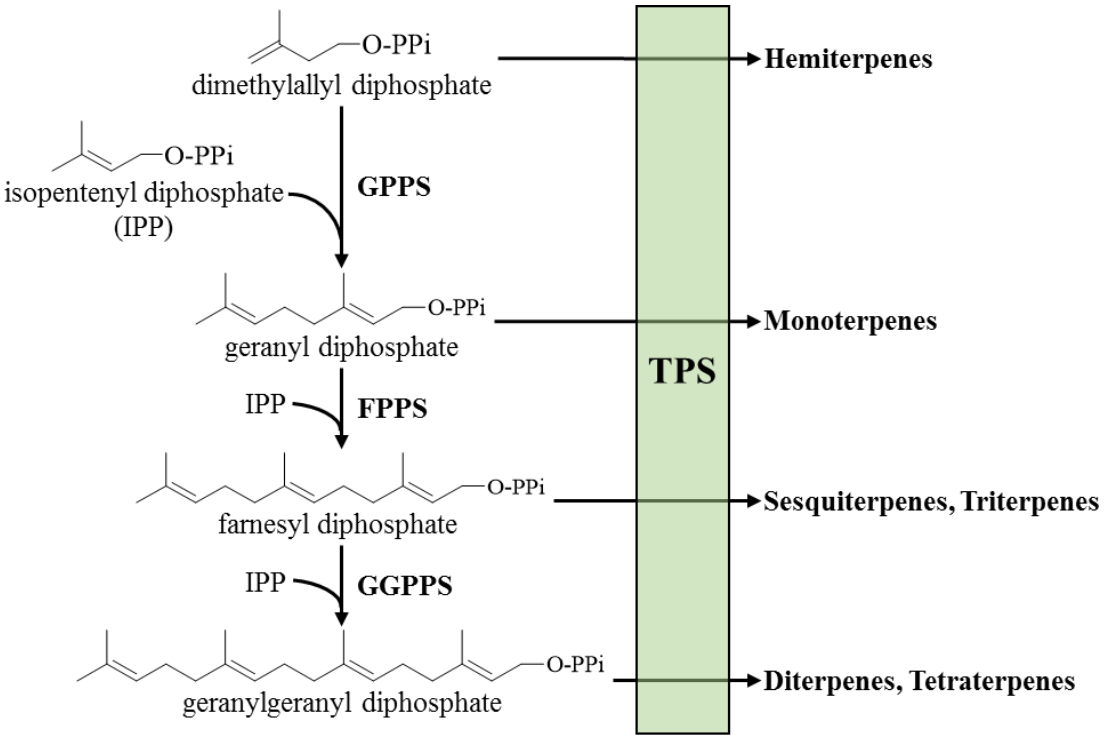


Figure 1.13: Biosynthesis of terpene precursors. Abbreviations: GPPS, geranyl diphosphate synthase; FPPS, farnesyl diphosphate synthase; GGPPS, geranylgeranyl diphosphate synthase; TPS, terpene synthase.

1.3.3 Monoterpene Synthases

Mono-TPS produce a variety of acyclic, monocyclic, and bicyclic structures from GPP (Davis and Croteau, 2000). Mono-TPS are primarily found in the TPS-b subfamily (angiosperm mono-TPS), the TPS-d subfamily (gymnosperm mono-TPS), the TPS-e/f subfamily (vascular plants), and the TPS-g subfamily (angiosperm mono-TPS) of the large plant TPS gene family (Chen *et al.*, 2011). Mono-TPS typically require a Mg^{2+} or Mn^{2+} as a co-factor (Croteau and Karp, 1979; Croteau *et al.*, 1980).

Plant mono-TPS are between approximately 600 – 650 amino acids in length (Bohlmann *et al.*, 1997; Colby *et al.*, 1993; Martin *et al.*, 2004). Mono-TPS contain three highly conserved motifs: the RRX_8W, DDXXD, and NSE/DTE motifs. Found in the N-terminus, the RRX_8W motif contains the tandem arginine pair thought to assist the initial diphosphate migration step of GPP cyclization. Found in the C-terminus, the DDXXD (Lesburg *et al.*, 1997; Tarshis *et al.*, 1994) and NSE/DTE (Caruthers *et al.*, 2000) motifs are critical in the bindings and positioning of the three metal ions needed for catalytic activity. In

addition to these, most mono-TPS contain a plastid-targeting sequence in the N-terminus (Whittington *et al.*, 2002).

To date, x-ray crystal structures of three different plant mono-TPS have been published: A *Salvia officinalis* (+)-bornyl diphosphate synthase (Whittington *et al.*, 2002), a *Salvia fruticosa* 1,8-cineole synthase (Hyatt *et al.*, 2007), and a *Mentha spicata* (4S)-limonene synthase (Kampranis *et al.*, 2007). Despite their low sequence similarity, these enzymes share a similar βα-didomain structure comprised entirely of α-helices and short connecting loops, suggesting the existence of a general monoterpene synthase fold (Lesburg *et al.*, 1997). Within this fold, the active site is found in a hydrophobic pocket of the α-domain and is composed of several α-helix loop regions. Upon binding the GPP substrate, these unstructured loops become ordered and form a protective cap over the active site, preventing premature quenching of carbocation reaction intermediates. While few functional elements have been found in the β-domain, mutational analysis has indicated this region might act as a scaffold to facilitate proper folding of the α-domain upon substrate binding (Kollner *et al.*, 2004).

The mechanism of mono-TPS follows a conserved core process (Fig. 1.14) (Cane *et al.*, 1982; Croteau and Felton, 1981; Croteau *et al.*, 1985a; Croteau *et al.*, 1985b; Croteau *et al.*, 1989; Wise *et al.*, 2001). For a mono-TPS to produce a cyclic monoterpene, it must first overcome the impediment to direct cyclisation caused by the geometry of the GPP C2-C3 double bond. This is achieved by initial ionization and isomerization of GPP to form linalyl diphosphate (LPP). C1-C6 ring closure is then achieved by an SN′ reaction and associated diphosphate departure to form the α-terpinyl cation (Cane *et al.*, 1995). Debate on what helps stabilize this carbocation still exists; current data suggests either the ionized diphosphate group or π-cation interactions facilitated by aromatic side-chains in the active site (Christianson, 2006). Further interactions between the cation intermediate, the mono-TPS, and additional substrates such as H_2O result in a series of hydride shifts, cyclizations, and/or hydroxylations to form the suite of potential products. For acyclic products, the reactions end through mono-TPS mediated deprotonation or water capture of either the geranyl cation or linalyl cation.

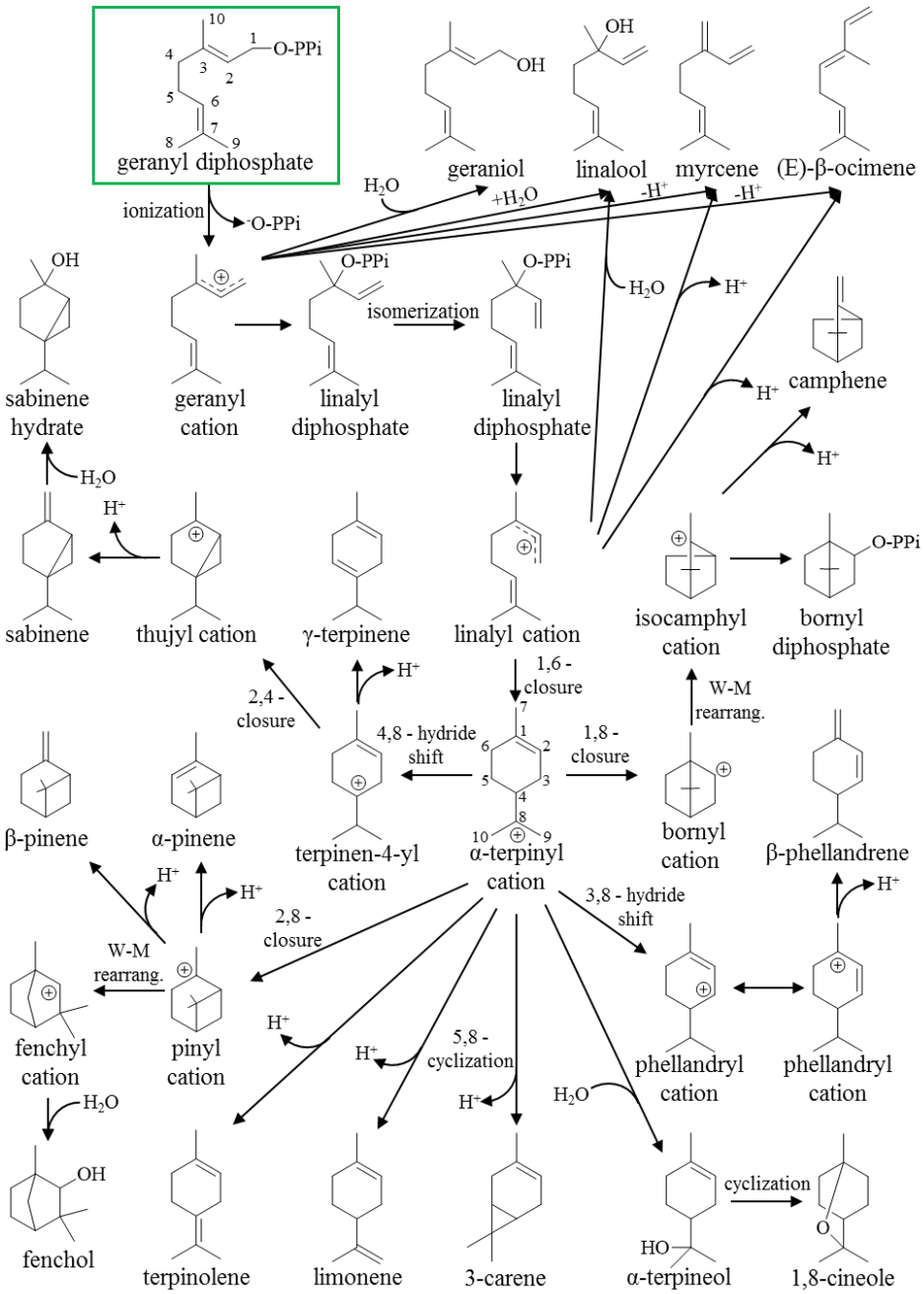


Figure 1.14: Monoterpene Synthase Catalytic Mechanism. The reaction mechanisms of all monoterpene synthases start with the ionization of the geranyl diphosphate substrate (green box). The resulting carbocation can undergo a range of cyclizations, hydride shifts and rearrangements before reaction is terminated by deprotonation or water capture. The formation of acyclic monoterpenes can proceed either through the geranyl cation or the linalyl cation. The formation of cyclic monoterpenes requires the preliminary isomerization of the geranyl cation to a linalyl intermediate capable of cyclization. The production of the initial cyclic species, the α-terpinyl cation, can then undergo further interactions between the monoterpene synthase and additional substrates, such as H_2O, to result in a series of hydride shifts, cyclizations, and/or hydroxylations to form the suite of potential products.

1.3.4 Monoterpene Synthases in Spruce and Insect Resistance

Conifers, and in particular species of spruce, are the most economically and ecologically dominant trees of the Canadian forest landscape. The white pine weevil (*Pissodes strobi*) is one of the most devastating pests of several spruce species (King *et al.*, 2004) (Fig. 1.15). Attack from weevils results in killing of the apical shoot tip, stem deformation, overall growth loss, and possible death of trees due to out-competition by surrounding vegetation.

Figure 1.15: Examples of white pine weevil damage to Sitka spruce.

The problem of weevil infestation is most severe with Sitka spruce (*Picea sitchensis*) in Western Canada. While most genotypes of Sitka spruce are susceptible to weevil attack, resistant trees have been identified and clonally replicated in field trials (King *et al.*, 2004; King and Alfaro, 2009). From these trials, the genotype H898 was identified as almost completely resistant. Conversely, the genotype Q903, originating from the Haida Gwaii Island of BC, was identified as highly susceptible to attack. In large replicated field trial, the levels of the diterpenoid resin acid dehydroabietic acid and the monoterpenes (+)-3-carene and terpinolene had a strong positive correlation with resistance to weevil attack (Robert *et al.*, 2010). Hall *et al.* (2011) used a combination of genomic, proteomic, and biochemical approaches to investigate the basis of variation of (+)-3-carene levels in two contrasting H898 and Q903 genotypes. This work identified that genotype-specific variations of gene copy number, transcript and protein expression, and catalytic efficiencies of members of a small family of (+)-3-carene synthase-like mono-TPS genes, containing the three (+)-3-carene synthases *PsTPS-3car1*, *PsTPS-3car2*, *PsTPS-3car3*, and the (–)-sabinene synthase *PsTPS-sab*, were responsible for the difference in (+)-3-carene levels (Hall *et al.*, 2011). Specifically,

the genomic presence, transcript and protein expression, and enzyme activity of the *PsTPS-3car2* gene accounted for much of the high levels of (+)-3-carene in the resistant genotype.

1.3.5 Terpene Synthase Evolution

Numerous structural and biochemical studies have showcased the ability of a protein to evolve novel activities or functions with only a small number of amino acid alterations (Aharoni *et al.*, 2005; Gerlt *et al.*, 2005; Khersonsky *et al.*, 2006). These residues associated with directing enzyme specificity are termed "plasticity residues" (Yoshikuni *et al.*, 2006) and are more often prevalent in or around enzyme active sites (Aharoni *et al.*, 2005; Aharoni *et al.*, 2004). This process of proteins developing promiscuous functions is believed to be a major driver of organisms developing new enzyme activity and specificity through divergent evolution (James and Tawfik, 2003). A unique aspect of TPS is that despite having a highly conserved active site scaffold composed largely of inert residues, most show promiscuous function. Catalytic specificity in TPS appears to be governed by the positioning of the polypeptide backbone and amino acid side chains on the active site surface, with supporting layers of surrounding residues playing a role in active site contour and dynamics (Greenhagen *et al.*, 2006). For this reason, TPS enzymes have been good candidates to investigate how plasticity residues contribute to divergent molecular evolution. The high sequence similarity, yet different product profiles of the different members of the Sitka spruce (+)-3-carene synthase-like family makes it an attractive target for investigating how plasticity and functional evolution can lead to an expansion of a species' specialized metabolism.

1.4 SCOPE OF THESIS

The central theme of this thesis is the elucidation and characterization of genes involved in specialized metabolism using two non-model plant systems. The first part of my thesis deals with MbA biosynthesis in *Crocosmia x crocosmiiflora*, a plant system with no prior supporting body of research on metabolism. This part of my thesis work will present examples of the three aforementioned phases of exploring how to harness specialized metabolism for human use (Fig. 1.2) in a new system. The second part of my thesis is dealing with elucidation of unique aspects of specialized metabolism of 3-carene biosynthesis Sitka spruce.

Chapter 2 presents an example of developing resources for a new plant system with the purpose of studying specialized metabolite biosynthesis. This chapter details work towards the development of *C. x crocosmiiflora* metabolite profile-, histological-, and transcriptome-based resources for exploring MbA biosynthesis. Chapters 3 and 4 present work characterizing part of the biosynthetic pathways of the MbA metabolite system. These chapters detail work towards the identification and characterization of candidate genes involved in the MbA biosynthetic pathway. Using the Sitka spruce system in Chapter 5, this work explores the amino acids critical for specific product profiles of a family of (+)-3-carene synthase-like enzymes to gain insight into the plasticity and functional evolution of this mono-TPS family. Overall, the research presented in this thesis acts to highlights opportunities, challenges, approaches and novel insights from non-model system research into specialized metabolite systems.

CHAPTER 2: DEVELOPMENT OF *CROCOSMIA* RESOURCES FOR THE ELUCIDATION OF THE MONTBRETIN A BIOSYNTHESIS PATHWAY

As part of the initial strategy to identify and characterize the biosynthetic pathways of montbretin A in *Crocosmia x crocosmiiflora*, I developed a series of biological, metabolite profiling, and transcriptome resources for this previously unstudied system. Metabolite profiling of montbretin A accumulation within *C. x crocosmiiflora* identified that montbretin A primarily accumulates in the corm, the below ground storage and overwintering organ, with minor accumulation in the flower, stem, and stolon organs. Matrix-assisted laser desorption ionization and additional metabolite profiling analyses showed that within the corm, montbretin A is primarily found in the peripheral tissues with levels significantly higher than in tissues of the central vascular cylinder. I used 16 organ-specific *C. x crocosmiiflora* samples to generate a first assembly of the *C. x crocosmiiflora* transcriptome containing 77,894 unigenes. *In silico* annotation revealed a lack of high-quality gene annotations of closely related species in the public database. Employing a homology-based search approach, I identified candidate genes for each step in the proposed montbretin A early biosynthetic pathway. Integrating montbretin A accumulation and gene expression data resulted in the identification of 14 genes for the proposed montbretin A late biosynthetic pathway. To the best of my knowledge, this is the first report of omics-based resource development for the *Crocosmia* genus.

2.1 INTRODUCTION

Historically, establishing the molecular and biochemical basis of specialized metabolite biosynthesis in plants often required enzyme purification and testing for specific enzymatic activity as a starting point for gene discovery. While resources for gene discovery in model organisms have proliferated in recent years, many specalized metabolites have a narrow taxonomical distribution (Sumner *et al.*, 2015). As such, elucidation of target biosynthetic pathways genes and enzymes in non-model systems usually cannot solely rely on a homology-based approach employing model systems. In these situations, a critical first step is the development of species-specific resources (Facchini *et al.*, 2012; Saito *et al.*, 2008).

The *Crocosmia* genus is an excellent example of a non-model plant system. At the beginning of this work, the majority of published research in this system was of a horticultural nature with only six non-horticultural papers available, all of which reported on new specialized metabolites or their functions (Asada *et al.*, 1989; Asada *et al.*, 1990; Asada *et al.*, 1994; Masuda *et al.*, 1987; Nagamoto *et al.*, 1988; Tarling *et al.*, 2008). From these papers, it was known that montbretin A (MbA) can be isolated from *Crocosmia* corms. However, it was unknown which organ(s) or tissue(s) produce MbA, if MbA is being transported to site(s) not involved in biosynthesis, or if MbA biosynthesis is constitutive or occurs under specific inducing conditions. A search of the public domain showed only nine reported *Crocosmia* nucleotide sequences, all of which are plastidial genes (Schaefer *et al.*, 2011; Souza-Chies *et al.*, 1997). Currently, few species closely related to *Crocosmia* have public, high-quality genomic data available. The closest genera with an available genome is the orchid *Phalaenopsis equestris*, which is a member of the Asparagales order with *Crocosmia* (Cai *et al.*, 2015) (Fig. S2.1). More distantly related, multiple grasses within the Commelinids order, such as *oryza*, *sorghum*, *brachypodium*, and *mays*, have genomes available (Goff *et al.*, 2002; Paterson *et al.*, 2009; Schnable *et al.*, 2009; Vogel *et al.*, 2010). The closest genera with comprehensive transcriptomic data available are *Crocus* (Baba *et al.*, 2015; Jain *et al.*, 2016) and *Iris* (Ballerini *et al.*, 2013), both members of the Iridaceae family with *Crocosmia*. This lack of scientific information about the *Crocosmia* genus or MbA makes the goal of elucidating genes involved in the biosynthetic pathway of MbA especially challenging.

Recent advances in the fields of genomics, transcriptomics, and metabolomics provide improved processes for establishing a pool of candidate genes of a novel biosynthetic process (Facchini *et al.*, 2012). The two strategies most commonly employed for identifying such candidates are homology to genes with similar functions and correlating metabolite abundance with transcript expression profiles. Multiple studies have shown these approaches can be used effectively for the identification of specific pathway genes in the investigation of specialized metabolite biosynthesis (Attia *et al.*, 2012; Keeling *et al.*, 2011; Liscombe *et al.*, 2009; Xiao *et al.*, 2013; Zerbe *et al.*, 2013).

The biosynthesis of MbA is proposed to occur in two parts: the early biosynthesis pathway (EBP) and the late biosynthesis pathway (LBP) (Fig. 1.8). The EBP employs genes of the phenylpropanoid, flavonoid, and nucleotide sugar metabolism pathways and results in

the formation of the individual building blocks and enzyme substrates of MbA: myricetin, UDP-glucose, UDP-rhamnose, UDP-xylose, and caffeoyl-CoA. Because genes with functions involved in the EPB are well characterized in other systems, it is hypothesized that identification of candidate genes through a homology-based approach will be an efficient method. The LBP involves the assembly of these individual components into MbA. Because most of the predicted functions of the LBP have not been reported in characterized genes from other species, it is proposed that the integration of transcript and targeted metabolite profiles will be an efficient method for identification of candidate genes. Employing these two approaches for identifying candidate MbA biosynthesis genes requires the development of *Crocosmia*-specific resources.

Metabolite profiling and transcriptome sequencing (RNA-seq) are two of the most commonly employed techniques for the identification of novel biosynthetic pathway genes. Metabolite profiling plays two major roles in the elucidation of specialized metabolite biosynthesis: (i) identification of spatial and temporal distribution patterns as effected, for example, by plant development and environmental cues, and (ii) identification of potential intermediates of the biosynthetic process. Complementary to this profiling, an understanding of the morphology of the plant can provide valuable information on the location of metabolite biosynthesis as well as the *in planta* biological function. Next-generation sequencing (NGS) technologies have revolutionized biological research by providing rapid and reliable sequence data. Of the many tools in the NGS portfolio, RNA-seq has proven particularly useful for non-model organisms lacking a reference genome (Wang *et al.*, 2009b). Not requiring prior knowledge of gene sequences, RNA-seq coupled with sequence assembly algorithms can be a powerful tool for the discovery of novel transcripts and *in silico* estimation of gene expression.

The goal of this chapter is to develop and characterize the first set of biological, metabolite profile, and transcriptome resources for *Crocosmia* to serve as a foundation for research into the MbA biosynthetic pathway. A combination of liquid chromatography- mass spectrometry (LC-MS), matrix-assisted laser desorption ionization (MALDI), and histological analyses were used to explore spatial and temporal MbA accumulation levels within *C. x crocosmiiflora*. RNA-seq was used to develop the first *C. x crocosmiiflora* draft transcriptome. Extensive manual annotation of the transcriptome and the integration with metabolite profiling resources was used to identify a pool of candidate genes involved in MbA biosynthesis. To

the best of my knowledge, this is the first report of omics-based resources developed for the *Crocosmia* genus.

2.2 EXPERIMENTAL

2.2.1 Plant Material

To start the *C. x crocosmiiflora* colonies, starter plants were obtained from the garden of Dr. Gary Brayer in Richmond, British Columbia, Canada on July 9th, 2010. The variety of the obtained *C. x crocosmiiflora* was identified by Dr. Gary Brayer as "Emily McKenzie" by comparing phenotype to (i) horticultural references (Goldblatt *et al.*, 2004; Kostelijk *et al.*, 1984) and (ii) against varieties of *C. x crocosmiiflora* commercially available. Individual corms were separated and potted in 4L planting pots with perennial soil at least 4 inches deep. Plants were grown in shaded areas of the University of British Columbia Horticulture Greenhouse (6394 Stores Road V6T 1Z4) outdoor patio year round. A sample has been submitted to the UBC Herbarium (http://www.biodiversity.ubc.ca/museum/herbarium/) under accession numbers V244885a and V244885b. Each November, the decaying above-ground tissues were cut approximately one inch above the soil. During the winter months (November – February), plants would be left outside with a layer of mulch on top of planter pots to act as insulators and prevent frost from reaching corms. Every second year in January, plants were removed from pots and repotted so each would contain only a single corm.

2.2.2 Metabolite Analysis

For temporal and spatial analysis of MbA levels in different parts of the plant (section 2.3.1), samples were collected at 17 time points between May 29th, 2012 and May 1st, 2013. At each time point, whole plants were dug up, soil was removed, and plants were separated into six organs: flower, stem, leaf, corm, stolon, and root (Fig. S2.2). Flower samples were cut at the base of the pedicel. Leaf samples were cut 2.5 cm away from the stem. Stem samples were cut 2.5 cm away from the corm and inflorescence and had all leaf tissue coating the stem removed. Stolon and root samples were cut one inch away from the corm. Corm samples had the remaining stem cut off as well as the tunic, basal plate, lateral growth tissue, and remaining stolon and root tissue removed. Once collected, samples were immediately frozen in liquid

nitrogen and stored at -80°C. Analyses were done with three biological replicates. Statistical analysis was performed using the Single Variable ANOVA data analysis in EXCEL.

For spatial analysis of MbA within corms, samples were collected on April 18th 2016. Corms were harvested as above, cut along the same plane into three segments. One segment was used for metabolite analysis while the other two were used for *in situ* and histological analyses (see sections 2.2.3 and 2.2.4). The segment was further divided into seven concentric sections using Boekel cork borers (http://www.boekelsci.com/) (Fig. 2.4). The first section was always the central vascular cylinder with the endodermis removed. Subsequent sections were produced using borers with a 3 mm increase in diameter. Once collected, samples were immediate frozen in liquid nitrogen and stored at -80°C.

Metabolites were extracted from homogenized tissue with 50% methanol (5 mL/g tissue) for 24 hours at 40°C. Samples were passed through a 0.22 µm hydrophilic polypropylene membrane filter (http://www.pall.com/). Metabolites were identified by liquid chromatography (LC) (Agilent 1100 Series)/mass spectrometry detector (MSD) Trap (XCTplus) by comparison of retention times and mass spectra with authentic standards. An Agilent ZORBAX SB-C18 column (4.6 mm internal diameter, 50 mm length, 1.8 µM pore size) was used with a temperature of 50°C and flow rate 0.8 mL min^{-1}. The mobile phase used was a combination of two solvents: solvent A (H_20 + 0.2% formic acid) and solvent B (acetonitrile + 0.2% formic acid). The mobile phase run was 95% solvent A by 0.5 min, 80% solvent A by 5 min, 10% solvent A by 7 min, and 95% solvent A by 7.10 min, and held for 2.9 min, giving a total run time 10 min. Diode array detector (DAD) monitored wavelengths at 266 nm and 326 nm. The mass spectrometer mode was negative electrospray with nebulizer pressure 60 psi, dried gas rate 12 L min^{-1}, dry temp 350°C, and a m/z scanning range between 50 – 2000. Quantification of MbA levels was based on external standard curves using purified MbA obtained from Dr. Stephen Withers.

2.2.3 Matrix-Assisted Laser Desorption Ionization (MALDI) Analysis

Using corm segments produced on April 18th, 2016 (section 2.2.2), MALDI imaging was performed at the University of Victoria Genome BC Proteomic Centre (http://www.proteincentre.com/). Samples were cut into 40 µm thick sections with a Microm HM500 at -20°C cryostat and thaw-mounted onto indium-tin oxide (ITO)-coated microscopic

glass slides. The matrix solution containing 2-mercaptobenzothiazole (2-MBT) was prepared at 10 mg/mL in 80% aqueous methanol containing 2% formic acid. Corm sections were spray coated using a Bruker Daltonics ImagePrep electronic matrix sprayer, with the application of MCAEF (Wang *et al.*, 2015). MALDI-mass spectrometry data was acquired in negative ion (-) mode on a Bruker Apex-Qe 12-Tesla hybrid quadrupole-fourier transform-ion cyclotron resonance instrument equipped with a 355-nm smartbeam UV laser. For tissue imaging, a laser raster step size of 300 μm and laser beam size of ca. 200 μm was used. Mass spectral datasets were processed with Bruker DataAnalysis, and the MS ion images were constructed with Bruker FlexImaging.

2.2.4 Corm Histology Analysis

Corm segments produced on April 18th, 2016 (section 2.2.2) were placed into Formalin-Acetic Acid-Alcohol fixative at 4^{o}C for three weeks. Samples were submitted to the Wax-It Service Laboratory (Wax-it Histology Services Inc., http://www.waxitinc.com/) for dehydration in an ethanol series and paraffin embedding. Paraffin-embedded samples were thin-sectioned (20 μm) and stained with three different methods. Sections were stained for the presence of lignin and suberin by staining in 0.1% (w/v) berberine hemisulphate (Sigma-Aldrich Co., http://www.sigmaaldrich.com/) in distilled water (dH_2O) for 1 hour and rinsed in dH_2O for 30 min. Sections were then counterstained by immersion in 0.5% (w/v) aniline blue (Sigma-Aldrich Co., http://www.sigmaaldrich.com/) in dH_2O for 30 min, rinsed in dH_2O for 30 min, and mounted on slides in 0.1% (w/v) FeCl3 in 50% glycerol (Cholewa and Griffith, 2004). Sections were stained for the presence of starch granules by staining in 10% Lugol's Solution (Sigma-Aldrich Co., http://www.sigmaaldrich.com/) for 5 minutes and rinsed in dH_2O for 30 min (Gurr, 1965). Sections were stained for the presence of lignin in the sclerified cells by staining with phloroglucinol in 20% HCl for 5 minutes and rinsed in dH_2O for 30 min (Jensen, 1962). Microscopy was performed using a Zeiss Axioplan 2 (Carl Zeiss, Ontario, Canada) equipped with an X-Cite Series 120 Q light source (Excelitas Technologies Corp., Massachusetts, USA) for epifluorescence. Sections were viewed with white and ultraviolet light. The ultraviolet filter set consisted of the BP365 exciter filter, the FT395 chromatic beam splitter, and the 397-barrier filter. Images were captured on Hamamatsu Orca Flash 4.0 LT camera (Hamamatsu, Shizuoka Pref., Japan)

2.2.5 RNA Isolation

Samples for RNA-sequencing were collected on July 29th, 2013 as described in section 2.2.2, immediately frozen in liquid nitrogen, and stored at -80°C until RNA was extracted. Total RNA was extracted as described (Mageroy *et al.*, 2015) and stored at -80°C. Total RNA concentration was determined using a NanoDrop 1000 (Thermo Scientific, http://www.thermoscientific.com/) and RNA Integrity Number (RIN) was assessed as a quality control with an Agilent 2100 Bioanalyzer and Agilent RNA 6000 Nano Kit LabChips (Agilent Technologies Inc., http://www.agilent.com/). Sixteen RNA samples (three each for flowers, leaves, stems, and stolons and four for corms) with RIN greater than 8.00 were sent to the McGill University and Génome Québec Innovation Centre (http://gqinnovationcenter.com) to generate approximately 75 gigabases (Gbp) of sequencing data using the Illumina HiSeq2000 platform with 150bp paired-end sequencing.

2.2.6 *De Novo* Transcriptome Assembly

16 paired-end RNA-Seq libraries from five different *C. x crocosmiiflora* organs were generated with the Illumina HiSeq2500 platform. Raw Illumina reads were cleaned using Trimmomatic (Bolger *et al.*, 2014) to remove the 15 bp adapter sequence at front ends of sequences and tested for quality using FastQC (www.bioinformatics.babraham.ac.uk/projects/fastqc/). The raw sequences were then filter for plastidial sequences using Bowtie. bbMerge, from the software suite BBMap (https://sourceforge.net/projects/bbmap/files/), was used to pre-assemble overlapping paired-end reads to generate longer, single end read for improved assembly contiguity. Merged single-end reads and un-merged paired-end reads of all tissues were then pooled together and assembled *de novo* by Trinity assembler version 2.1(Grabherr *et al.*, 2011). Predicted peptides were established by TransDecoder (Haas *et al.*, 2013). Coding sequences (CDS) at 98% nucleotide similarity or greater were deemed to be allelic variants and were clustered, by Cd-hit, for downstream differential expression analysis (Li and Godzik, 2006).

2.2.7 Assessment of *Crocosmia x crocosmiiflora* Unigene Dataset

The Core Eukaryotic Gene Mapping Approach (CEGMA) dataset (Parra *et al.*, 2007) and the Benchmarking Universal Single-Copy Orthologs (BUSCO) dataset (Simao *et al.*, 2015) were used to quantitatively estimate the completion of *C. x crocosmiiflora* unigene dataset. *C. x crocosmiiflora* orthologs of the 248 and 956 highly conserved ortholog sets, respectively, were identified by performing a BLASTx of these datasets against the unigene set with an e-value threshold of $1e^{-20}$.

2.2.8 Annotation and Classification of Unigenes Dataset

The unigene dataset was annotated against the NCBI Non-Redundant (NR) and Cluster of Orthologous Group for eukaryotic complete genomes (KOG) (Koonin *et al.*, 2004) databases using a BLASTx search with an e-value threshold of $1e^{-10}$. Additionally, Gene Ontology (GO) terms were assigned based on the best BLAST search hits against the NR database using the Blast2GO program (https://www.blast2go.com/) (Conesa *et al.*, 2005) with an e-value threshold of $1e^{-10}$. Unigenes were assigned to the "biological process", "molecular functions", and "cellular components" ontologies. The distribution of those unigenes' functional category was summarized using WEGO software (Ye *et al.*, 2006). To analyze specialized metabolism pathways active in the *C. x crocosmiiflora* samples, gene annotations of unigene set were compared to the reference canonical pathways in KEGG.

2.2.9 cDNA Cloning of Early Biosynthetic Pathway Genes

Using the isolated RNA that was submitted for RNA sequencing (Section 2.2.5), cDNA was synthesized using a Maxima First Strand cDNA Synthesis Kit (ThermoFisher, https://www.thermofisher.com) and quantified using a NanoDrop 1000 (Thermo Scientific, http://www.thermoscientific.com/). Genes of interest were amplified and cloned into pJet1.2 vectors (Fermentas, http://www.fermentas.com/) from cDNA using primers found in Table S2.1. Sequences and gene insertion orientation were verified by Sanger sequencing.

2.2.10 Haystack Analysis

Identification of unigenes with expression pattern correlating to MbA accumulation patterns was performed using the Haystack program (http://haystack.mocklerlab.org/)

(Michael *et al.*, 2008; Michael *et al.*, 2008; Mockler *et al.*, 2007). MbA accumulation levels within each of the 16 samples submitted for sequencing was analyzed as outlined in in section 2.2.2 and used as the model file. The reads per kilobase of transcript per million mapped reads (RPKM) for each unigene within each of the 16 samples was used as the corresponding data file. Parameters were as follows: correlation cut-off of 0.8, fold cut-off of 2, P-value cut-off of 0.05, and background cut-off of 1.

2.3 RESULTS

2.3.1 Temporal Accumulation Patterns of Montbretin A within *C. x crocosmiiflora*

Previous work reported corms as the predominant site of MbA accumulation in plants of the genus *Crocosmia* (Andersen *et al.*, 2009). To confirm this and explore how MbA accumulation patterns change within *C. x crocosmiiflora* over the course of one year, accumulation levels within six of the major organs were profiled every three weeks for twelve months (Fig. 2.1; Table S2.2). Results showed that throughout the year, MbA primarily accumulated in the corms with levels ranging between 1.85 – 4.31 mg/g fresh tissue (FT). The only other organ containing notable levels of MbA was the flowers with levels ranging between 0.027 – 0.145 mg/g FT (0.75% – 4.48% of corm levels). The stem and stolon contained only trace amounts of MbA with between 0.003 – 0.047 mg and 0.002 – 0.013 mg/g FT (0.75% – 1.45% and 0.06% – 0.44% of corm levels, respectively). No MbA was detected in the leaf or root. Single factor ANOVA analyses identified that levels of MbA were not statistically different between the corm and stolon organs. The same analysis on the stem and flower organs identified a statistical difference in MbA levels throughout the year. These results confirmed that despite high levels of biological variation, corms are the primary site of MbA storage in *C. x crocosmiiflora*.

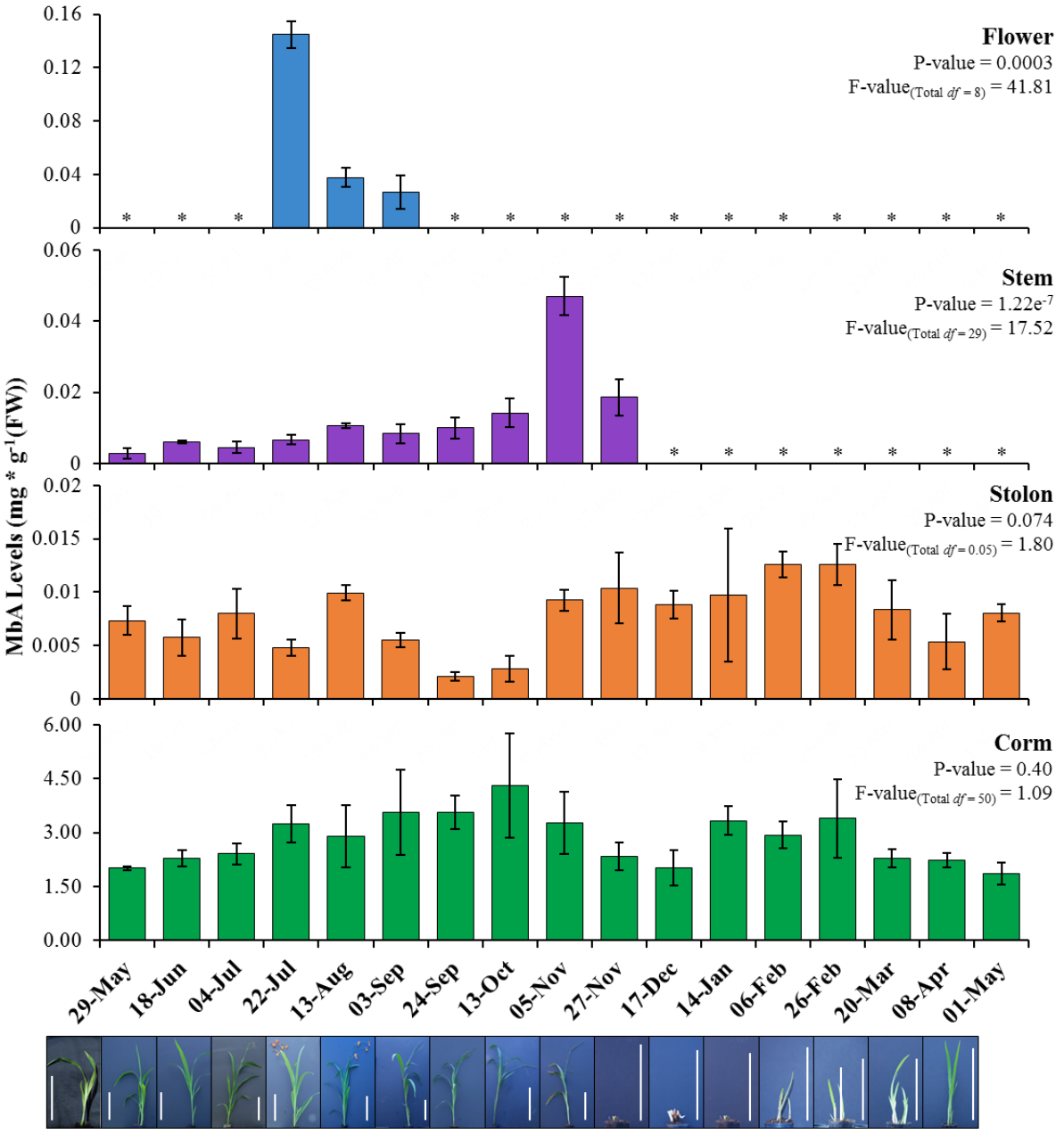


Figure 2.1: Temporal analysis of montbretin A accumulation within *C. x crocosmiiflora*. Pictures below the x-axis are of one of the biological replicates harvested at the corresponding time. White bar in each picture represents 30 cm. Larger versions of plant pictures can be seen in Supplemental Fig. S2.2. Results are shown as the average of three biological replicates. Error bars represent standard error. "*" denotes organ was not available for sampling at this time point. P-value and F-value were calculated using a single factor ANOVA in data analysis function of Excel. F-values reported were based on an $\alpha = 0.05$. "Total *df*" = total degrees of freedom.

2.3.2 MbA Accumulation Patterns in *Crocosmia x crocosmiiflora* Corms

Corms were shown to be the predominant site of MbA accumulation in *C. x crocosmiiflora*. However, it was not known whether MbA is uniformly distributed throughout the corm or if its accumulation is localized to specific regions. To investigate this, a triad of histological, MALDI, and MbA profiling analyses were performed on three segments originating from the same corm.

To better understand corm histology, above-ground plant material was removed and corms were extracted from soil (Fig. 2.2.a). Remaining leaves as well as roots and stolons were carefully removed from the extracted corms. These corms represented a below-ground part of the stem, as is apparent from spiral pattern of leaf scars (Fig. 2.2.b), specialized for storage of nutrient reserves, overwintering and vegetative propagation. The major tissue types observed after sectioning and staining included the epidermis, cortex, endodermis, and central vascular cylinder (Fig. 2.2.c-g). The outermost cell layer of the corms consisted of the epidermis (Fig. 2.2.e), which is covered by a thick cuticle. While cuticles are not composed of lignin or suberin, one of their major chemical component, cutin, shares certain chemical similarities with suberin (Kolattukudy, 1980; Ma *et al.*, 2004) and showed a positive reaction when stained with berberine-aniline blue. The tissue underneath the cuticle represents the epidermis and hypodermis (Fig. 2.2.e). In contrast to the very regular shape of cells that form the epidermis, hypodermis cells are more irregular shaped as is commonly observed in monocots (Evert, 2006). The following inner tissue layers of the corm show characteristic features of a differentiated cortex, endodermis, and central vascular cylinder. The endodermis was identified by the typical, lignified thickened cell walls (Fig. 2.2.f) separating the cortex from the central vascular cylinder. The central vascular cylinder contains numerous concentric amphivasal vascular bundles, each of which was composed of phloem surrounded by a ring of lignified xylem (Fig. 2.2.f-g) (Evert, 2006). When stained with Lugol's solution, the parenchyma cells within the cortex and the central vascular cylinder (CVC) appeared to store substantial amounts of starch (Fig. 2.2.h).

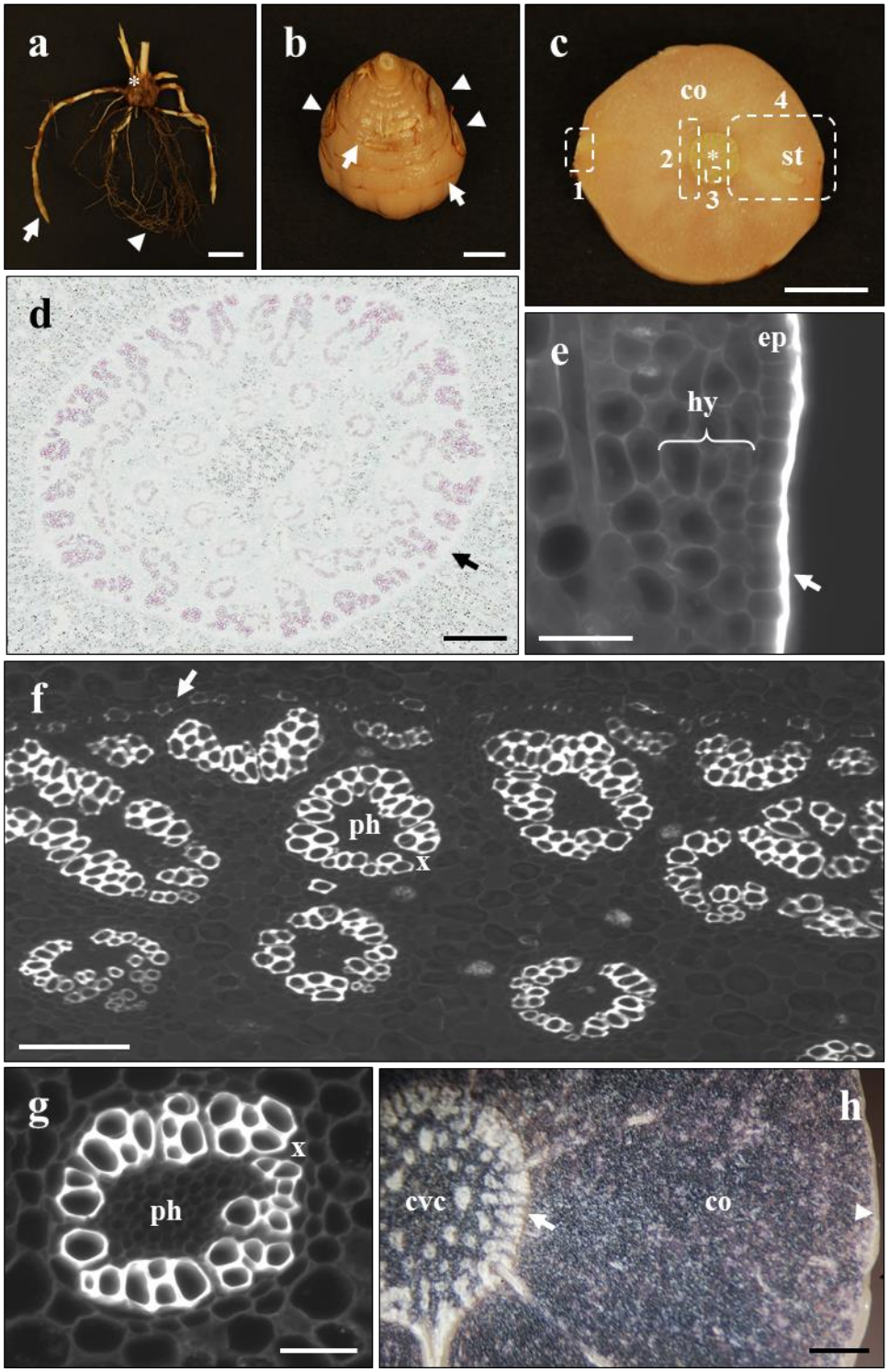


Figure 2.2: Anatomy of *C. x crocosmiiflora* corm. (a) Underground *C. x crocosmiiflora* organs including stolon (arrow), root (arrowhead), and corm (asterisk). Scale bar = 3 cm. (b) Individual corm stripped of stem, stolon, and root tissue. Leaf scars (arrows) and stolon outgrowth nodes (arrowhead) can be seen. Scale bar = 1 cm. (c) Unstained transverse section of corm. The areas encompassed within the white boxes correspond to areas visualized in this figure: box "1" corresponds to figure e, box "2" corresponds to figure f, box "3" corresponds

to figure g, and box "4" corresponds to figure h. Scale bar = 1 cm. (d) Phloroglucinol-HCl/toluidine blue stained transverse section of corm oriented at the central vascular cylinder. Lignified elements are stained pink and nucleic acids dark blue. The circular endoderm can be seen unstained (arrow). Scale bar = 500 µM. (e-g) Transverse sections of corm stained with berberine/aniline blue identify lignin and suberin. Samples are examined under ultraviolet light with epifluorescence optics. (e) Outermost layers of the cortex. Exterior to the compactly arranged hypodermis (hy) with little intercellular space and the single cell layer epidermis (ep), the cuticle is revealed by staining (arrow). Scale bar = 50 µM. (f) Innermost layers of the cortex. Staining reveals the presence of the single cell layer endodermis (arrow) with typical thickened cell wall through lignification. Within the central vascular cylinder, staining highlights horizontally oriented lignified xylem elements (x) in the axial concentric amphivasal vascular bundles where a ring of xylem surrounds the phloem (ph). Scale bar = 250 µM. (g) Close-up view of xylem (x) and phloem (ph) elements within a vascular cylinder. Scale-bar = 50 µM. (h) Transverse section of corm stained with Lugol solution. Dark staining indicates the presence of starch in the storage parenchyma cells of the cortex (co) and the central vascular cylinder (cvc). The epidermis (arrowhead) and endodermis (arrow) can be seen unstained. Scale bar = 2 mm.

MALDI imaging analysis showed MbA accumulation levels were lowest in the tissues of the CVC. Exterior the CVC, MbA accumulation showed a general trend of increasing towards the periphery of the corm section (Fig. 2.3).

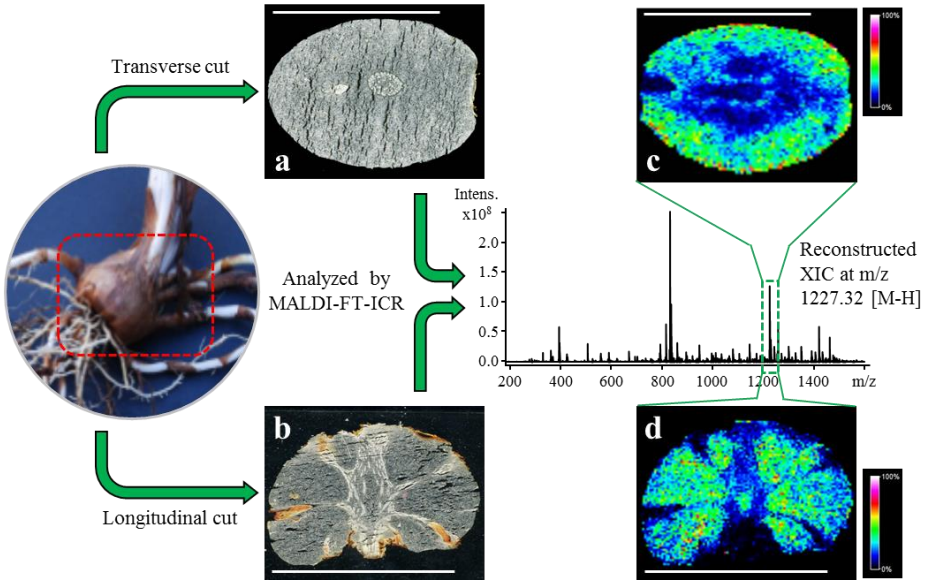


Figure 2.3: MALDI imaging of montbretin A accumulation patterns within corm segments. *C. x crocosmiiflora* corms were cut along the transverse and longitudinal planes into segments. Segments were analyzed by MALDI-FT-ICR. MALDI images were reconstructed from extracted ion chromatogram (XIC) at m/z 1227.32 corresponding to negative ionization of MbA. (a-b) optical images of transverse and longitudinal corm segments, respectively. Central vascular cylinder is indicated by the black arrow. (c-d) MALDI images of transverse and longitudinal corm segments respectively. White bars in pictures represent 3 cm.

These findings were supported by MbA profiling analysis of seven different sections of corm segments (Fig. 2.4; Table S2.3). In the CVC (section 1), MbA levels were significantly lower compared to the other six sections with an average of 0.30 mg/g FT. MbA levels of those segments outside the CVC (sections 2-7) had averages between 2.374 – 3.863 mg/g FT with the average level increasing towards the outermost section (Fig. 2.4).

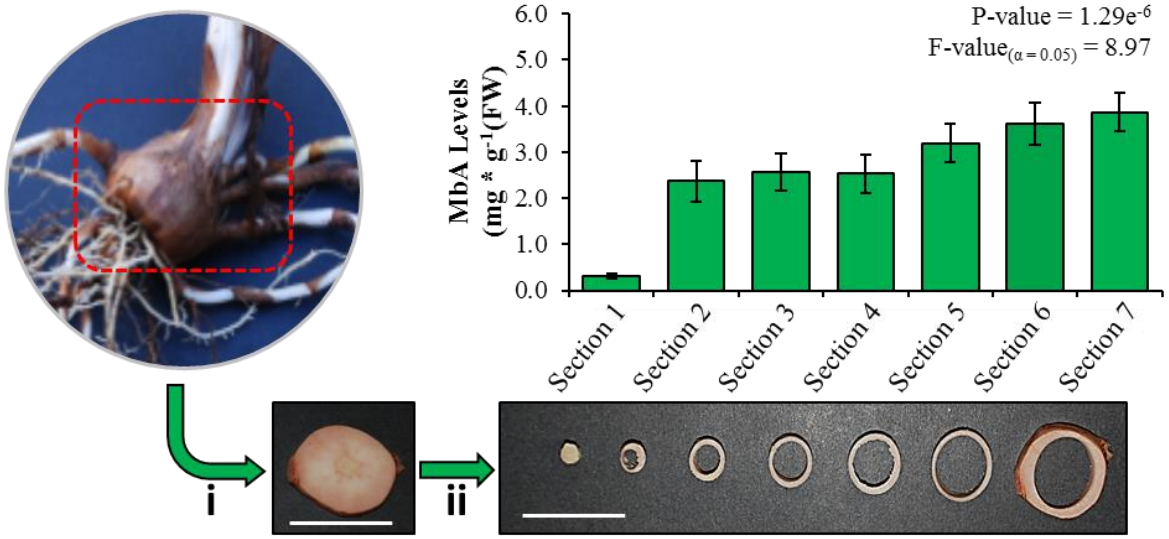


Figure 2.4: In-depth spatial analysis of montbretin A accumulation within corm segments. Corms were cut along the transverse plane into individual segments(i). Segments were then cut into seven sections with cork borers(ii). The picture just below x-axis of histogram show examples of sections used for analysis. White bars in pictures represent 3 cm. Results are shown as the average of eight biological replicates. Error bars represent standard error. P-value and F-value were calculated using a single factor ANOVA in data analysis function of Excel.

Collectively, these results suggest that MbA is tightly restricted to a specific region within the corm, but is primarily stored outside of the CVC with average levels highest towards the peripheral tissue.

2.3.3 Transcriptome Sequencing and Assembly

To generate a robust transcriptome resource for *C. x crocosmiiflora*, normalized cDNA libraries constructed from RNA isolated from multiple organs and biological replicates were

used to generate paired-end sequences with the Illumina HiSeq2000 platform. After quality check for and if needed, subsequent removal, of low quality sequences, low complexity reads, or contaminations, sequencing of 16 cDNA libraries resulted in 263,941,221 high-quality 150 bp paired-end reads (Table 2.1). Reads had their 15 bp adaptor sequence trimmed, and assembled into 470,118 contigs with an average length of 556 bp and an N50 of 655 bp. After further gap-filling, a unigene set was produced by extracting the coding sequences and removing redundant sequences as defined by greater than 98% sequence identity. This set contained 77,894 unigenes with an average length of 710 bp, an N50 of 891 bp, and a normally observed length distribution pattern (Table 2.1, Fig. 2.5).

Table 2.1: Sequencing and assembly results for *Crocosmia x crocosmiiflora* flower, leaf, stem, stolon, and corm organ.

Organ	# of Raw Reads	# of High-Quality Reads	Total Base-Pairs (High Quality)	Libraries from All Organs were Pooled and Redundancies Removed	# of Unigenes	Avg. Read Length	N50
Flower	53,613,896	51,285,139	15,385,554,300				
Leaf	83,751,469	49,701,603	14,910,561,000				
Stem	71,973,353	53,304,715	15,991,429,200		77,894	710 bp	891 bp
Stolon	51,588,144	50,343,863	15,103,171,200				
Corm	59,624,047	59,305,901	17,791,786,800				

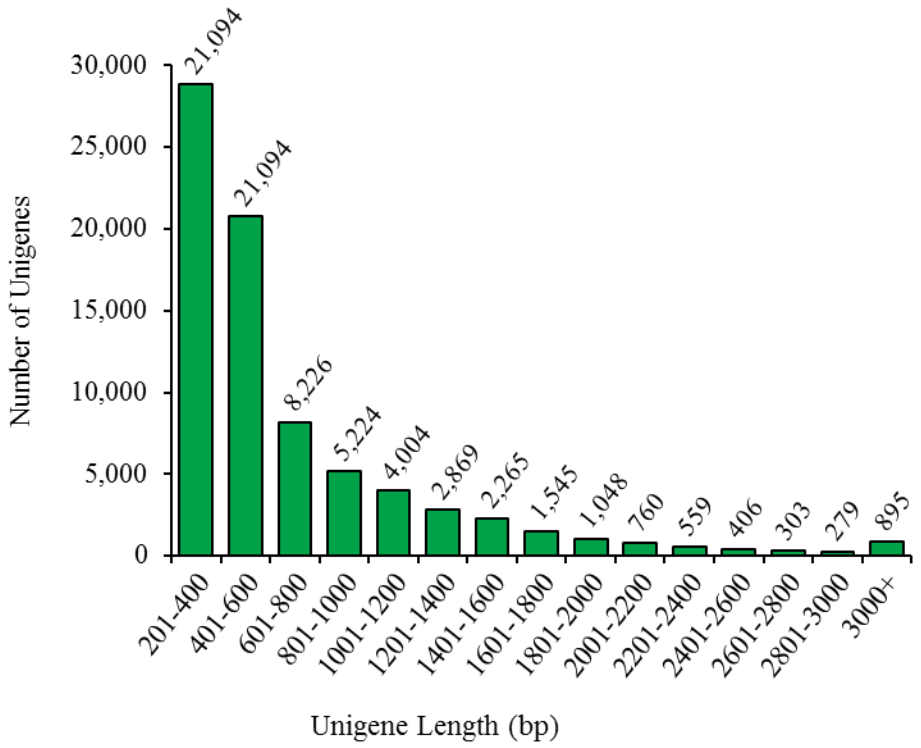


Figure 2.5: Length distribution of unigenes.

Two separate approaches were used to assess the completeness of the *C. x crocosmiiflora* unigene set: The Core Eukaryotic Gene Mapping Approach (CEGMA) (Parra *et al.*, 2007) and the Benchmarking Universal Single-Copy Orthologs (BUSCO) set (Simao *et al.*, 2015). Using a BLASTx search of the CEMGA dataset of 248 conserved core eukaryotic genes from *Arabidopsis thaliana*, representing an unbiased set of proteins conserved in eukaryotes, against the *C. x crocosmiiflora* unigene set identified 274 (59.8%) full-length orthologs, 167 (36.5%) partial-length orthologs and 17 (3.7%) orthologs not identified. Using a BLASTx search of the BUSCO dataset of 956 conserved genes from plant genomes, representing a set of near universally distributed single-copy genes, against the *C. x crocosmiiflora* unigene set identified 547 (57.2%) full-length orthologs, 397 (41.5%) partial-length orthologs and 12 (1.3%) orthologs not identified.

Collectively, these results suggest that the transcriptome resource developed has a typical average read length for a plant transcriptome and covers a good representation of both core eukaryotic genes and highly conserved plant single-copy ortholog genes.

2.3.4 Functional Annotation of Unigene Set

To enable the investigation into specialized metabolite biosynthesis in *C. x crocosmiiflora* and to contribute genomic information from an underrepresented phylogenetic clade to the public repository, *in silico* functional annotation on the unigene dataset was performed. BLASTx searches against public protein databases were used to annotate the unigene dataset. Searches against the NCBI NR database were successful for 61,094 (78.4%) unigenes with an e-value threshold of $1e^{-10}$. While this represented a significant portion of the unigene set, 39,279 hits had an e-value less than $1e^{-100}$ (Fig. 2.6). In-depth analysis of the BLAST search results showed that the top five hits for most unigenes were to sequences which only had a general characterization and provided little or no details towards specific functions.

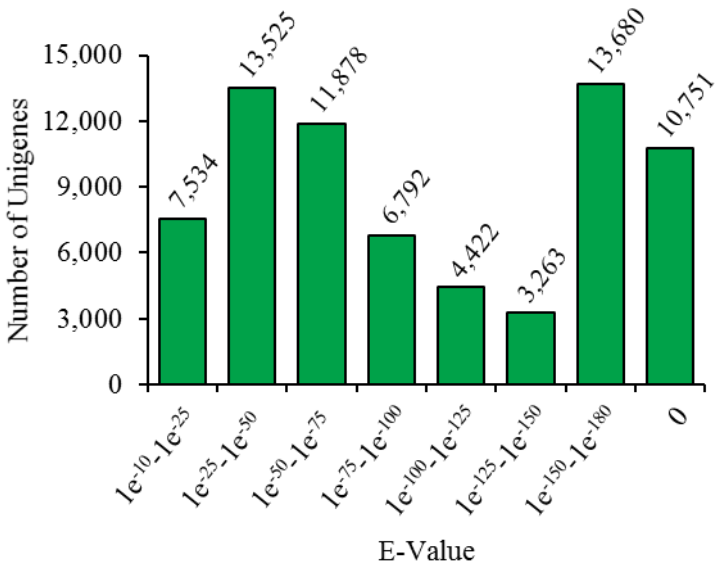


Figure 2.6: E-value distribution of unigenes.

Because annotation based on sequence homology to genes in the public database proved challenging, orthology-based annotation was used to gain insight into the molecular or biochemical functions contained within this unigene set. Using the Blast2Go software (https://www.blast2go.com/), unigenes were assigned GO terms (Conesa *et al.*, 2005). This resulted in 91,612 GO annotations assigned to 30,246 (38.8%) unigenes across three high-level categories: "biological process", "molecular function", and "cellular component" (Fig. 2.7). The cellular process category consisted of 31,687 GO annotations assigned to 9,798 (12.6%) unigenes. In this category, the most common annotations were "cell" (9,546 terms; 30.1%),

"cell part" (9,447 terms; 29.8%), and organelle (5,386 terms; 17.0%). The biological process category consisted of 39,404 GO terms assigned to 15,061 (19.3%) unigenes. In this category, the most common annotations were "metabolic process" (12,624 terms; 32.0%) and "cellular process" (11,294 terms; 28.7%). The molecular function category consisted of 22,777 GO terms assigned to 15,468 genes (19.9%). In this category, the most common annotations were "binding" (9,727 terms; 42.7%) and "catalytic activity" (9,646 terms; 42.46%).

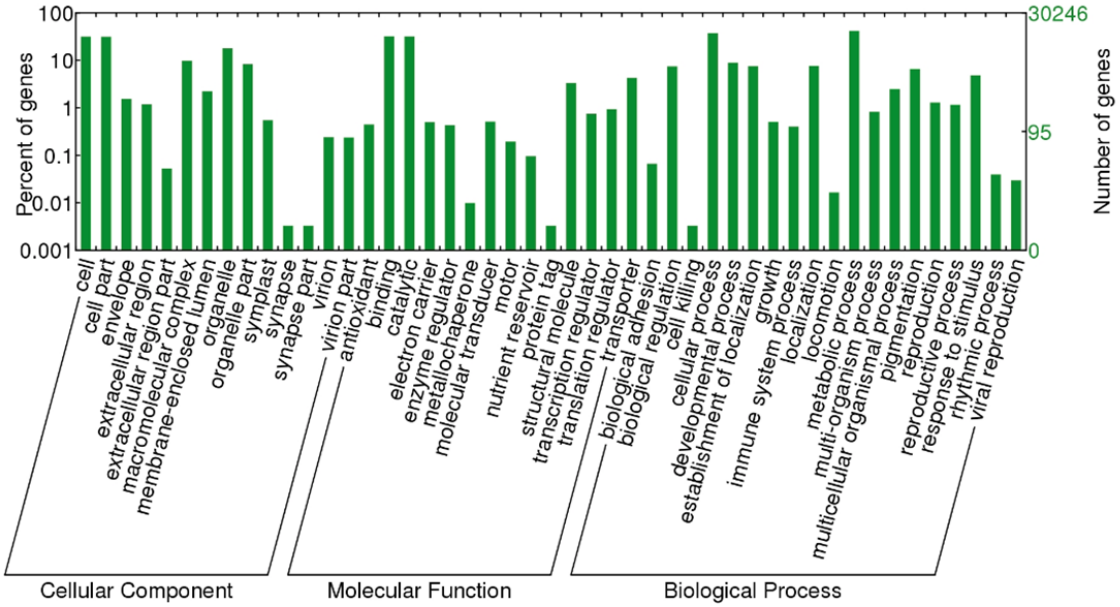


Figure 2.7: Gene Ontology classification of unigenes. Results are grouped into three high-level categories: cellular process, molecular function, and biological process.

Unigenes were further annotated based on EuKaryotic Orthologous Groups (KOG); a phylogenetic classification of the orthologous groups of proteins encoded in complete genomes of seven eukaryotes (Tatusov *et al.*, 2001). Overall, 19,366 unigenes (24.9%) were assigned 21,452 functional annotations classified over 25 KOG groups (Fig. 2.8). Among these 25 groups, most annotations were to "general function prediction only" (4,714; 22.0%), "signal transduction mechanisms" (2,197; 10.2%), "posttranslational modifications, protein turnover, chaperones" (1,887; 8.8%), "carbohydrate transport and metabolism" (1,291; 6.0%), and "transcription" (1,124; 5.2%).

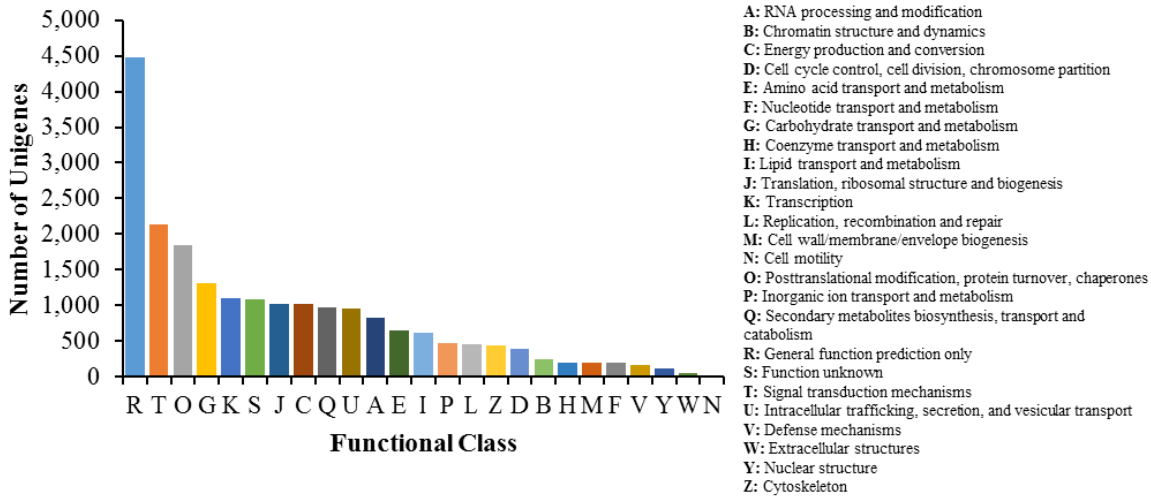


Figure 2.8: KOG functional classification of unigene dataset. All unigenes were aligned to KOGs database at NCBI to predict and categorize possible functions.

Interestingly, a large number of unigenes were assigned the "metabolic processes" GO term (Fig. 2.7) and the "secondary metabolite biosynthesis, transport, and catabolism" KOG classification (Q; Fig. 2.8). With the goal of developing this unigene set as a resource for the discovery of specialized metabolism biosynthesis genes, the canonical KEGG pathways were used to identify the actively expressed specialized metabolism pathways in the unigene set (Table 2.2). In total, 923 unigenes were found to be involved in the biosynthesis of various specialized metabolism pathways. Of the pathways, the cluster for "phenylpropanoid biosynthesis" [PATH: 00940] represented the largest group with 146 unigenes (15.8%), followed by "phenylalanine, tyrosine, and tryptophan biosynthesis" [PATH: 00400] with 88 unigenes (9.5%), and "flavonoid biosynthesis" [PATH: 00941] with 71 unigenes (7.7%).

Table 2.2: Summary of expressed unigenes related to specialized metabolites in *C. x crocosmiiflora* unigene set.

Path Number	Gene Pathway	Number of Genes
ec00073	Cutine, suberine and wax biosynthesis	20
ec00100	Steroid Biosynthesis	63
ec00130	Ubiquinone and other terpenoid-quinone biosynthesis	58
ec00140	Steroid Hormone Biosynthesis	31
ec00232	Caffeine metabolism	3
ec00400	Phenylalanine, tyrosine and tryptophan biosynthesis	88
ec00402	Benzoxazinoid biosynthesis	4
ec00403	Indole diterpene alkaloid biosynthesis	0
ec00760	Nicotinate and nicotinamide metabolism	23
ec00900	Terpenoid backbone biosynthesis	82
ec00901	Indole alkaloid biosynthesis	21
ec00902	Monoterpenoid biosynthesis	37
ec00904	Diterpenoid biosynthesis	25
ec00905	Brassinosteroid biosynthesis	0
ec00906	Carotenoid biosynthesis	44
ec00908	Zeatin Biosynthesis	32
ec00909	Sesquiterpenoid and triterpenoid biosynthesis	21
ec00940	Phenylpropanoid biosynthesis	146
ec00941	Flavonoid biosynthesis	71
ec00942	Anthocyanin biosynthesis	7
ec00943	Isoflavonoid biosynthesis	58
ec00944	Flavone and flavonol biosynthesis	2
ec00945	Stilbenoid, diarylheptanoid and gingerol biosynthesis	0
ec00950	Isoquinoline alkaloid biosynthesis	37
ec00960	Tropane, piperidine and pyridine alkaloid biosynthesis	33
ec00965	Betalain biosynthesis	0
ec00966	Glucosinolate biosynthesis	17
ec01058	Acridone alkaloid biosynthesis	0
Total		923

Collectively, these results suggest that *in silico* functional characterization of the *C. x crocosmiiflora* expressed unigene set is challenging due to the absence of well-characterized genes from closely related species within the public domain. However, annotations based on orthology provided insights into the diversity of functions contained within the unigene set and identified a substantial set of genes generally annotated with metabolism and specialized metabolism.

2.3.5 Identification of Montbretin A Biosynthetic Pathway Genes

Based on its chemical structure, the biosynthesis of MbA was proposed to occur in two parts, the EBP and LBP (Fig. 1.8). Responsible for producing the precursors for MbA, the EPB is proposed to be composed of genes from the phenylpropanoid, flavonoid, and nucleotide sugar pathways. The individual building blocks of MbA produced in the EBP are then used by enzymes of the LBP to assemble them into the complex MbA molecule. To establish a pool of candidate genes putatively involved in MbA biosynthesis, a homology-based approach was used to identify EBP genes and a correlative transcript-metabolite profiling approach was used to identify candidate LBP genes.

A set of customized reference sequence databases containing phenylpropanoid, flavonoid, and nucleotide sugar genes were employed in BLASTx searches against the unigene set to identify putative EBP genes. After confirmation of correct annotation by verifying with BLASTx searches against the NCBI NR database, 51, 19, and 33 genes were identified as putative members of the phenylpropanoid, flavonoid, and nucleotide sugar pathways, respectively (Fig. 2.9). Of these putative genes, 17, 11, and 25 genes respectively appear to be full-length. Focusing on genes useful for developing a sustainable production platform, 13, 7, and 15 putative genes of the respective pathways were amplified and cloned into a sequencing vector (Fig. 2.9).

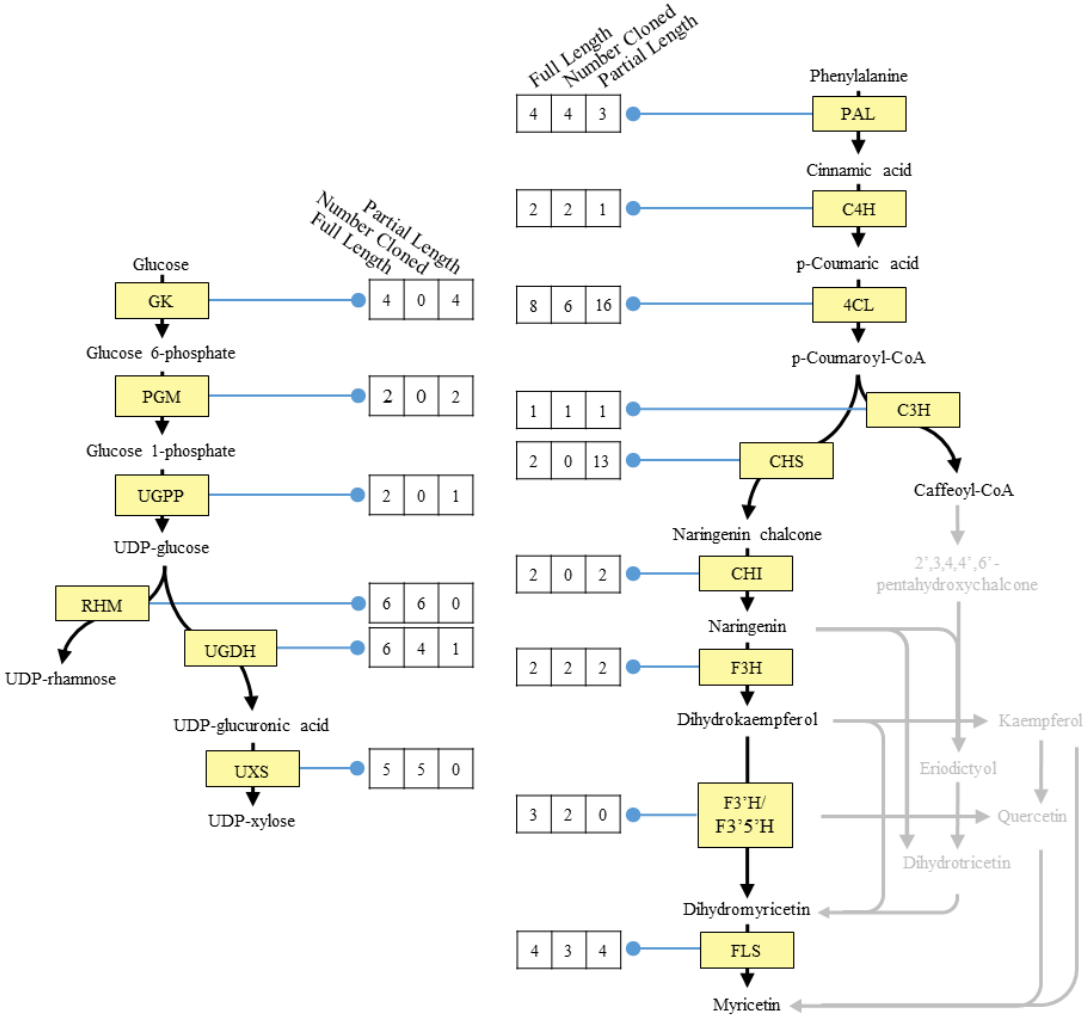


Figure 2.9: Identified and cloned putative genes likely involved in the montbretin A Early Biosynthetic Pathway

Based on the structure of MbA, the LPB likely contains activities of five UDP-glycosyltransferases (UGTs) which belong to glycosyltransferase family 1. To identify a candidate list of putative LPB UGTs, the program Haystack (http://haystack.mocklerlab.org/) (Mockler *et al.*, 2007) was used to identify unigenes whose expression pattern correlated with MbA accumulation. In this approach, MbA levels from each organ used for sequencing were determined (Table 2.3) and used as the input model to search against. As a result, 1,967 unigenes (2.5% of all unigenes) were identified to have expression patterns with high correlation to these MbA accumulation levels. Using the Plant Secondary Product Glycosylation (PSPG) motif, a conserved 44 amino acid motif found in most plant GT1 UGTs

(Vogt and Jones, 2000), a BLASTx search of the 1,967 unigenes identified 14 putative GT1 UGTs. Collectively, these results show that the proposed methods for identifying a pool of candidate genes of the MbA biosynthesis were successful.

Table 2.3: Montbretin A accumulation levels in *C. x crocosmiiflora* organs used for sequencing.

Sample	Total Average (mg/g FW)
flower - 1	0.051
flower - 2	0.053
flower - 3	0.047
leaves - 1	0.000
leaves - 2	0.000
leaves - 3	0.000
stem - 1	0.029
stem - 2	0.004
stem - 3	0.032
corm - 1	1.21
corm - 2	2.06
corm - 3	1.50
corm - 4	0.99
stolon - 1	0.012
stolon - 2	0.001
stolon - 3	0.001

2.4 DISCUSSION

After establishing a colony of *C. x crocosmiiflora*, I used histology, metabolite profiling, and transcriptome analyses to develop a set of resources to assist in identifying genes involved in the biosynthesis of MbA. Metabolite profiling showed that MbA is primarily accumulated within the non-vascular tissues of the corms. Analysis of these organs over a twelve-month period showed that MbA accumulation levels do not change significantly. RNA-seq and subsequent analysis produced a high-quality transcriptome. Employing these resources, a homology-based search approach identified multiple putative candidate genes for the EPB and an integrated analysis between MbA accumulation levels and gene expression identified putative candidate genes for the LBP.

Metabolite profiling showed that MbA is primarily found in the corm organ with minor levels found in the flower, stem, and stolon organs. While a single variable ANOVA analysis

showed profile levels over the 12 months were not statistically different, a large biological variation in MbA was observed. Plant secondary metabolite biosynthesis may be affected by multiple factors, such as plant interactions with pests, pathogens, and herbivores (Keeling and Bohlmann, 2006b; Miller *et al.*, 2005; Miller *et al.*, 2005; Zulak *et al.*, 2009; Zulak *et al.*, 2009), growth stages (Liu *et al.*, 1998), light (Binder *et al.*, 2009; Fischbach *et al.*, 1999; Fischbach *et al.*, 1999), temperature (Kovács *et al.*, 2010; Kovács *et al.*, 2010; Zhang *et al.*, 1997), nutrient availability (Bongue-Bartelsman and Phillips, 1995), salinity (Ali and Abbas, 2003), or water availability (Wang *et al.*, 2010a). Although the effects of environmental factors cannot be fully excluded, because plants were all subject to the same conditions, it is likely that MbA variation is due to genetic heterogeneity or age of corms. With the expansion of the originally collected material through propagation, future investigation into these unknown factors is a next step towards understanding MbA biosynthesis.

The histology of *C. x crocosmiiflora* corms identified them as modified stems containing traits of both monocotyledon stems, such as a cuticle, epidermis, and hyperdermis, as well as features reminiscent of monocotyledon roots, such as a vascular structure surrounded by an endodermis (Fig. 2.2.d-g) (Evert, 2006). Interestingly, berberine-aniline blue staining did not reveal the presence of suberized lamellae or a Casparian band typically indicative of an exodermis as normally seen in roots (Evert, 2006) or other monocotyledon corms (Cholewa and Griffith, 2004). The histology of *C. x crocosmiiflora* corms did not reveal any obvious specialized structures, such as idioblasts, which might be associated with biosynthesis or accumulation of MbA. Comparison of corm anatomy and MbA accumulation showed a strong correlation between starch storage patterns and MbA accumulation patterns. The mechanisms that lead to this pattern of accumulation may involve localized biosynthesis or transport, but are not yet known. The development of molecular probes specific for MbA biosynthesis will enable future studies on the pattern of localized MbA accumulation associated with starch storage. As the corm is the critical storage and overwintering organ from which *Crocosmia* plants regenerate and propagate during the growing season, protecting this organ is vital for the plant's asexual reproduction and propagation. Accumulation of MbA in the peripheral tissues of the corms may inhibit the digestion of starch contained in corm material ingested by herbivores, preventing or reducing nutrient acquisition. Herbivores may associate feeding on *Crocosmia* corms with absence of nutrient uptake and select against feeding. If this can be

proven, it would support a chemoecological role of MbA as a defense compound. Similar trends for defense metabolites have been observed such as toxic glycoalkaloids being primarily accumulated in the peripheral tissues of potato (*Solanum tuberosum*) tubers (Ha *et al.*, 2012) and hydroxynitrile glucosides, a storage form of cyanide, being predominantly stored in the outer cell layers of manioc (*Manihot esculenta*) tubers (Li *et al.*, 2013).

Transcriptome sequencing is often a method of choice at the foundation of gene discovery in non-model organisms (Strickler *et al.*, 2012). However, without reference genomic data, producing an accurate assembly able to provide insight into gene sequences and associated expression levels needed for downstream application of the transcriptomic data can be challenging (Feldmeyer *et al.*, 2011). A goal of this work was to produce a high-quality, comprehensive transcriptomic resource for *C. x crocosmiiflora*. Accordingly, ~5 Gbp of sequencing for each of the 16 *C. x crocosmiiflora* samples was performed. Through *de novo* assembly and subsequent refinement, a unigene set of 77,894 sequences was produced. When compared to a representative sampling of recent *de novo* transcriptomic studies which used similar assembly pipelines (Fu *et al.*, 2013; Lang *et al.*, 2015; Liu *et al.*, 2013; Mudalkar *et al.*, 2014; Wang *et al.*, 2010b; Wu *et al.*, 2014; Xu *et al.*, 2012; Zhang *et al.*, 2012; Zhou *et al.*, 2015), the *C. x crocosmiiflora* unigene set showed an average unigene length in the top quartile and a unigene set size that is higher than the number of actively expressed genes found in other species. Selective comparison against recent transcriptomic datasets of the closest relatives with data available, *Crocus* (Baba *et al.*, 2015; Jain *et al.*, 2016) and *Iris* (Ballerini *et al.*, 2013), show similarly large unigene set size and high average read length. While all three genera are members of the Iridaceae family, the *Crocus* and *Iris* genera are members of the Iridoideae subfamily while *Crocosmia* is of the Ixiodeae subfamily. Analysis of approximate genome size of genera of these subfamilies show members of Iridoideae contain genomes estiamated between approximately 10.66 – 70.81 Gbps while genera in Ixioideae contain genomes estiamted between appromately 1.08 – 5.28 Gbps (Goldblatt, 1971; Goldblatt *et al.*, 1984). While care should be taken in using a genome size to make predictions about estiamated transcriptome size, the estimated order of magnitude difference in genome size between the Iridoideae and Ixiodeae subfamilies suggest multiple factors could also be causing a higher than expected number of unigenes in the *C. x crocosmiiflora* unigene set. Total RNA for sequencing in this work was assembled from 16 biological samples which could have

negatively affected read assembly by increasing false positives and limiting ability to align longer reads (Chen *et al.*, 2010; Góngora-Castillo and Buell, 2013; Góngora-Castillo and Buell, 2013). Second, the high dynamic range of transcript expression levels could be problematic for comprehensive *de novo* sequencing and assembly by increasing false positives (Adamidi *et al.*, 2011). Third, depending on frequency, alternative splicing and fusion events can restrict the assembly of short sequences into longer ones, resulting in a single gene being assembled as multiple small fragments (Asmann *et al.*, 2011; Asmann *et al.*, 2011; Maher *et al.*, 2009). To assess completeness, a CEGMA analysis estimated that approximately 95% of the genes expressed in the *C. x crocosmiiflora* samples are present in the unigene set with approximately 60% contained as full-length sequences. This was further supported through identification of putative MbA EBP genes. Of these 103 genes, 53 (51.4%) were identified as full-length transcripts. Collectively, these results support this unigene set as an initial, high-quality representation of the genes expressed in *C. x crocosmiiflora*. As such, this data can be seen as a first draft of the *C. x crocosmiiflora* transcriptome and is the first genomic resource for the *Crocosmia* genus.

Without annotated reference sequences from closely related species, gaining functional information from even a high-quality transcriptome assembly remains challenging. To further develop the *Crocosmia* transcriptome into a resource for specialized metabolism gene discovery, *in silico* annotation was performed. While a BLASTx search of the unigene set against the NCBI NR database yielded hits for 61,094 (78.4%) of the unigenes, most comparable studies showed a lower percentage (30 – 60%) under identical parameters (Lang *et al.*, 2015; Lang *et al.*, 2015; Liu *et al.*, 2013; Liu *et al.*, 2013; Mudalkar *et al.*, 2014; Mudalkar *et al.*, 2014; Wang *et al.*, 2010b; Wang *et al.*, 2010b; Wu *et al.*, 2014; Wu *et al.*, 2014; Xu *et al.*, 2012; Xu *et al.*, 2012; Zhou *et al.*, 2015; Zhou *et al.*, 2015), specific functional annotations proved to be challenging. Of the successful BLAST search hits, 39,729 showed an e-value of less than $1e^{-100}$ and many of the top hits against unigenes, regardless of e-values, contained annotations that did not provide insight beyond general enzyme class. Difficulties with this homology-based annotation might be due to species-specific sequence divergence (Logacheva *et al.*, 2011; Strickler *et al.*, 2012; Strickler *et al.*, 2012). Additionally, the presence of false positive and short sequence transcripts could be contributing to this lack of effectivity (Grabherr *et al.*, 2011). Results from the orthology-based GO and KOG analyses

provided a representation of the gene functions contained within the transcriptome. However, they also highlighted the challenge of providing specific gene annotations for *C. x crocosmiiflora* based on in silico comparisons with public domain sequence resources. For example, compared to other transcriptomic studies the KOG annotation "general function prediction only" was 5 – 10% more common in the *C. x crocosmiiflora* transcriptome than any other (Fu *et al.*, 2013; Fu *et al.*, 2013; Lang *et al.*, 2015; Lang *et al.*, 2015; Liu *et al.*, 2013; Liu *et al.*, 2013; Mudalkar *et al.*, 2014; Mudalkar *et al.*, 2014; Wang *et al.*, 2010b; Wang *et al.*, 2010b; Wu *et al.*, 2014; Wu *et al.*, 2014; Xu *et al.*, 2012; Xu *et al.*, 2012). The public domain lacks annotated genomic data from species closely related to the *Crocosmia* genus and showcases the need and value of functionally characterizing *C. x crocosmiiflora* genes in future work.

The high number of unigene annotations involved in specialized metabolism as well as members of the KEGG specialized metabolism pathways suggests the *C. x crocosmiiflora* transcriptome is a suitable resource for the identification and subsequent characterization of genes involved in the biosynthesis of MbA. Interestingly, the "phenylpropanoid biosynthesis" and "flavonoid biosynthesis" pathways, two of the major ones involved in the EBP of MbA, were found to contain high numbers of unigenes in the *Crocosmia* transcriptome. Employing the sequence homology and integrated metabolite to transcriptome analysis approaches proved successful in identifying candidate genes of the MbA biosynthetic pathway. Through BLAST searches, 103 candidates were identified for the 16 genes involved in the MbA EBP. Similarly, the integration of MbA accumulation and unigene expression data resulted in the identification of 14 GT1 UGTs as candidate genes in the LBP. In order to identify these genes as members of the MbA biosynthetic process, additional functional characterization work must be performed.

2.5 CONCLUSION

In summary, this work established *C. x crocosmiiflora* resources for the elucidation of the MbA biosynthetic pathway. Temporal- and spatial-based metabolite profiling of MbA showed that MbA is primarily accumulated in corms at levels that do not change significantly throughout the year. MALDI and metabolite profiling identified that within corms, MbA primarily accumulates outside of the CVC. These findings provided first insights into

accumulation patterns and potential insights into biosynthesis in the plant. The accumulation in corms may also indicate a function of MbA as a defensive compound protecting the storage and overwintering organs of the perennial plant against herbivory. Transcriptomic sequencing and subsequent high-level gene annotation resulted in the first draft transcriptome assembly for *Crocosmia* containing 77,894 unigenes. *In silico* functional annotation of this transcriptome has proven to be difficult due to the lack of closely related, functionally characterized reference sequences. However, orthology-based annotation highlighted the diversity of functions contained within the transcriptome. Analysis of this transcriptome has led to the identification of putative genes for potentially every step of the proposed MbA biosynthetic pathway. In addition, the amplification of a variety of full-length putative EBP and LBP genes not only verified the quality of the transcriptome but also laid the foundation for future functional characterization of these genes and their encoded enzymes.

CHAPTER 3: FUNCTIONAL CHARACTERIZATION OF *CROCOSMIA x CROCOSMIIFLORA* NUCLEOTIDE SUGAR INTERCONVERSION ENZYMES INVOLVED IN MONTBRETIN A BIOSYNTHESIS

Nucleotide 5′-diphosphate sugars (NDP-sugars) are important metabolites used by glycosyltransferases in the biosynthesis of glycoconjugates and specialized metabolites. Of the NDP-sugars found in plants, the most common are the uridine 5′-diphosphate sugars (UDP-sugars). The diversity of UDP-sugars observed in plants is largely a result of genes of the 11 families of NDP-sugar interconversion enzymes (NSEs), which act on the few different sugar molecules produced by photosynthesis. Genes of two of these families, the UDP-xylose synthase (UXS) and UDP-rhamnose synthase (RHM) families, are thought to be involved in the biosynthetic pathway of montbretin A (MbA), a specialized metabolite found in *Crocosmia x crocosmiiflora*. In this chapter, I functionally characterized the *CxcUXS* and *CxcRHM* gene families identified in the *C. x crocosmiiflora* transcriptome. Within the *CxcUXS* family, four genes were functionally identified as UDP-xylose synthases and one as a putative UDP-4-keto pentose synthase. While the *in planta* role of a UDP-4-keto pentose synthase is unclear, site-directed mutagenesis showed a potential evolutionary path by which this function might have evolved from a UXS. Within the *CxcRHM* family, five genes were identified as UDP-rhamnose synthases and one as a 3,5-epimerase/4-keto-reductase. Kinetic and relative activity analyses of the different *Cxc*UXS and *Cxc*RHM members identified *Cxc*UXS4 and *Cxc*RHM1 as the most efficient enzymes of these two families. In addition to identifying specific NSE genes involved in the modular biosynthesis of MbA, these enzymes may enable the downstream characterization of UDP-glycosyltransferases (UGT) of the family 1 glycosyltransferases (GT1), involved in MbA biosynthesis through NSE-UGT coupled assays.

3.1 INTRODUCTION

Nucleoside 5′-diphosphate sugars (NDP-sugars) are essential to many biosynthetic systems. These activated sugars contain a high-energy bond between the sugar and NDP moieties, allowing them to serve as sugar donors in the biosynthesis of components of plant cell walls and a diverse array of specialized metabolite glycosides. To date, 30 NDP-sugars have been identified in plants with the majority produced from either UDP-α-D-glucose (UDP-

Glc) or GDP-α-D-mannose through a series of reduction, oxidation, decarboxylation, and/or epimerization reactions performed by NDP-sugar interconversion enzymes (NSEs) (Bar-Peled and O'Neill, 2011; Bowles *et al.*, 2005; Bowles *et al.*, 2006; Seifert, 2004). While the bulk of NDP-sugars in plants are incorporated into a variety of glycoconjugates such as glycolipids, glycoproteins, and polysaccharides, NDP-sugars are also commonly used to modify an array of specialized metabolites which may alter their stability, solubility, and bioactivity (Bowles *et al.*, 2005; Carpita and Gibeaut, 1993; Driouich *et al.*, 1993; Gibeaut and Carpita, 1994).

Among the suite of NDP-sugars found in plants are UDP-α-D-xylose (UDP-Xyl) and UDP-β-L-rhamnose (UDP-Rha). UDP-Xyl contributes, for example, to pectic polysaccharides in both primary and secondary cell walls through xylogalacturonan, N-linked oligosaccharides, and rhamnogalacturonan II, as well as to the glycosylation of specialized metabolites (Durand *et al.*, 2009; Jensen *et al.*, 2008; Northcote, 1963; Strasser *et al.*, 2008). UDP-Xyl is biosynthesized from UDP-glucuronic acid (UDP-GlcA) by UDP-xylose synthase (UXS) (EC # 1.1.1.305) (Fig. 3.1), which is active in the cytosol and Golgi apparatus (Harper and Bar-Peled, 2002; Pattathil *et al.*, 2005). UXS binds NAD^+, a cofactor, which facilitates the oxidation of UDP-GlcA to the first intermediate UDP-4-keto-glucuronic acid (UDP-4K-GlcA), which then undergoes decarboxylation to form the second intermediate UDP-4-keto-pentose (UDP-4KP) (Harper and Bar-Peled, 2002). Finally, using the bound NADH, UXS reduces UDP-4KP to form UDP-Xyl and NAD^+. To date, UXSs have only been isolated and functionally characterized in a small number of plant species (Guyett *et al.*, 2009; Harper and Bar-Peled, 2002; Pan *et al.*, 2010; Pattathil *et al.*, 2005; Yin and Kong, 2016; Zhang *et al.*, 2005). Within the UXS family, both transmembrane and soluble isozymes have been identified. Phylogenetic analysis has identified three distinct clades for UXSs where clade A and B possess type II membrane domains with their catalytic domain facing the membrane lumen, while clade C has been found to be a soluble enzyme in the cytosol (Harper and Bar-Peled, 2002; Pattathil *et al.*, 2005).

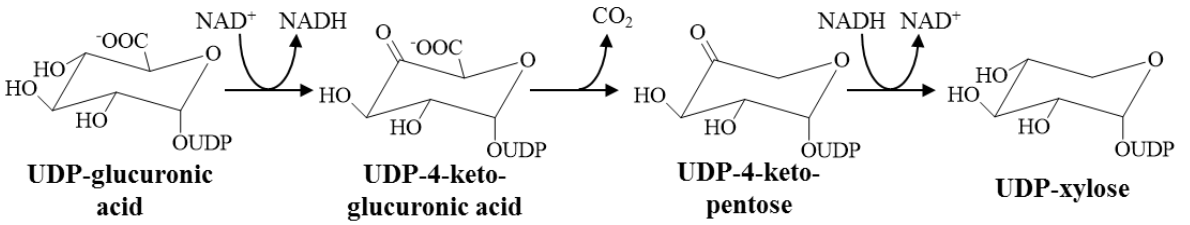


Figure 3.1: Biosynthesis of UDP-xylose from UDP-glucuronic acid. In the presence of NAD^+, UDP-GlcA is oxidized to UDP-4K-GlcA and NADH, followed by a decarboxylation to form UDP-4KP. After subsequent C-5 protonation, the still bound NADH is used to protonate the C-4 keto to form UDP-Xyl and regenerate NAD^+.

In plants, UDP-Rha serves as a building block of the pectic polysaccharides rhamnogalacturonan I and II, as well as in the formation of a variety of specialized metabolite glycosides (Ridley *et al.*, 2001). UDP-Rha is biosynthesized from UDP-Glc through a three-step reaction (Fig. 3.2) (Kamsteeg *et al.*, 1978). While bacteria employ three sequential enzymes in the biosynthesis of NDP-Rha, rfbB, rfbC and rfbD (Reeves *et al.*, 1996; Stevenson *et al.*, 1994), plants employ a single tri-functional enzyme, UDP-rhamnose synthase (RHM), which contains two sequential active sites (Oka *et al.*, 2007). Using NAD+ as a cofactor, RHM's N-terminal active site catalyzes the initial 4,6-dehydratase reaction (EC 4.2.1.46) of UDP-Glc to form the UDP-4-keto-6-deoxy-glucose (UDP-4K6DG) intermediate (Oka *et al.*, 2007; Watt *et al.*, 2004). After leaving the N-terminal active site, the UDP-4K6DG intermediate enters the C-terminal active site where it undergoes the 3,5-epimerase (EC 5.1.3.13) and 4-keto-reductase (EC 1.1.1.133) reactions in the presence of the NADPH cofactor to form UDP-Rha. Plants have also been found to contain an enzyme possessing both 3,5-epimerase and 4-keto-reductase activities, often referred to as the bifunctional UDP-4-keto-6-deoxy-glucose 3,5-epimerase/UDP-4-keto-rhamnose reductase (UER) or NRS/ER (Seifert, 2004; Watt *et al.*, 2004). To date, there are only a few reports of isolation and functional characterization of RHMs (Kim *et al.*, 2013; Martinez *et al.*, 2012; Oka *et al.*, 2007; Watt *et al.*, 2004; Yin *et al.*, 2016).

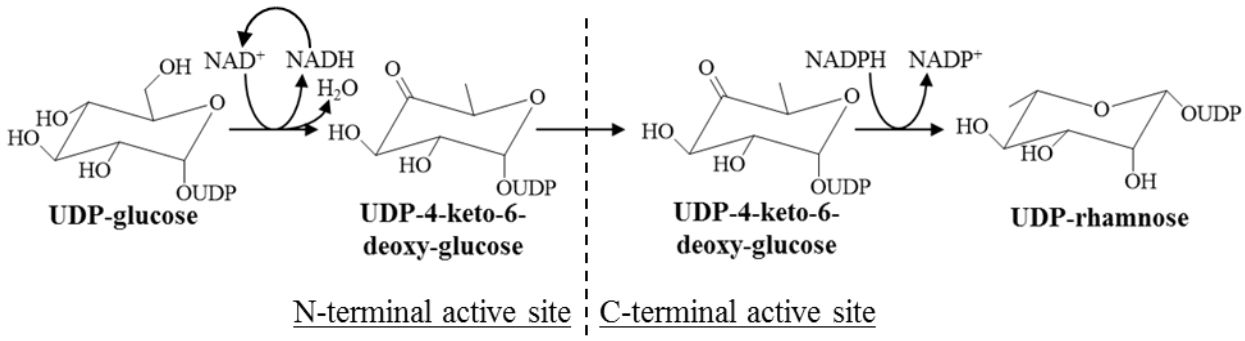


Figure 3.2: Biosynthesis of UDP-rhamnose from UDP-glucose. In the presence of NAD^+, UDP-Glc is oxidized to UDP-4K6DG. UDP-4K6DG then leaves the N-terminal active site and enters the C-terminal site. After subsequent isomerization of UDP-4K6DG, NADPH is used to reduce the C-4 keto to form UDP-Rha.

Montbretin A (MbA), a specialized metabolite with pharmacological benefits found in the non-model plant *C. x crocosmiiflora*, is a glycosylated flavonoid containing two glucose, two rhamnose, and a xylose moiety. These sugars play two important roles in the activity of MbA as an inhibitor of human pancreatic amylase (HPA): (i) they properly orient the inhibitory myricetin and caffeic acid moieties in the HPA active site and (ii) they help stabilize MbA in the active site through hydrogen bonding, increasing its inhibitory effect (Williams *et al.*, 2015).

An important consideration in the potential development of MbA as a type 2 diabetes therapeutic is the challenge for large-scale production. To this end, any *in vivo* or *in vitro* production system would require large amounts of UDP-Glc, UDP-Rha, and UDP-Xyl, for complete MbA biosynthesis. Due to difficulties with their chemical synthesis or enzymatic biosynthesis, nucleotide sugars are expensive reagents. In pursuing options for scalable *in vivo* biosynthesis, an economic approach would be to harness a host organism's endogenous pool of nucleotide sugars by modifying them with relevant NSEs as needed by metabolic engineering. Using this approach, several studies successfully produced glycosylated specialized metabolites by transforming exogenous NSE genes into microorganisms (Han *et al.*, 2014; Kim *et al.*, 2012). Regardless if MbA production will be achieved through increasing biosynthesis in *C. x crocosmiiflora* or a heterologous microbial system, understanding of the *C. x crocosmiiflora* NSE genes responsible for the biosynthesis of the required UDP-sugars of MbA biosynthesis will provide a valuable source of information and a potentially valuable set of tools.

To this end, this chapter expands on the functional annotation of the *C. x crocosmiiflora* transcriptome. Using genomic and biochemical approaches, I identified and functionally characterized the *CxcUXS* and *CxcRHM* gene families. These findings help lay a foundation for the comprehensive understanding of the MbA biosynthesis in *C. x crocosmiiflora* and may provide a resource to be used in synthetic biology and metabolic engineering endeavours requiring increased pools of UDP-Xyl or UDP-Rha.

3.2 EXPERIMENTAL

3.2.1 Subcloning of NSE cDNAs

Using cDNA previously generated from pooled *C. x crocosmiiflora* flower, stem, leaf, stolon, and corm organ RNA, as well as primers found in Table S2.1, cDNA corresponding to the five *C. x crocosmiiflora* UXS, five *C. x crocosmiiflora* RHM, and one UER identified in the transcriptome were amplified by PCR and cloned into pJet1.2 vectors (Fermentas, http://www.fermentas.com/). The *CxcUXS* and *CxcRHM* sequences that resulted from Sanger sequencing of these clones were used to predict transmembrane domains with the TMHMM (Krogh *et al.*, 2001; Sonnhammer *et al.*, 1998) and TMpred (Hofmann and Stoffel, 1993) programs. These sequences in the pJet1.2 vectors were then used as templates for sub-cloning the full-length and N-terminally truncated sequences into the pET28b(+) expression vector (EMD Chemicals, http://www.emdmillipore.com) in-frame with an N-terminal His_6-tag using primers shown in Table S3.1. Sequences and cDNA insertion orientation were verified by Sanger sequencing.

3.2.2 Alignments and Phylogenetic Analysis of NSE Sequences

For the purpose of phylogenetic analyses, amino acid sequence alignments were generated using ClustalW (Thompson *et al.*, 1994). Phylogenetic analysis was performed using a maximum likelihood algorithm in the MEGA 7.0 (http://www.megasoftware.net) (Kumar *et al.*, 2016) using uniform rate variation among sites, LG substitution model, BIONJ/NJ starting tree, and 1000 bootstrap repetitions. Amino acid sequence alignments were visualized using CLC Bio Main Workbench (https://www.qiagenbioinformatics.com/products/clc-main-workbench) while phylogenetic

trees weres visualized using the Interactive Tree of Life software (http://itol.embl.de/) (Letunic and Bork, 2011).

3.2.3 NSE Enzyme Assays

cDNAs in pET28b(+) expression vector were transformed into *E. coli* C43 (www.overexpress.com) containing the pRARE 2 plasmid isolated from Rosetta 2 cells (Novagen) to negate codon bias. Individual colonies were inoculated into 50 mL of Terrific Broth containing kanamycin (50 mg/L) and chloramphenicol (50 mg/L) and cultured at 37ºC and 180 rpm until an OD_{600} = ~0.8 was reached. Cultures were then cooled to 16ºC, induced by addition of isopropyl β-D-1-thiogalactopyranoside (final concentration 0.1 mM), and grown for 16 h at 180 rpm before harvesting. Recombinant protein was extracted and Ni^{2+} affinity purified as previously described (Roach *et al.*, 2014) with minor changes. Protein extracts were desalted on Sephadex PD minitrap G-25 columns (GE Healthcare, http://www.gehealthcare.com) pre-equilibrated with 50 mM sodium phosphate buffer. The buffer pH used for the initial characterization of enzyme activities was 7.5, and was modified for subsequent characterization and kinetic studies. Protein concentrations were determined using a bicinchoninic acid (BCA) protein quantification assay kit (Thermo Fisher, www.thermofisher.com) employing a standard curve and SDS-PAGE with measurement of protein band intensity performed with the program ImageJ (http://rsbweb.nih.gov/ij/).

After protein quantification, purified protein of putative *Cxc*UXS and *Cxc*RHM were assayed in triplicate to test activity towards UDP-GlcA and UDP-Glc, respectively. For initial *Cxc*UXS assays, 100 µL reactions containing 1 mM UDP-GlcA, 1 mM NAD^+, and 5 µg of purified protein, in a 50 mM sodium phosphate buffer (pH 7.5) was incubated overnight at 30ºC. For initial *Cxc*RHM assays, 100 µL reactions containing 1 mM UDP-Glc, 1 mM NAD^+, 1 mM NADPH, and 50 µg of purified protein, in a 50 mM phosphate buffer (pH 7.5) was incubated overnight at 30ºC. Reactions were terminated by treatment at 100ºC for 1 minute and the addition of 50 µL of chloroform to precipitate protein. Soluble fractions were separated by centrifugation at 1000 x *g* for 10 minutes, and then analyzed by liquid chromatography-mass spectrometry (LC-MS) (section 3.2.5).

To determine the optimal enzymatic conditions for *Cxc*UXS and *Cxc*RHM enzymes, enzyme assays were performed at a variety of different temperatures and pHs. For *Cxc*UXS

temperature optimization, assays were performed in 100 µL volume containing 1 mM UDP-GlcA, 1 mM NAD^+, 0.5 µg of purified protein, in 50 mM sodium phosphate buffer (pH 7.5) for 10 minutes at varying temperatures. For *Cxc*UXS pH optimization, assays were performed in 100 µL volume containing 1 mM UDP-GlcA, 1 mM NAD^+, 0.5 µg of purified protein, in 50 mM sodium phosphate buffer with varying pHs for 10 minutes at 10°C. For *Cxc*RHM temperature and pH optimization assays, the same conditions were employed except the 100 µL assay volume contained 1 mM UDP-Glc, 1 mM NAD^+, 1 mM NADPH, and 25 µg of purified protein and assays were performed for 1 hour. Reactions were terminated by treatment at 100°C for 1 minute and the addition of 50 µL of chloroform to precipitate protein. Soluble fractions were separated by centrifugation at 1000 x *g* for 10 minutes, and then analyzed by LC-MS (section 3.2.5).

To determine enzyme kinetic parameters of *Cxc*UXS2 – *Cxc*UXS5, assays were performed with nine different concentrations of UDP-GlcA or NAD+, both ranging from 0.01 mM to 2 mM. Enzyme concentrations in each assay were 184 – 286 pM for *Cxc*UXS2, 65.2 – 81.6 pM for *Cxc*UXS3, 88.2 – 67.3 pM for *Cxc*UXS4, and 33.0 – 94.5 pM for *Cxc*UXS5. All assays were incubated for 10 min at 30°C. Reactions were terminated by treatment at 100°C for 1 minute and the addition of 50 µL of chloroform to precipitate protein. Soluble fractions were separated by centrifugation at 1000 x *g* for 10 minutes, then analyzed by LC-MS (section 3.2.5). Kinetic analysis was performed by non-linear regression using the EXCEL template ANEMONA (Hernandez *et al.*, 1998).

3.2.4 NSE-UGT Coupled Enzyme Assays

NSE-UGT coupled assays were performed using *A. thaliana* UGT78D1, obtained from the JBEI collection, and *C. x crocosmiiflora* UGT7 (section 4.3.1). Using full-length cDNA for *A. thaliana* UGT78D1 (shipped in pDEST-GT vector) obtained from JBEI (Lao *et al.*, 2014) or the *C. x crocosmiiflora* draft transcriptome, cDNA for each gene were cloned into the pASK-IBA37plus expression vector (IBA Life Sciences, www.iba-lifesciences.com) in-frame with an N-terminal His_6-tag using primers shown in Table S3.1. Sequences and cDNA insertion orientation were verified by Sanger sequencing.

Plasmids were transformed into *E. coli* C43 (www.overexpress.com) containing the pRARE 2 plasmid isolated from Rosetta 2 cells (Novagen) to negate codon bias. Individual

colonies were inoculated into 50 mL of Terrific Broth containing ampicillin (100 mg/L) and chloramphenicol (50 mg/L) and cultured at 22°C and 200 rpm until an OD_{600} = ~0.5 was reached. Cultures were then cooled to 18°C, induced by addition of anhydrous tetracycline (final concentration 0.45 µM), and grown for 16 h at 180 rpm before harvesting. Protein was extracted from cell pellets by resuspending in 4 mL of ice-cold, 50 mM Tris-HCl extraction buffer (pH 7.5, 10% glycerol, 10 mM $MgCl_2$, 5 mM DTT, 0.2 mg/mL lysozyme, protease inhibitor, and 0.1 µL/mL benzonase). After five cycles of freeze-thawing in liquid nitrogen, lysed cells were clarified by centrifugation. Soluble protein was desalted on Sephadex PD minitrap G-25 columns (GE Healthcare, http://www.gehealthcare.com) pre-equilibrated with a 50 mM Tris-HCl buffer (pH 7.5, 10% glycerol, 1 mM DTT). The presence of the recombinant proteins of interest were confirmed by Western blots of SDS-PAGE gels. Western blots were performed using His-Tag Antibody HRP Conjugate Kit (EMD Millipore, www.emdmillipore.com) and visualized with Clarity Western ECL Blotting Substrate (Bio-rad, www.bio-rad.com).

Assays with protein extracts were performed in triplicate. For the *Cxc*UXS-*Cxc*UGT7 coupled reactions, 100 µL of unpurified *Cxc*UGT7 protein extraction was combined with 1 mM UDP-GlcA, 1 mM NAD^+, and 5 µg of purified *Cxc*UXS protein (as described in section 3.2.3) and incubated overnight at 30°C. For the *Cxc*RHM-*At*UGT78D1 coupled reactions, 100 µL of unpurified *At*UGT78D1 protein extraction was combined with 1 mM UDP-Glc, 1 mM NAD^+, 1 mM NADPH, and 50 µg of purified *Cxc*RHM protein (as described in section 3.2.3) and incubated overnight at 30°C. Reactions were terminated by treatment at 100°C for 1 minute and the addition of 50 µL of chloroform to precipitate protein. Soluble fractions were separated by centrifugation at 1000 g for 10 minutes at 4°C and analyzed by LC-MS (section 3.2.5).

3.2.5 Liquid Chromatography-Mass Spectrometry (LC-MS) Analysis

Reaction products from enzyme assays were analyzed by liquid chromatography (LC) (Agilent 1100 Series)/mass spectrometry detector (MSD) Trap (XCTplus) by comparison of retention times and mass spectra with authentic standards or enzymatic products of previously characterized enzymes.

For identification and functional characterization of NSE, enzyme products were analyzed on a Varian Inertsil ODS-3 column (4.6 mm internal diameter, 250 mm length, 5 μm pore size) with a temperature of 55°C and flow rate of 1.0 mL min^{-1}. The mobile phase used was a combination of two solvents: solvent A (H_20 + 2% formic acid) and solvent B (acetonitrile + 2% formic acid). The mobile phase run was 95% solvent A by 3.0 min, 5% solvent A by 10.0 min, 5% solvent A by 12.0 min, and 95% solvent A by 12.1 min, and held for 6.9 min, giving a total run time of 17 min. A diode array detector (DAD) was used to monitor wavelengths from 190 nm - 400 nm. The mass spectrometer mode was negative electrospray with nebulizer pressure 60 psi, dried gas rate 12 L min^{-1}, dry temp 350°C, and a m/z scanning range between 50 – 800.

For NSE-UGT coupled assays, enzyme products were analyzed on an Agilent ZORBAX SB-C18 column (4.6 mm internal diameter, 50 mm length, 1.8 μM pore size) with a temperature of 50°C and flow rate of 0.8 mL min^{-1}. The mobile phase used was a combination of two solvents: solvent A (H_20 + 0.2% formic acid) and solvent B (acetonitrile + 0.2% formic acid). The mobile phase run was 95% solvent A by 0.5 min, 80% solvent A by 5 min, 10% solvent A by 7 min, and 95% solvent A by 7.10 min, and held for 2.9 min, giving a total run time 10 min. A DAD was used to monitor wavelengths at 266 nm and 326 nm. The mass spectrometer mode was negative electrospray with nebulizer pressure 60 psi, dried gas rate 12 L min^{-1}, dry temp 350°C, and a m/z scanning range between 50 – 2000.

3.2.6 Differential Gene Expression Analysis

Quantification of transcript abundance was performed using Sailfish (Version 0.10) (Patro *et al.*, 2014). Subsequent statistical analysis was performed with limma (Ritchie *et al.*, 2015) and edgeR (Robinson *et al.*, 2010) in R. A gene was considered differentially expressed when the absolute log_2 fold change was equal to or greater than 1, and the adjusted p-value was equal to or less than 0.05.

3.3 RESULTS

3.3.1 Sequence Analysis of *C. x crocosmiiflora* UDP-Xylose Synthases

Analysis of the *C. x crocosmiiflora* draft transcriptome against the NCBI nr database revealed five sequences with high sequence similarity to previously characterized *A. thaliana*

UXSs. The five sequences were amplified by PCR from cDNA generated from the previously isolated *C. x crocosmiiflora* RNA (section 2.3.5) and the corresponding cDNAs were cloned into the pJET1.2 vector. The resulting clones were designated *CxcUXS1 – CxcUXS5*. In comparing the predicted amino acid sequences of these five putative UXSs to their closest related *A. thaliana* sequences, *Cxc*UXS1 and *Cxc*UXS2 showed 63.6% and 71.1% sequence identity, respectively, to *At*UXS2 while *Cxc*UXS3 – *Cxc*UXS5 showed 87.5%, 86.9%, and 85.3% identity, respectively, to *At*UXS3. These five sequences showed between 52.5% – 90.5% identity and 57.9% – 92.0% similarity at the nucleotide level, and between 55.1% – 90.7% identity and 61.8% – 92.2% similarity at the amino acid level (Table 3.1).

Table 3.1: Sequence pairwise comparisons of percent identity and similarity between *Cxc*UXS. Right-hand corner of matrix corresponds to nucleotide coding sequence similarity; left-hand corner corresponds to amino acid sequence similarity. Number outside and inside of parentheses corresponds to identity and similarity, respectively.

	***Cxc*UXS1**	***Cxc*UXS2**	***Cxc*UXS3**	***Cxc*UXS4**	***Cxc*UXS5**
***Cxc*UXS1**	–	67.1 (71.9)	54.9 (62.0)	54.5 (61.4)	54.5 (61.5)
***Cxc*UXS2**	68.1 (75.9)	–	52.8 (58.8)	52.7 (58.3)	52.5 (57.9)
***Cxc*UXS3**	56.9 (64.4)	55.1 (62.2)	–	90.5 (92.0)	80.2 (85.6)
***Cxc*UXS4**	56.7 (64.0)	55.3 (61.8)	90.7 (92.2)	–	81.2 (86.2)
***Cxc*UXS5**	55.8 (63.7)	55.5 (62.5)	85.5 (89.5)	88.2 (91.8)	–

An amino acid sequence alignment highlights conserved regions among the *Cxc*UXSs and two previously characterized UXSs from *A. thaliana* (Fig. 3.3). Conserved regions include the GXXGXXG motif, which is critical in binding the NAD^+ cofactor (Rossmann and Argos, 1978), and the YXXXK motif, which contains part of the Ser-Tyr-Lys catalytic triad critical in the oxidoreduction reaction (Duax *et al.*, 2000). Analysis of the five *Cxc*UXS sequences using the TMHMM (v2.0; (Krogh *et al.*, 2001)) and TMpred (Hofmann and Stoffel, 1993) programs predicted that *Cxc*UXS1 and *Cxc*UXS2 are likely type II membrane proteins (Fig. S3.1). The N-termini of these enzymes were predicted to contain a 30 and 39, respectively, amino acid long cytosolic domain, followed by a 22 amino acid long hydrophobic membrane-spanning domain. *Cxc*UXS3 – *Cxc*UXS5 were predicted to lack a predicted transmembrane domain, suggesting they are cytosolic proteins.

Figure 3.3: Amino acid sequence alignment of the *C. x crocosmiiflora* UXSs. The alignment includes protein sequences of the *Cxc*UXS family as well as *At*UXS2 and *At*UXS3 (Harper and Bar-Peled, 2002). Amino acids highlighted with blue background are those identified as different from the consensus. The conserved GXXGXXG motif, YXXXK motif, and catalytic serine are identified by the green, blue, and orange boxes, respectively.

Phylogenetic analysis of the *Cxc*UXS family with other characterized and putative plant UXSs showed the members of the *C. x crocosmiiflora* family fall into all three previously defined clades (Fig. 3.4) (Harper and Bar-Peled, 2002; Pattathil *et al.*, 2005). *Cxc*UXS1 and *Cxc*UXS2 were identified as members of the transmembrane clade A UXSs and clade B UXSs, respectively, while *Cxc*UXS3 – *Cxc*UXS5 were identified as members of the soluble clade C UXSs.

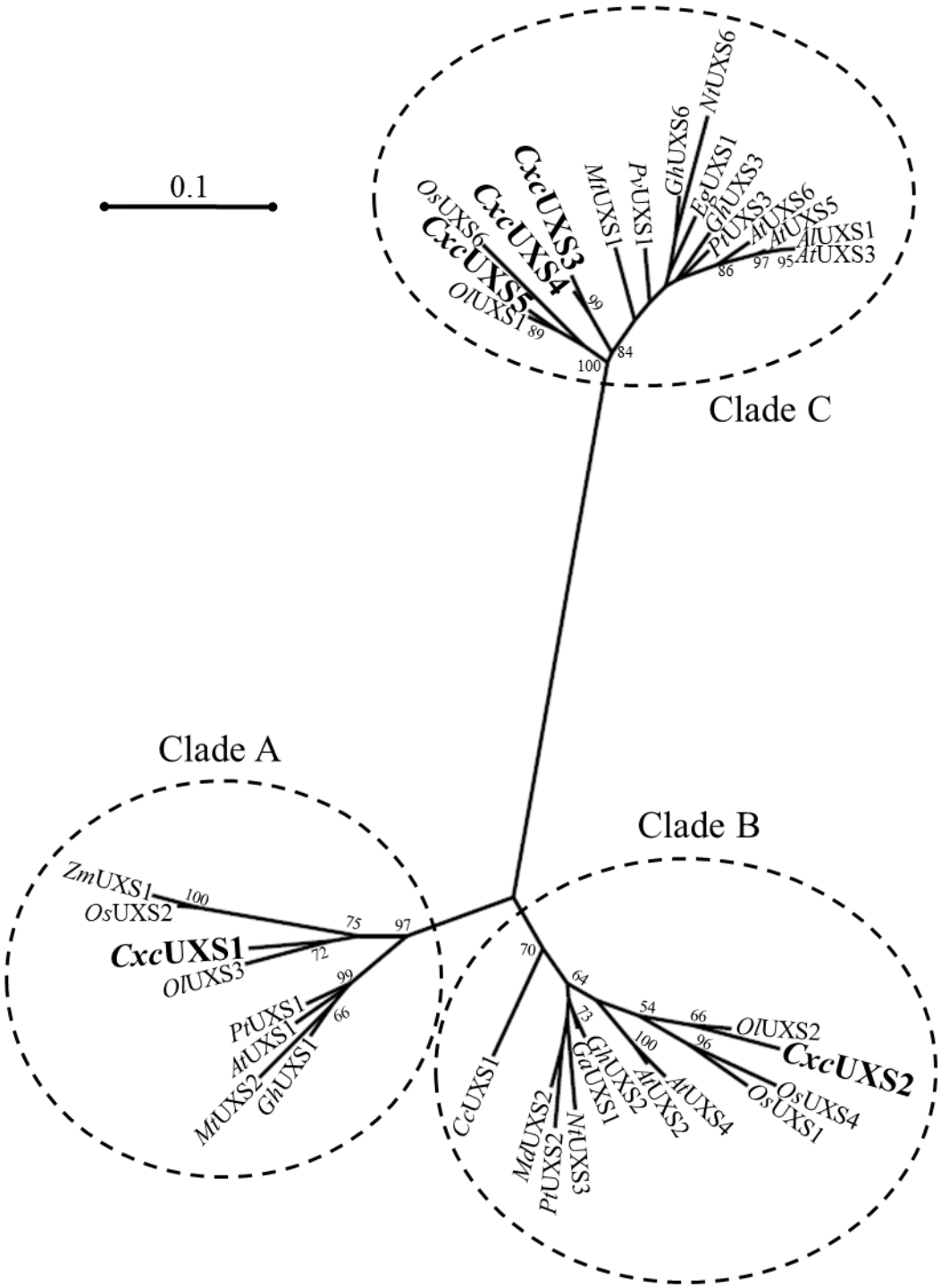


Figure 3.4: Phylogenetic analysis of *C. x crocosmiiflora* UDP-xylose synthases. The maximum-likelihood tree was produced with the *Cxc*UXS family as well as characterized and putative plant UXS obtained from the NCBI nr database using the MEGA 7.0 program (bootstrap value set at 1,000). Bootstrap values over 50% are indicated at the nodes. The black bar represents 0.1 amino acid substitutions per site. Sequences and alignment used in production of this tree can be found in Table S3.2 and Fig. S3.2, respectively.

3.3.2 Expression of Recombinant Proteins and Identification of *Cxc*UXSs as UDP-Xylose Synthases

For functional characterization of the putative *CxcUXS*, the corresponding cDNAs were cloned into the pET28b(+) vector for expression of the recombinant proteins with N-terminal His_6-tags. Recombinant proteins were expressed in *E. coli* and Ni^{2+} affinity purified. The resulting soluble proteins for *Cxc*UXS3 – *Cxc*UXS5 were detected by Western blotting with bands that matched the predicted molar mass of 38.4 – 39.7 kDa (Fig. S3.3a). Recombinant expression of *Cxc*UXS1 and *Cxc*UXS2 resulted in *E. coli* culture pellets that were approximately 30% dry cell weight compared to culture pellets of *E. coli* expressing *Cxc*UXS3 – *Cxc*UXS5 under identical culture conditions (Table S3.3). This observation may suggest that the putative N-terminal transmembrane domains of *Cxc*UXS1 and *Cxc*UXS2 had a negative effect on the host and protein expression in this system. To address this problem, the cDNAs of *Cxc*UXS1 and *Cxc*UXS2 were truncated, each at two different positions, to produce the recombinant N-terminal His_6-tag proteins without the putative N-terminal transmembrane domains: *Cxc*UXS1(Δ1–68), *Cxc*UXS1(Δ1–81), *Cxc*UXS2(Δ1–89), and *Cxc*UXS2(Δ1–98). Expression and purification of the truncated versions of *Cxc*UXS1 and *Cxc*UXS2 resulted in soluble proteins that were detected by Western blotting with bands that matched the predicted molar mass of 37.7 – 39.8 kDa (Fig. S3.3b–c). Based on BCA protein assay quantification and SDS-page purity analysis of protein purified from the 50 mL *E. coli* cultures, it was estimated that between approximately 15 and 40 mg of purified *Cxc*UXS protein could be isolated from a 1 L *E. coli* culture.

Enzyme assays (n = 3 replicates) with each of the purified *Cxc*UXS proteins were performed with UDP-GlcA as the substrate and NAD^+ as cofactor followed by LC-MS analysis of assay products. Assays were performed against a negative control of protein derived from *E. coli* expression an empty pET28b(+) vector, and authentic standards were used where available to identify reaction products. Products of enzyme assays with the truncated *Cxc*UXS2 and *Cxc*UXS3 – *Cxc*UXS5 proteins appeared in the LC-MS analysis as a single peak that was absent in the negative controls with a retention time and mass spectra that corresponded to the UDP-xylose standard (Fig. 3.5). *Cxc*UXS2 – *Cxc*UXS5 were thereby identified as UDP-Xyl synthases.

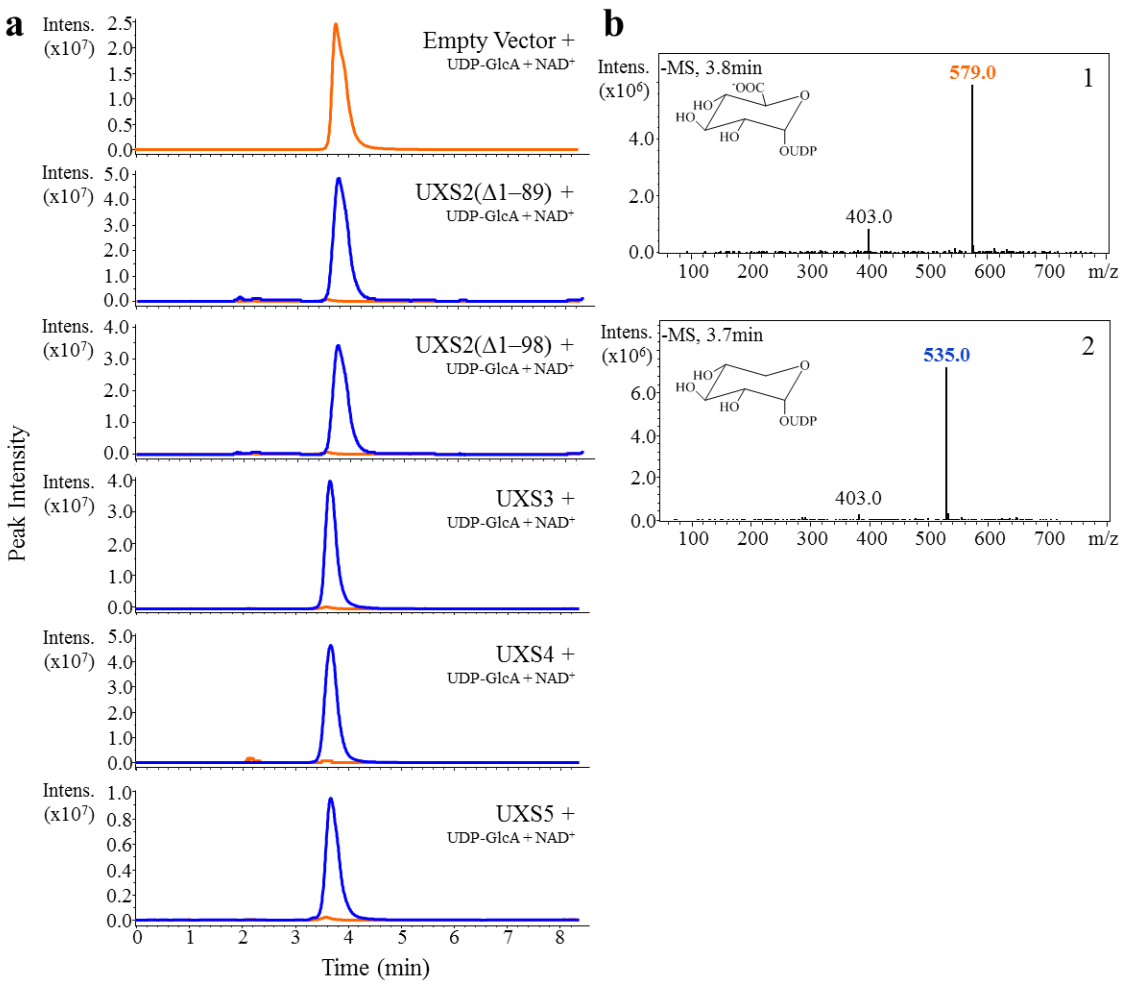


Figure 3.5: Representative regions of extracted ion LC-MS chromatograms and corresponding mass spectra of *Cxc*UXS2 – *Cxc*UXS5 enzyme assays. (a) Purified protein derived from *E. coli* expressing a control vector, *Cxc*UXS2(Δ1–89), *Cxc*UXS2(Δ1–98), *Cxc*UXS3, *Cxc*UXS4, and *Cxc*UXS5 were incubated overnight with 1 mM UDP-glucose and 1 mM NAD^+. Orange and blue traces represent extracted ion chromatograms for m/z of 579.0 [M-H](superscript -) and 535.0 [M-H](superscript -), respectively. Based on comparison of retention times and mass spectra against analytical standards, the peaks identified in orange and blue traces were confirmed to be UDP-glucuronic acid and UDP-xylose, respectively. (b) Mass spectra of enzyme assay products. The spectra presented in "1" and "2" are the background subtracted mass spectra for chromatographic peaks corresponding to UDP-glucuronic acid and UDP-xylose (theoretical molecular weight of each is 579.28 and 535.28), respectively. Within reach spectra, the ion with m/z of 403.0 [M-H](superscript -) corresponds to UDP.

Enzyme assays with both truncated *Cxc*UXS1 heterologous proteins showed one product peak not found in controls with a m/z of 551.0 [M-H](superscript -) (Fig. 3.6). This peak did not match the expected UDP-Xyl product of a UXS as had been previously seen. This product has not previously been reported for other clade A UXSs (Harper and Bar-Peled, 2002;

Pattathil *et al.*, 2005; Yin and Kong, 2016). Based on MS/MS analysis this product was tentatively identified as a UDP-gem-diol pentose (Fig. S3.4), however to my knowledge a UDP-sugar with a mass of 552 (551.0 when negatively ionized) has not been previously reported. This product may be explained by the observation that the expected reaction intermediate of UDP-Xyl biosynthesis, UDP-4KP, exists as both a gem-diol hydrate pentose and 4-keto pentose in an aqueous solution (Gu *et al.*, 2010). This observation is in line with the presence of both 533.0 and 551.0 ions [M-H](superscript -) in the MS/MS analysis, suggesting the presence of both gem-diol pentose and 4-keto pentose. It is therefore possible that the observed product is UDP-4KP. This suggests that the truncated versions of *Cxc*UXS1 do not perform the final enzymatic step reducing UDP-4KP into UDP-Xyl, but instead release the intermediate. Alternatively, *Cxc*UXS1 may be a *bona fide* UDP-4KP synthase with high similarity to clade A UXSs. To test if the observed product may be due to problems caused by the Ni^{2+} affinity purification or the presence of imidazole in the lysis buffer, as has been previously reported problems for other plant UXS enzyme activity (Yin and Kong, 2016), purified and unpurified enzyme assays were performed both with and without imidazole. Under these conditions, the product profile remained the same.

Previous work on the catalytic mechanism of human UXS showed the hydroxyl unit on the catalytic tyrosine of the YXXXK motif as critical in the reduction of the UDP-4KP intermediate (Eixelsberger *et al.*, 2012). This work showed that site-directed mutagenesis of this tyrosine to a phenylalanine eliminated the enzyme's reductase activity, resulting in a product profile exclusively producing UDP-4KP. Phylogenetic alignment of putative and characterized plant UXS shows *Cxc*UXS1 possess a unique YXXXK different from any publicly available plant UXS with a glutamic acid residue in the 252 positions instead of a glycine (Fig. S3.5). The potential effect of an amino acid with such different steric and charge characteristics compared to glycine, as well as its positioning in comparison to Tyr^{249}, suggests that Glu^{252} could affect *Cxc*UXS1's reductase activity through disruption of the catalytic Tyr^{249}. To test the effect of a sequence variation in the 252-position, the mutants *Cxc*UXS1(Δ1–68; E252G) and *Cxc*UXS1(Δ1–81; E252G) were produced and tested for activity and product outcome. The LC-MS product profile of *Cxc*UXS1(Δ1–68; E252G) and *Cxc*UXS1(Δ1–81; E252G) showed UDP-Xyl as their major product and the putative UDP-gem-diol pentose as its minor product (Fig. 3.6).

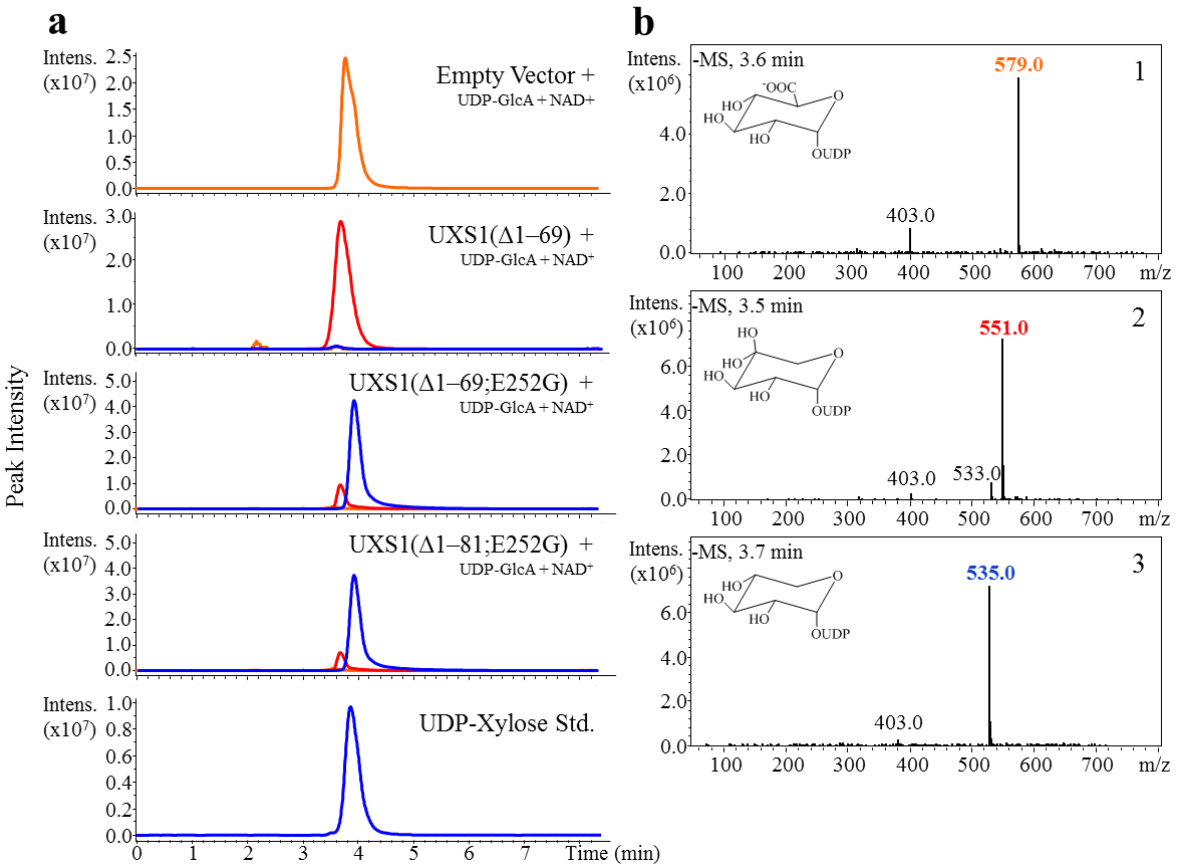


Figure 3.6: Representative regions of extracted ion LC-MS chromatograms and corresponding mass spectra of *Cxc*UXS1 enzyme assays. (a) Purified protein derived from *E. coli* expressing a control vector, *Cxc*UXS1(Δ1–69), and *Cxc*UXS1(Δ1–69; E252G) were incubated overnight with 1 mM UDP-glucuronic acid and 1 mM NAD^+. Orange, red, and blue traces represent extracted ion chromatograms for m/z of 579.0 [M-H](superscript -), 551.0 [M-H](superscript -), and 535.0 [M-H](superscript -), respectively. Based on comparison of retention time and mass spectra against analytical standards, the peaks identified in orange of blue traces were confirmed to be UDP-glucuronic acid and UDP-xylose, respectively. Based on MS/MS analysis (Fig. S3.3) and observations from Gu *et al.* (2010), the red peak likely correlates to UDP-4-keto-pentose. (b) Mass spectra of enzyme assay products. The spectra presented in "1", "2", and "3" are the background subtracted mass spectra for chromatographic peaks corresponding to UDP-glucuronic acid, a putative UDP-gem-diol pentose, and UDP-xylose (theoretical molecular weight of each is 579.28, 552.27, and 535.28), respectively. Within reach spectra, the ion with m/z of 403.0 [M-H](superscript -) corresponds to UDP.

While these findings suggest *Cxc*UXS1 is a UDP-4KP synthase with high similarity to plant UXS and present a potential mechanism in which such an enzyme could evolved from a UXS, additional work is needed to confirm if the *in planta* activity of *Cxc*UXS1 is a UDP-4KP synthase and that the observed results were not caused by partial enzyme inactivity due to *in vitro* effects.

3.3.3 Characterization of *Crocosmia x crocosmiiflora* UDP-Xylose Synthase Properties

To identify the most efficient *Cxc*UXS for potential application in metabolic engineering of MbA precursors, I determined the optimal temperature and pH conditions for each *Cxc*UXS and their basic enzyme kinetic parameters. Due to the same results obtained for both N-terminal truncated recombinant enzymes of *Cxc*UXS1 and *Cxc*UXS2, only *Cxc*UXS1(Δ1–69) and *Cxc*UXS2(Δ1–89) were used in these characterization analyses. Relative enzyme activity profiles showed a broad peak for the temperature dependent activities of all *Cxc*UXS with the highest levels of activity observed at 30ºC or 37ºC (Fig. 3.7; Table S3.4).

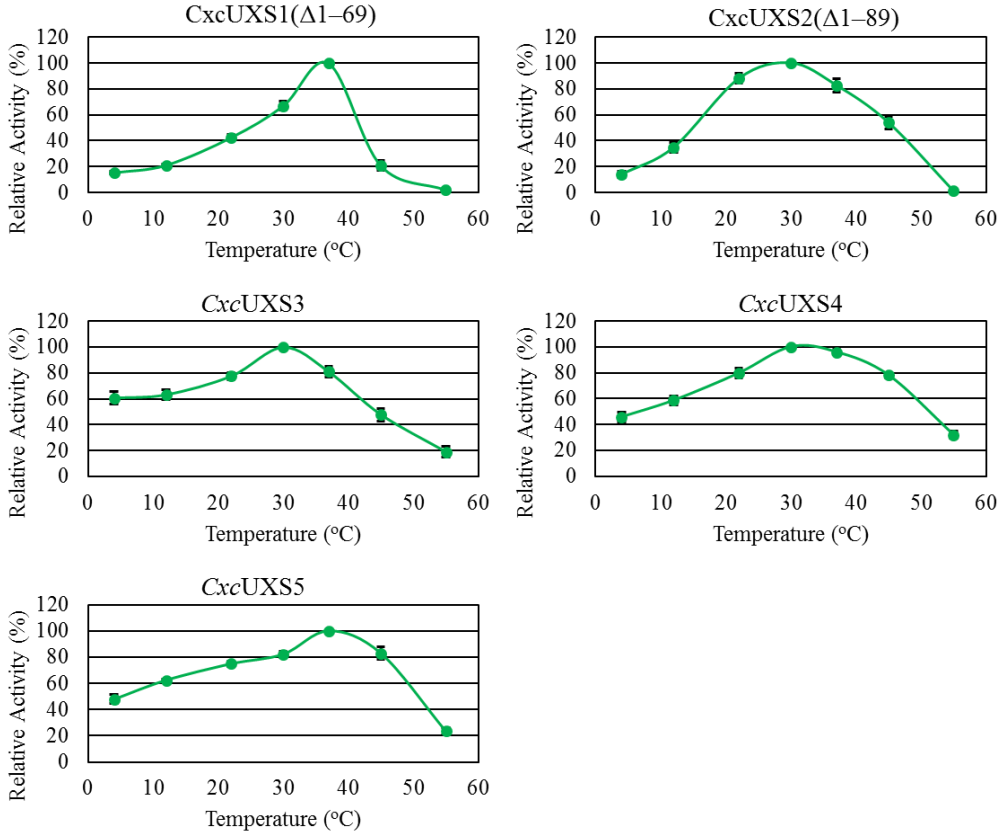


Figure 3.7: Temperature optimum of *Cxc*UXS activity *in vitro*. The activity of the recombinant *Cxc*UXSs was analyzed at different temperatures. Assays were performed with 5 replicates for each enzyme and each of the seven different temperatures. Results are shown as the calculated mean value with error bars representing standard error. 100% relative activity corresponds to the level of activity observed at the optimum tested temperature for a given enzyme.

In assessing optimal pH, relative activity assays identified a broad range of activities between pH 5.5 – 7.5 for all *Cxc*UXS with the highest level of activity observed at pH 7.0 or pH 7.5 (Fig. 3.8; Table S3.4).

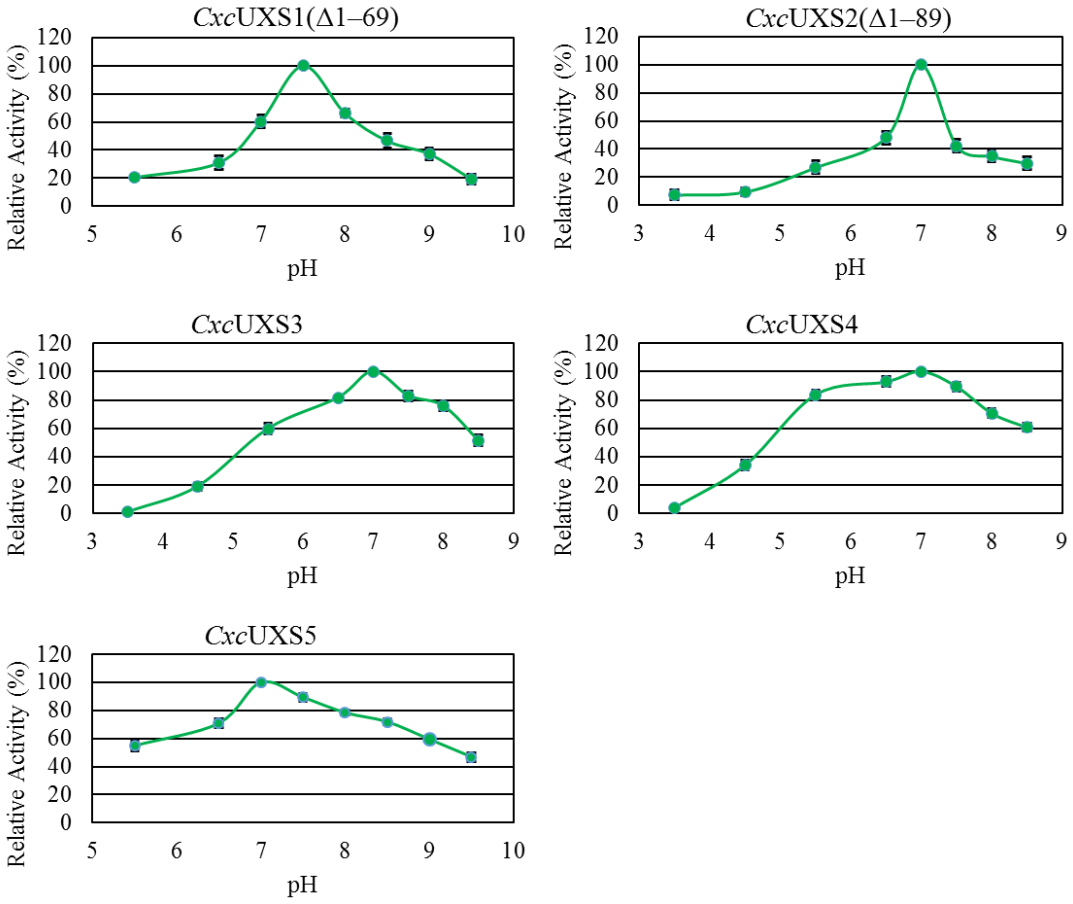


Figure 3.8: pH optimum of *Cxc*UXS activity *in vitro*. The activity of the recombinant *Cxc*UXSs was analyzed at different pHs. Assays were performed with 5 replicates for each enzyme and each of the eight different pH conditions. Results are shown as the calculated mean value with error bars representing standard error. 100% relative activity corresponds to the level of activity observed at the optimum tested pH for a given enzyme.

Kinetic characteristics of both the UDP-GlcA substrate and NAD^+ cofactor were determined for the four characterized UDP-xylose synthases under their optimal pH and temperature conditions (Table 3.2). This analysis showed similar apparent $K_{M,\ NAD+}$ values of 41.2 μM for *Cxc*UXS2(Δ1–89), 43.1 μM for *Cxc*UXS3, 40.6 μM for *Cxc*UXS4, and 52.0 μM for *Cxc*UXS5. The apparent K_m, UDP-GlcA values were 73.1 μM for *Cxc*UXS2(Δ1–89), 207 μM for *Cxc*UXS3, 147 μM for *Cxc*UXS4, and 126 μM for *Cxc*UXS5. Analysis of k_{cat} values shows statistically similar values for each enzyme whether determined through analysis of

NAD+ or UDP-GlcA, with the soluble enzymes apparently possessing higher kinetic efficiencies with $k_{cat,}$ UDP-GlcA values of 3.56 sec^{-1} for *Cxc*UXS3, 5.34 sec^{-1} for *Cxc*UXS4, and 0.8 sec^{-1} for *Cxc*UXS5 compared to 0.36 sec^{-1} for *Cxc*UXS2(Δ1–89).

Table 3.2: Enzymatic and kinetic properties of *Cxc*UXS1 – 5.

Property	*Cxc*UXS1 (Δ1–69)	*Cxc*UXS2 (Δ1–81)	*Cxc*UXS3	*Cxc*UXS4	*Cxc*UXS5
Optimal pH	7.5	7.0	7.0	7.0	7.0
Optimal Temperature (°C)	37.0	30.0	30.0	30.0	37.0
$V_{max;\ UDP\text{-}glcA}$ (pmoles sec^{-1} ug^{-1})		8.47 ± 0.22	89.8 ± 12.0	139 ± 27.1	19.3 ± 1.05
$V_{max;\ NAD+}$ (pmoles sec^{-1} ug^{-1})		8.24 ± 0.67	83.0 ± 11.4	118 ± 11.6	16.9 ± 1.88
$k_{cat;\ UDP\text{-}glcA}$ (sec^{-1})		0.36 ± 0.01	3.56 ± 0.48	5.34 ± 1.04	0.80 ± 0.04
$k_{cat;\ NAD+}$ (sec^{-1})		0.35 ± 0.03	3.30 ± 0.45	4.55 ± 0.44	0.70 ± 0.08
$K_{m;\ UDP\text{-}glcA}$ (μM)		73.1 ± 0.03	207 ± 0.16	147 ± 0.22	125 ± 0.07
$K_{m;\ NAD+}$ (μM)		41.2 ± 0.09	43.1 ± 0.15	40.6 ± 0.10	52.0 ± 0.13
$k_{cat}/K_{m\ ;\ UDP\text{-}glcA}$ (sec^{-1} μm^{-1})		0.005	0.017	0.036	0.006
$k_{cat}/K_{m\ ;\ NAD+}$ (sec^{-1} μm^{-1})		0.008	0.077	0.112	0.013

These results identify overall similar kinetic properties for *Cxc*UXS3 – *Cxc*UXS5 and confirm two general observations with plant UXS: cytosolic UXS appear to have higher kinetic efficiencies *in vitro* compared to transmembrane UXS, and a higher affinity for the NAD^+ cofactor compared to UDP-GlcA.

3.3.4 Sequence Analysis of *Crocosmia x crocosmiiflora* UDP-Rhamnose Synthases

Analysis of the *C. x crocosmiiflora* draft transcriptome against the NCBI nr database revealed five sequences with high sequence similarity to previously characterized *A. thaliana* RHM and one sequence with high sequence similarity to a previously characterized *A. thaliana* UER. The six sequences were amplified by PCR from cDNA generated from the previously isolated *C. x crocosmiiflora* RNA (section 2.3.5) and the corresponding cDNAs were cloned into the pJET1.2 vector. The resulting clones were designated *CxcRHM1 – CxcRHM5* and *CxcUER1*. Comparison of the predicted amino acid sequences with their closest related sequence from *A. thaliana* showed *Cxc*RHM1 – *Cxc*RHM5 to be 83.0%, 82.7%, 82.3%, 81.7%, and 81.7%, respectively, identical to *At*RHM2, while *Cxc*UER1 showed 82.0% identity to *At*UER. The five *Cxc*RHM sequences showed between 78.0% – 94.7% identity and 83.0%

– 96.8% similarity at the nucleotide level, and between 85.0% – 97.5% identity and 89.9% – 98.2% similarity at the amino acid level (Table 3.3).

Table 3.3: Sequence pairwise comparisons of percent identity and similarity between *Cxc*RHM. Right-hand corner of matrix corresponds to nucleotide coding sequence similarity; left-hand corner corresponds to amino acid sequence similarity. Number outside and inside of parentheses corresponds to identity and similarity, respectively.

	***Cxc*RHM1**	***Cxc*RHM2**	***Cxc*RHM3**	***Cxc*RHM4**	***Cxc*RHM5**
***Cxc*RHM1**	–	94.0 (95.3)	78.9 (83.8)	78.0 (83.0)	79.5 (84.1)
***Cxc*RHM2**	95.5 (97.3)	–	79.0 (83.3)	78.5 (83.4)	79.6 (83.9)
***Cxc*RHM3**	85.9 (90.7)	86.1 (90.8)	–	95.2 (96.8)	94.7 (96.2)
***Cxc*RHM4**	85.0 (89.9)	85.0 (89.9)	97.5 (98.2)	–	90.2 (93.3)
***Cxc*RHM5**	87.0 (91.3)	86.5 (91.0)	94.9 (97.0)	92.7 (95.4)	–

An amino acid sequence alignment highlights conserved regions among the *Cxc*RHMs, *Cxc*UER1, and the previously characterized RHM2 from *A. thaliana* in both the dehydratase and epimerase/reductase domains. The N-termini of the *Cxc*RHMs contains the dehydratase domain and three important motifs: the GXXGXXA motif, which is critical for binding the NAD^+ cofactor (Rossmann and Argos, 1978), as well as the YXXXK motif (Duax *et al.*, 2000) and TDE motif (Duax *et al.*, 2000; Watt *et al.*, 2004), which contain the highly conserved amino acids responsible for the dehydratase reaction. The C-termini of the *Cxc*RHMs and *Cxc*UER1 contain the epimerase/reductase domain and two important motifs: the GXXGXXG motif, which is critical in binding the NAD^+ cofactor (Rossmann and Argos, 1978), as well as the YXXXK motif, which makes up part of the Ser-Tyr-Lys catalytic triad critical in the oxidoreduction reaction (Duax *et al.*, 2000). Analysis of the *Cxc*RHM and the *Cxc*UER1 sequences using the TMHMM (v2.0; (Krogh *et al.*, 2001)) and TMpred (Hofmann and Stoffel, 1993) programs predicted that all proteins are soluble as opposed to membrane bound.

Figure 3.9: Amino acid sequence alignment of the *C. x crocosmiiflora* RHM and UER. The alignment includes protein sequences of the *Cxc*RHM and *Cxc*UER1, as well as the *At*RHM2 (Oka *et al.*, 2007). (a) Conserved sequences of the N-terminal dehydratase domain. (b) Conserved sequences of the C-terminal epimerase/reductase domain. Amino acids highlighted with blue background colour are those different from the consensus. The green boxes in each domain identify the conserved GXXGXX(G/A), which are involved in binding the NAD^+ cofactor. The blue and orange box in each domain identify residues critical for enzymatic reactions; blue boxes correspond to the YXXXK motif while the orange box identifies the TDE motif.

Phylogenetic analysis of the five *Cxc*RHM with other characterized and putative plant RHM showed no obvious major clusters (Fig. 3.10). Across the different species, RHM enzymes were more likely to cluster with enzymes from the same plant species.

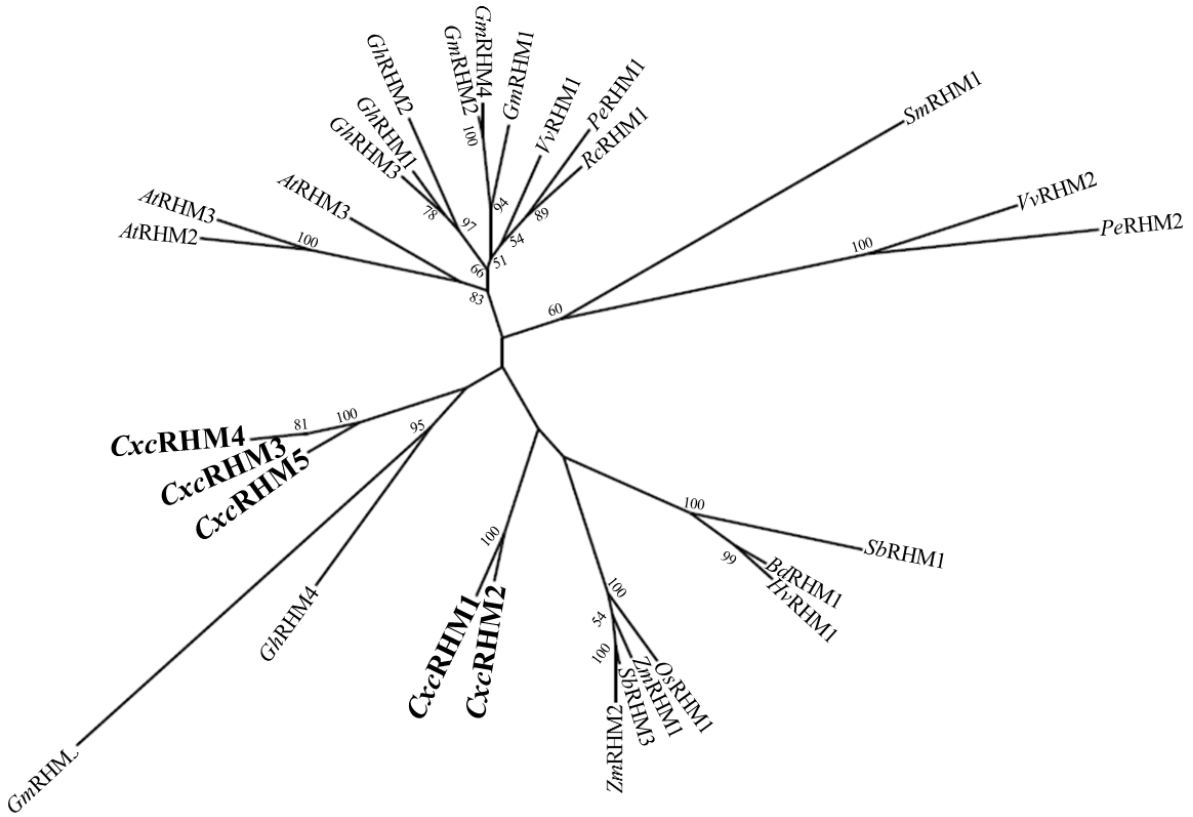


Figure 3.10: Phylogenetic analyses of *C. x crocosmiiflora* RHM. The maximum-likelihood tree was produced with five *Cxc*RHM with other characterized and putative plant RHM obtained from the NCBI nr database using the MEGA 7.0 program (bootstrap value set at 1,000). Bootstrap values over 50% are indicated at the nodes. The black bar represents 0.06 amino acid substitutions per site. Protein alignment and sequences are given in Fig. S3.6 and Table S3.4.

3.3.5 Identification of *Cxc*RHM as UDP-Rhamnose Synthases and *Cxc*UER1 as UDP-4-Keto-6-Deoxy-Glucose 3,5-Epimerase/UDP-4-Keto-Rhamnose Reductase

For functional characterization, the *CxcRHM* and *CxcUER1* cDNAs were cloned (section 2.3.5) into the pET28b(+) vector for expression of the corresponding proteins with an N-terminal His_6-tag. Recombinant proteins were expressed in *E. coli* and Ni^{2+} affinity purified. The resulting soluble proteins for *Cxc*RHM1 – *Cxc*RHM5 were detected by Western blotting with bands that matched the predicted molar mass of 75.7 – 76.1 kDa, and *Cxc*UER1 appeared with a band that agreed with the expected molar mass of 35.6 kDa (Fig. S3.7). Based on BCA protein assay quantification and SDS-page purity analysis of protein purified from the 50 mL *E. coli* cultures, it was estimated that between approximately 1.5 and 10 mg of purified *Cxc*RHM protein could be isolated from a 1 L *E. coli* culture.

Enzyme assays of each recombinant enzyme with either UDP-Glc or UDP-4K6DG as a substrate, as well as NAD^+ and NADPH as cofactors, followed by LC-MS analysis of reaction products were performed against relevant controls to identify the functions of these enzymes. Because analytical standards for UDP-4K6DG and UDP-Rha were not available, the products of the previously characterized *At*RHM2 dehydratase domain (*At*RHM2-N) and *At*RHM2 epimerase/reductase domain (*At*RHM2-C) were used to produce UDP-4K6DG (observed m/z of 547.0 [M-H](superscript -)) and UDP-Rha (observed m/z of 549.0 [M-H](superscript -)) respectively (Fig. S3.8). LC-MS analysis of the products of enzyme assay with *Cxc*RHM1 – *Cxc*RHM5 revealed two peaks that were not present in the negative ion controls with a m/z and retention time corresponding to the *At*RHM2-produced UDP-Rha and UDP-4K6DG (Fig. 3.11). *Cxc*RHM1 – *Cxc*RHM5 were thereby identified as UDP-Rha synthases.

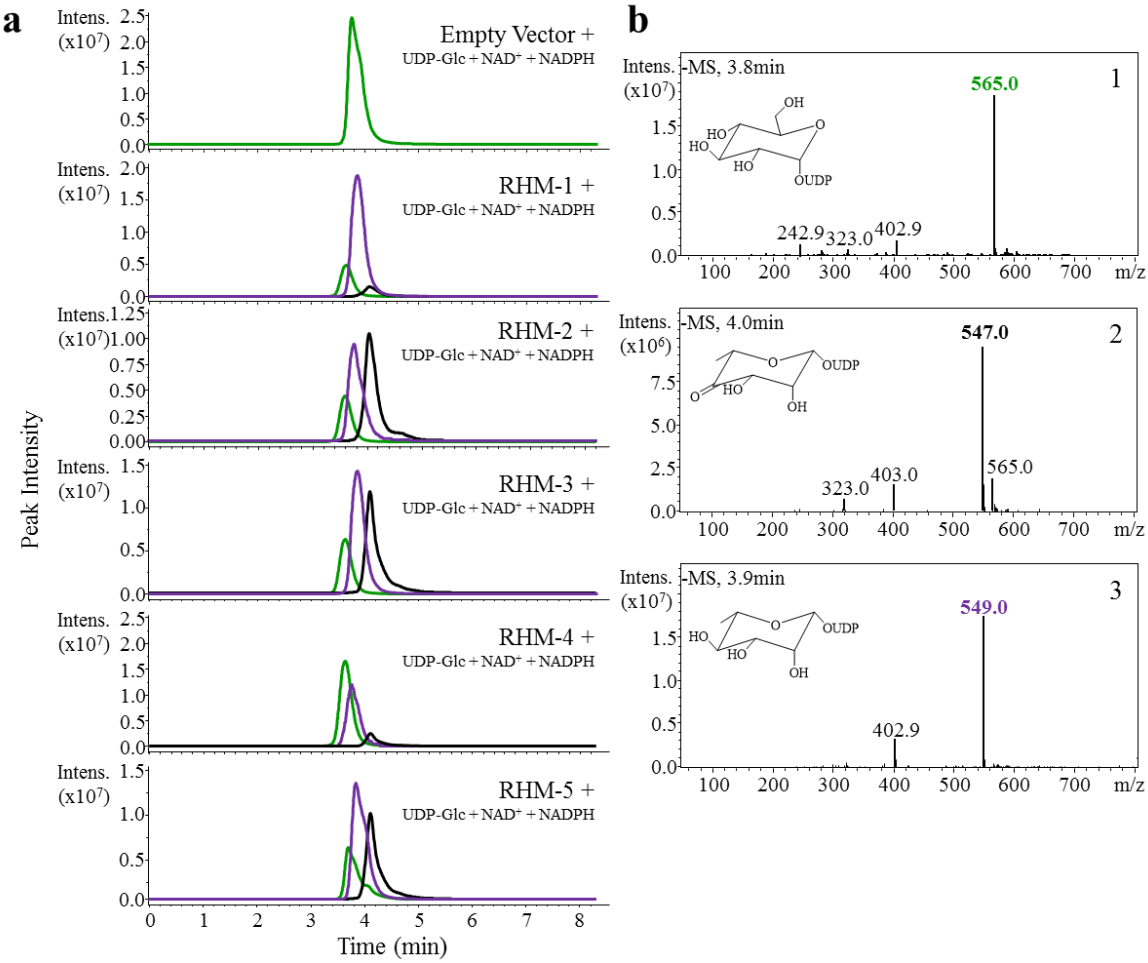


Figure 3.11: Representative regions of extracted ion LC-MS chromatograms and corresponding mass spectra of *Cxc*RHM1 – *Cxc*RHM5 enzyme assays. (a) Purified protein derived from *E. coli* expressing a control vector, *Cxc*RHM1, *Cxc*RHM2, *Cxc*RHM3, *Cxc*RHM4, and *Cxc*RHM5 were incubated overnight with 1 mM UDP-glucose, 1 mM NAD^+, and 1 mM NADPH. Green, black, and purple traces represent extracted ion chromatograms for m/z of 565.0 [M-H](superscript -), 547.0 [M-H](superscript -), and 549.0 [M-H](superscript -), respectively. Based on comparison of retention times and mass spectra against analytical standards, the peaks identified in green traces were confirmed to be UDP-glucose. Based on previously reported NMR analysis, peaks in the black and purple traces correspond to UDP-4-keto-6-deoxy-glucose and UDP-rhamnose, respectively (Oka *et al.*, 2007). (b) Mass spectra of enzyme assay products. The spectra presented in "1", "2", and "3" are the background subtracted mass spectra for chromatographic peaks corresponding to UDP-glucose, UDP-4-keto-6-deoxy-glucose, and UDP-rhamnose (the theoretical molecular weight of each is 565.30, 547.29, and 549.30), respectively. UDP-4-keto-6-deoxy-glucose is predicted to exist as both a keto and gem-diol pentose in aqueous solution, as is suggested by the presence of an ion with m/z of 565.0 [M-H](superscript -). Within reach spectra, the ion with m/z of 403.0 [M-H](superscript -) corresponds to UDP.

LC-MS analysis of assay product showed that *Cxc*UER1 converted the UDP-4K6DG produced by *At*RHM2-N into a product with a retention time and mass spectrum that matched those of UDP-Rha produced by coupled *At*RHM2-N and *At*RHM2-C (Fig. 3.12). *Cxc*UER1 was thereby identified as a UDP-4-keto-6-deoxy-glucose 3,5-epimerase/UDP-4-keto-rhamnose 4-keto-reductase.

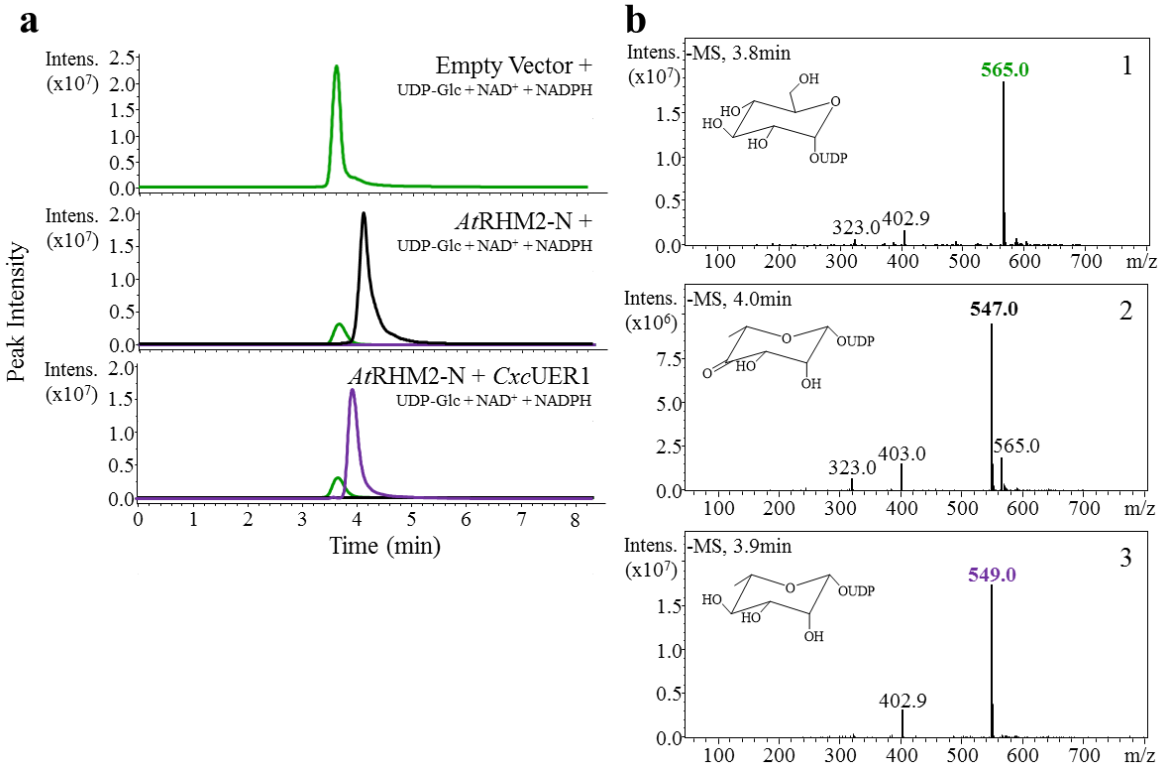


Figure 3.12: Representative regions of extracted ion LC-MS chromatograms and corresponding mass spectra of *Cxc*UER enzyme assays. (a) Purified protein derived from *E. coli* expressing a control vector, *At*RHM2-N alone (Oka and Jigami, 2007), or *At*RHM2-N combined with *Cxc*UER1 were incubated overnight with 1 mM UDP-glucose, 1 mM NAD^+, and 1 mM NADPH. Green, black, and purple traces represent extracted ion chromatograms for m/z of 565.0 [M-H](superscript -), 547.0 [M-H](superscript -), and 549.0 [M-H](superscript -), respectively. Based on comparison of retention times and mass spectra against analytical standards, the peaks identified in green traces were confirmed to be UDP-glucose. Based on previously reported NMR analysis, peaks in the black and purple traces correspond to UDP-4-keto-6-deoxy-glucose and UDP-rhamnose, respectively (Oka *et al.*, 2007). (b) Mass spectra of enzyme assay products. The spectra presented in "1", "2", and "3" are the background subtracted mass spectra for chromatographic peaks corresponding to UDP-glucose, UDP-4-keto-6-deoxy-glucose, and UDP-rhamnose (the theoretical molecular weight of each is 565.30, 547.29, and 549.30), respectively. UDP-4-keto-6-deoxy-glucose is predicted to exist as both a keto and gem-diol pentose in aqueous solution, as is suggested by the presence of an ion with m/z of 565.0 [M-H](superscript -). Within reach spectra, the ion with m/z of 403.0 [M-H](superscript -) corresponds to UDP.

3.3.6 Characterization of *Crocosmia x crocosmiiflora* RHM and UER Enzyme Properties

While comparison of kinetic parameters would be the ideal method for comparing enzyme efficiency, identifying these for the *Cxc*RHM and *Cxc*UER enzymes would be difficult as each member of the *Cxc*RHM family contains two active sites, which act sequentially in the biosynthesis of UDP-Rha. The UDP-4K6DG intermediate that is used as the substrate for the *Cxc*RHM C-terminal and *Cxc*UER active sites is not commercially available and can only be obtained through enzymatic reaction and purification. Given these limitations, I focused on comparing relative activities of the different *Cxc*RHMs, temperature and pH optima.

Relative enzyme activity profiles showed a broad peak for the temperature-dependent activities for all *Cxc*RHM with the highest levels of activity observed at 30°C or 37°C (Fig. 3.12; Table S3.6).

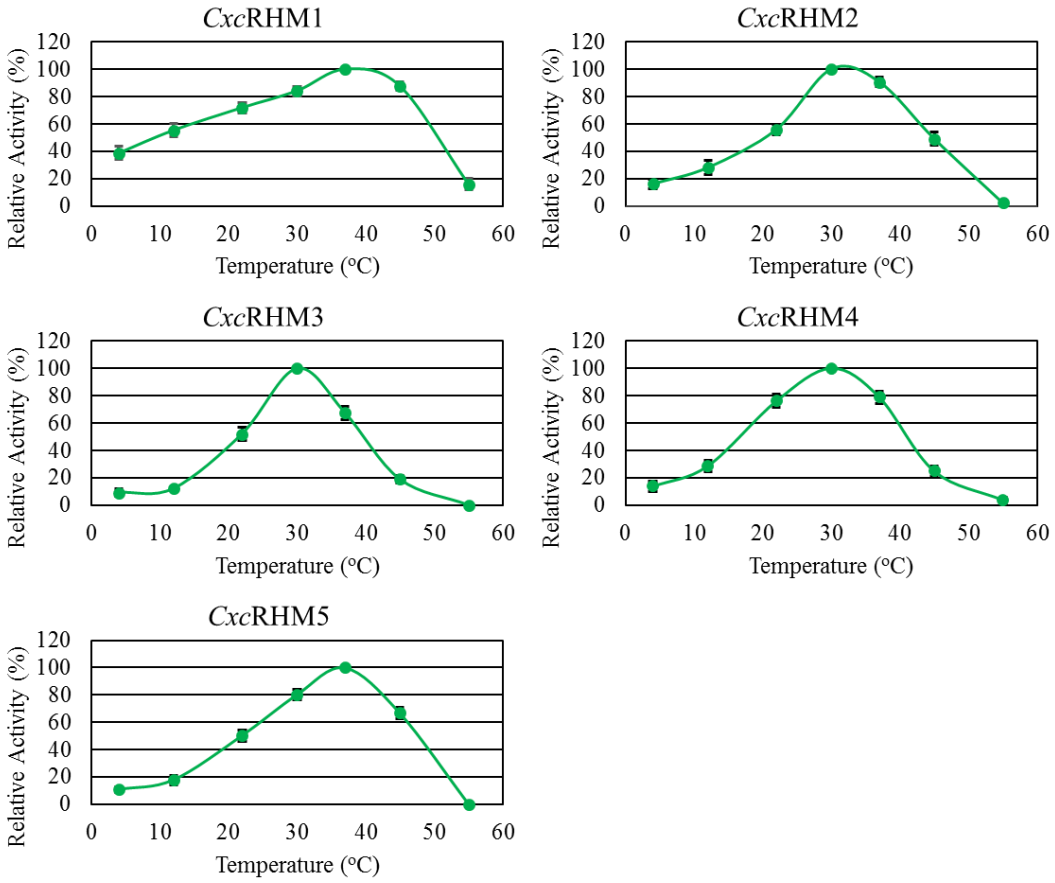


Figure 3.13: Temperature optimum of *Cxc*UXS activity *in vitro*. The activity of the recombinant *Cxc*RHMs was analyzed at different temperatures. Assays were performed with 5 replicates for each enzyme and each of the seven different temperatures. Results are shown as the calculated mean value with error bars representing standard error. 100% relative activity corresponds to the level of activity observed at the optimum tested temperature for a given enzyme.

In assessing optimal pH, relative activity assays identified a broad range of activities between pH 5.5 – 7.5 for all *Cxc*RHMs with the highest level of activity observed at pH 7.0, 7.5, 8.0, and 9.0 for *Cxc*RHM1 – 5 (Fig. 3.13; Table S3.6).

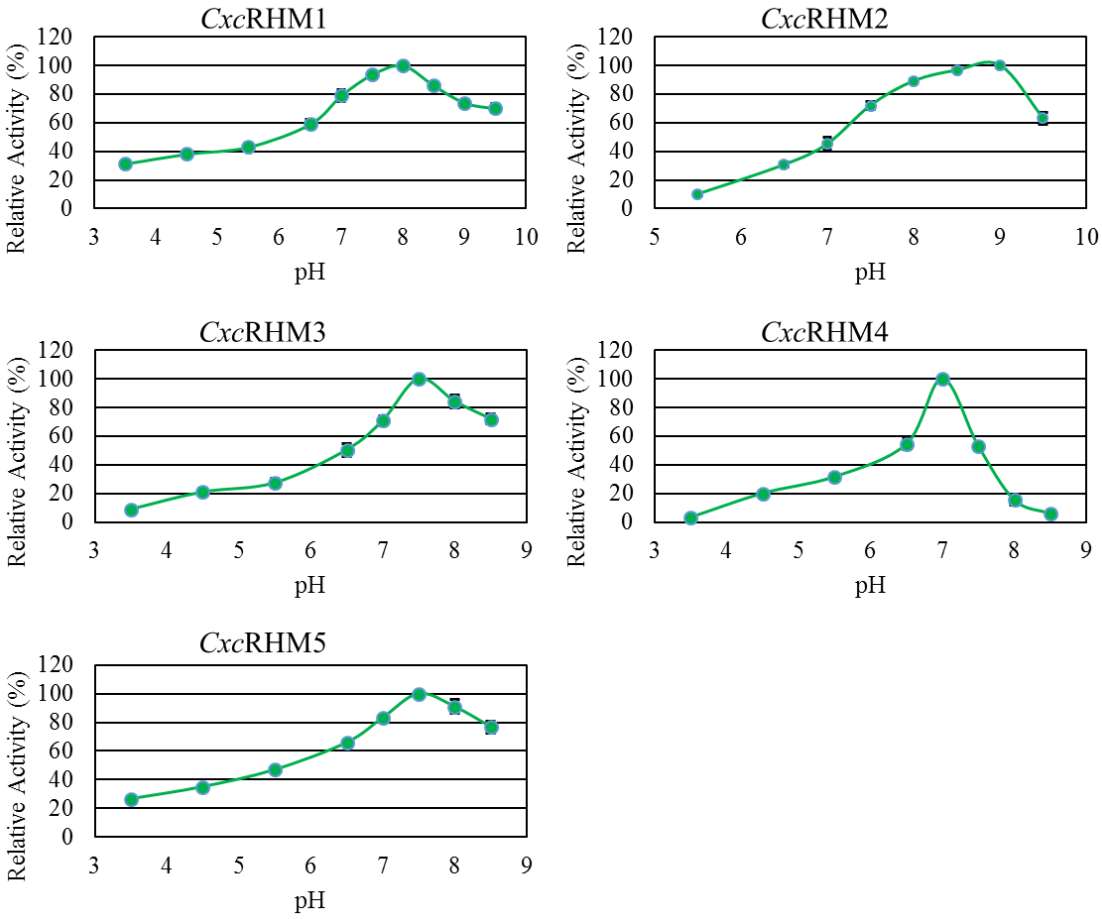


Figure 3.14: pH optimum of *Cxc*RHM activity *in vitro*. The activity of the recombinant *Cxc*RHMs was analyzed at different pHs. Assays were performed with 5 replicates for each enzyme and each of the eight different pH conditions. Results are shown as the calculated mean value with error bars representing standard error. 100% relative activity corresponds to the level of activity observed at the optimum tested pH for a given enzyme.

Performing assays at optimum temperature and pH, *Cxc*RHM1 appeared to have the highest relative activity (100%) for producing UDP-Rha, with comparative relative activities of *Cxc*RHM2, *Cxc*RHM3, *Cxc*RHM4, and *Cxc*RHM5 observed at 58.1 ± 4.9%, 62.7 ± 3.5%, 2.6 ± 0.4%, and 79.5 ± 3.8% activity, respectively (Table 3.4). Each of the *Cxc*RHM also produced UDP-4K6DG at a range of different levels. Based on peak integration of the extracted ion chromatographs for UDP-Rha and UDP-4K6DG in each assay, the relative

amount of UDP-4K6DG produced compared to UDP-Rha was 4.8 ± 2.4, 204 ± 44.0, 97.3 ± 14.9, 21.7 ± 5.3, and 82.9 ± 19.6% for *Cxc*RHM1 – *Cxc*RHM5, respectively (Table 3.4). These results suggest that of the enzymes tested, *Cxc*RHM1 is the most active at producing UDP-Rha.

Table 3.4: Enzymatic properties and relative activities of *Cxc*RHM1 – 5.
[1]Based on the relative activity at optimal conditions compared to *Cxc*RHM1.
[2]Relative production level listed are based on comparison of UDP-4K6DG and UDP-Rha levels produced by each enzyme as determined by LC-MS peak integration.

Property	*Cxc*RHM1	*Cxc*RHM2	*Cxc*RHM3	*Cxc*RHM4	*Cxc*RHM5
Optimal pH	8.0	9.0	7.5	7.0	7.5
Optimal Temperature (°C)	37.0	30.0	30.0	30.0	37.0
Relative Activity[1]	100 ± 0.0	58.1 ± 4.9	62.7 ± 3.5	2.6 ± 0.4	79.5 ± 3.8
Relative UDP-4K6DG Production[2]	4.81 ± 2.40	204.2 ± 44.0	97.3 ± 14.9	21.7 ± 5.3	82.9 ± 19.6

Because the UDP-4K6DG intermediate is the N-terminal active site's product and the substrate for the C-terminal site, *Cxc*RHM enzyme assay product profiles containing high levels of UDP-4K6DG suggest the N-terminal active site are less efficient than the enzyme's C-terminal active site. On this basis, it can be suggested that *Cxc*RHM1 and *Cxc*RHM4 possess C-terminal active sites which have similar or higher efficiencies than their N-terminal active sites, while *Cxc*HM2, *Cxc*RHM3, and *Cxc*RHM5 appear to have C-terminal active sites with lower efficiencies than their N-terminal active sites.

3.3.7 NSE-UGT Coupled Assays

With the successful identification of *C. x crocosmiiflora* UXSs and RHMs, enzyme assays coupled with GT1 UGTs were performed to assess if and how these newly identified UXSs and RHMs enzymes may be used in subsequent work characterizing novel *C. x crocosmiiflora* GT1 UGTs.

To test the ability of the *Cxc*UXS family enzymes and a UDP-xylosyltransferase to work in the same reaction, enzyme assays of each recombinant *Cxc*UXS with UDP-GlcA, and the NAD^+ cofactor, were combined with *C. x crocosmiiflora* UGT7. *Cxc*UGT7, was idependently shown to xylosylate myricetin (see section 4.3.3). The products of the coupled assays were analyzed by LC-MS against negative controls consisting of both purified enzyme

extracts of *E. coli* expressing an empty pET28b(+) vector and enzyme extracts of *E. coli* expressing an empty pASK-IBA37plus vector. In addition to UDP-Xyl, LC-MS analysis of these coupled assays revealed a new product absent in any of the controls, with retention time and mass spectra suggesting a myricetin xyloside (Fig. 3.14). This new product suggests that *Cxc*UGT7 was able to utilize UDP-Xyl formed in the coupled reaction as a UDP-sugar donor. The exact site of xylosylation of myricetin will require additional development of an authenticated standard or NMR analysis, which is not part of this chapter.

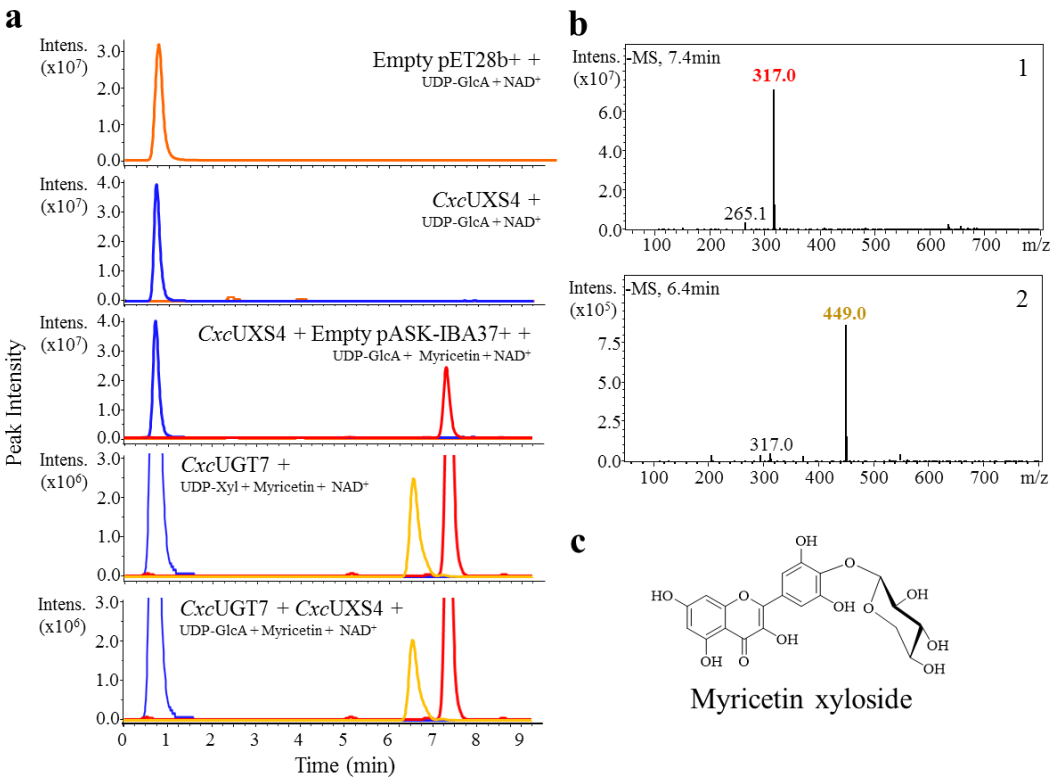


Figure 3.15: Representative regions of extracted ion LC-MS chromatograms and corresponding mass spectra of products formed in enzyme assays of *Cxc*UXS4 coupled with *Cxc*UGT7. (a) Protein derived from *E. coli* expressing both empty pET28b(+) and pASK-IBA37plus control vectors, *Cxc*UXS4, and *Cxc*UGT7 were incubated overnight with 1 mM UDP-glucuronic acid, 1 mM NAD^+, and 100 µM myricetin. Orange, blue, gold, and red traces represent extracted ion chromatograms for m/z of 579.0 [M-H](superscript -), 535.0 [M-H](superscript -), 449.0 [M-H](superscript -), and 317.0 [M-H](superscript -), respectively. Based on comparison of retention times and mass spectra against analytical standards, the peaks identified in orange, blue, and red traces were confirmed to be UDP-glucuronic acid, UDP-xylose, and myricetin, respectively. (b) Mass spectra of enzyme assay products. The spectra presented in "1" and "2" are the background subtracted mass spectra for chromatographic peaks corresponding to myricetin and an unknown myricetin xyloside (theoretical molecular weight of each is 317.24 and 450.35). (c) Representative structure of a myricetin xyloside. The exact xylosylation position of myricetin in the reaction product shown in (a) and (b) remains to be determined.

To test the ability of the *Cxc*RHM family enzymes and a UDP-rhamnosyltransferase to work in the same assay, enzyme assays of each recombinant *Cxc*RHM with UDP-Glc, and the NAD^+ and NADPH cofactors, were combined with *At*UGT78D1, a previously characterized UDP-rhamnosyltransferase. The products of the couple assays were analyzed by LC-MS against negative controls consisting of both purified enzyme extracts of *E. coli* expressing an empty pET28b(+) vector and enzyme extracts of *E. coli* expressing Gan empty pASK-IBA37plus vector. In addition to UDP-Rha and UDP-4K6DG, LC-MS analysis revealed a new product in the coupled assays, which was absent in any of the controls, with retention time and mass spectra identical to the myricetin-3-O-rhamnoside analytical standard (Fig. 3.15). The presence of this new compound shows *At*UGT78D1 had access to UDP-Rha and could utilize it as a sugar donor.

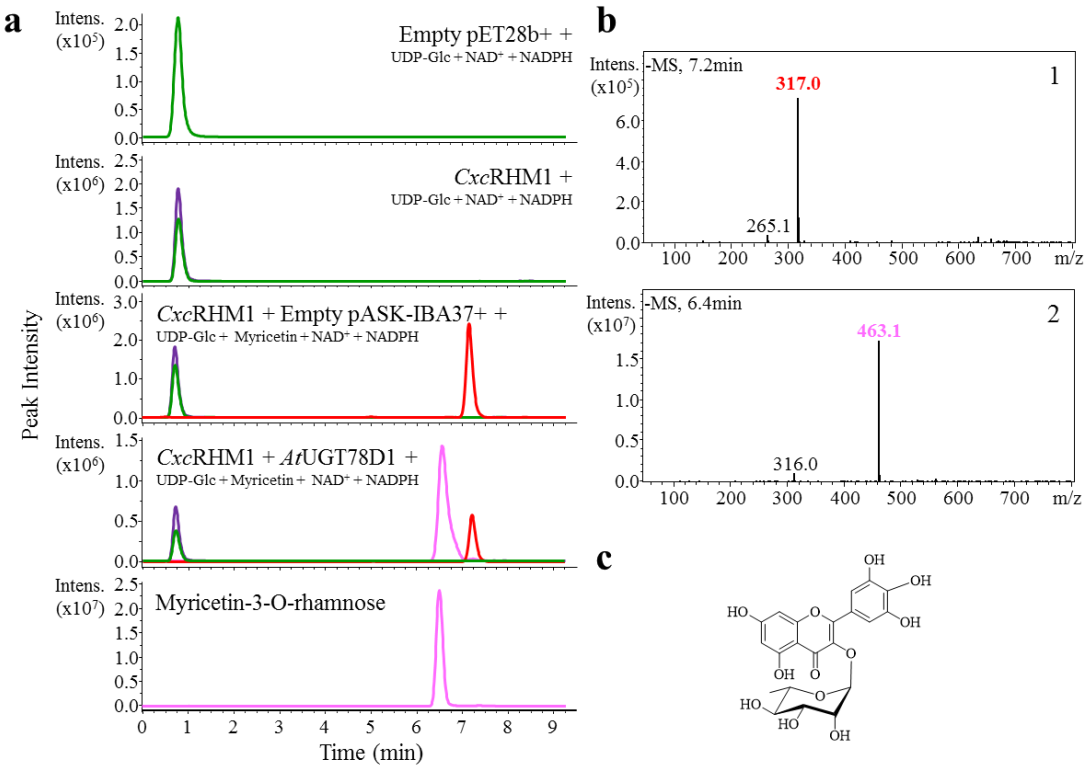


Figure 3.16: Representative regions of extracted ion LC-MS chromatograms and corresponding mass spectra of products formed in enzyme assays of *Cxc*RHM1 coupled with *At*UGT78D1. (a) Protein derived from *E. coli* expressing both empty pET28b(+) and pASK-IBA37plus control vectors, *Cxc*RHM1, and *At*UGT78D1 were incubated overnight with 1 mM UDP-glucose, 1 mM NAD^+, 1 mM NADPH, and 100 µM myricetin. Green, purple, pink, and red traces represent extracted ion chromatograms for m/z of 565.0 [M-H](superscript -), 549.0 [M-H](superscript -), 463.0 [M-H](superscript -), and 317.0 [M-H](superscript -), respectively. Based on comparison of retention times and mass spectra against analytical standards, the peaks identified in green, pink, and red traces were confirmed to be UDP-glucose, myricetin-3-O-rhamnoside, and myricetin, respectively. (b) Mass spectra of enzyme assay products. The spectra presented in "1" and "2" are the background subtracted mass spectra for chromatographic peaks corresponding to myricetin and myricetin-3-O-rhamnoside (theoretical molecular weight of each is 317.24 and 463.34). (c) Structure of myricetin-3-O-rhamnoside.

3.4 DISCUSSION

Using transcriptome mining and biochemical approaches, I investigated the *C. x crocosmiiflora* UXS and RHM gene families. Plant species that have had their UXS families previously characterized show between four and seven different UXSs (Bindschedler *et al.*, 2005; Harper and Bar-Peled, 2002; Yin and Kong, 2016; Zhang *et al.*, 2005). Analysis of the *C. x crocosmiiflora* transcriptome showed a similar pattern with the identification of at least five putative *Cxc*UXS across the three known phylogenetic clades. Comparison with non-

plant species suggests an expansion of the UXS gene family in plants, which may have begun with the duplication of a common UXS ancestor and led to the present multigene UXS family containing both transmembrane and cytosolic proteins (Du *et al.*, 2013; Gu *et al.*, 2010). The presence and conservation of a multi-clade UXS gene family across different plant species is indicative of potentially non-redundant or only partly redundant functions of the individual UXS members.

Characterization of *Cxc*UXS functions and analysis of their transcript expression in the transcriptome data, which showed some level of differential expression (Fig. S3.9), supports the possibility of the *Cxc*UXSs having different roles *in planta*. *Cxc*UXS2 was the only transmembrane UXS identified in the *C. x crocosmiiflora* transcriptome. The presence of a transmembrane NSE is interesting as plants typically use nucleotide sugar transporters (NSTs) to transport nucleotide sugars from the cytosol into the Golgi apparatus and endoplasmic reticulum lumen (Bakker *et al.*, 2005; Baldwin *et al.*, 2001; Handford *et al.*, 2004; Norambuena *et al.*, 2005; Norambuena *et al.*, 2002; Rollwitz *et al.*, 2006). While a transmembrane UXS and UDP-Xyl NST would appear redundant, multiple such genes have been identified in *A. thaliana* (Ebert *et al.*, 2015). The need for this transporter could be due to the common phenomenon that transmembrane UXSs, like *Cxc*UXS2, appear to have the lowest kinetic efficiencies (Oka *et al.*, 2007; Pattathil *et al.*, 2005; Zhang *et al.*, 2005). Accordingly, the presence of a NST could better support Golgi apparatus activities when higher levels of UDP-Xyl are needed, such as during active cell wall biosynthesis when xylan production is occurring. An alternative explanation can be derived from the observation that a mutant *A. thaliana* lacking one of these UDP-Xyl NST resulted in a viable phenotype with a reduced glucuronoxylan biosynthesis, but no affect on xyloglucan biosynthesis (Ebert *et al.*, 2015). The specific reduction on xylan biosynthesis suggest the UDP-Xyl produced by transmembrane UXSs could be selectively employed in specific activities within the Golgi apparatus. While the cytosolic *Cxc*UXSs appear to have lower substrate affinity, all three of these enzymes possess higher k_{cat} values than the transmembrane enzyme. Expression analysis shows that transcripts encoding *Cxc*UXS3 and *Cxc*UXS4 have the highest relative expression levels in the flower, stem, and stolon, while *Cxc*UXS5 has the relative highest expression in the corm.

*Cxc*UXS1 appears to function as a UDP-4-keto pentose synthase (UDP-4KPS), despite it clustering with the clade A transmembrane UXSs. Exploring possible reasons for this dichotomy in phylogenetic assignment and the observed UDP-4KP function, sequence analysis identified a unique glutamic acid in the 252 amino acid position. This bulky and charged residue could affect the active site by altering the positioning of the catalytic Tyr^{249}, affecting its ability to interact with the UDP-4KP intermediate. Previous work on human UXS has shown that if the UDP-4KP intermediate is no longer able to interact with the hydroxyl group on the catalytic tyrosine, the UDP-Xyl reaction will prematurely terminate to form UDP-4KP as the product (Eixelsberger *et al.*, 2012). Such a functional link between Tyr^{249} and Glu^{252} in *Cxc*UXS1 was supported by the observation that the mutants *Cxc*UXS1(Δ1–69; E252G) and *Cxc*UXS1(Δ1–81; E252G) showed altered product profiles with UDP-Xyl as the predominant product. Accordingly, future work is warranted to explore the effect Glu^{252} has on *Cxc*UXS1's active site and UDP-4KP biosynthesis.

While it is possible that the observed *Cxc*UXS1 activity as a UDP-4KPS is the result of a potentially improperly folded protein produced in *E. coli*, possibly due to the N-terminal truncation and His_6-tag addition, attempts to alleviate these concerns included employing unpurified protein assays, eliminating imidazole from all buffers used, increased concentration of protein stabilizing agents, as well as variations in enzyme assay pH, temperature, buffers, substrate concentrations, and cofactor concentrations all resulted in the same product profile. These results support the possibility that *Cxc*UXS1 is a UDP-4KPS that evolved from a UDP-xylose synthase. While UDP-4KP has been identified in other organisms (Gu *et al.*, 2010), a possible biological role for this enzyme in the plant's metabolism needs to be identified before the characterization can be definitive. Ideally, a biological role could be tested by altering the expression of *Cxc*UXS1 in *Crocosmia*; however, a transformation system to affect gene expression in this species is not yet available. Expression of *Cxc*UXS1 in better characterized plant model systems, such as *A. thaliana* or tobacco, may provide some information about a possible role in planta.

Based on the relatively high protein sequence identity to *Arabidopsis* RHMs, I predicted similar functions for *Cxc*RHM1 – *Cxc*RHM5. This general prediction was confirmed by functional characterization when all five *Cxc*RHMs were found to form both UDP-Rha and UDP-4K6DG as primary and secondary products. Previous work on the RHM family showed

that the N-terminal domain of the enzyme is responsible for the initial 4,6-dehydration reaction while the C-terminal domain is responsible for the 3,5-epimerization and 4-keto-reduction reactions (Oka *et al.*, 2007; Watt *et al.*, 2004). Having explored all putative RHM identified within the *C. x crocosmiiflora* transcriptome, it is interesting to note the variability between the ratios of UDP-Rha and UDP-4K6DG produced across the different *Cxc*RHMs (Table. 3.4). This observation suggested the two active sites have different specific activities in the different enzymes. This results in a range of combinations where the 3,5-epimerase/4-keto-reductase activity outpaces the 4,6-dehydratase, producing product profiles with almost no observed UDP-4K6DG intermediate, or the opposite where the levels of the intermediate appear to be double that of the final product. Analysis of the *C. x crocosmiiflora* transcriptome also identified a *Cxc*UER abundantly expressed across all organs tested. The presence of a gene containing the 3,5-epimerase and 4-keto-reductase activity suggests a possible supporting role, in a functional context, for *Cxc*RHMs whose C-terminal appear to have lower kinetic efficiencies. This would help prevent the accumulation of the UDP-4K6DG intermediate in the biosynthesis of UDP-Rha. The possible significance of such a supporting role of a *Cxc*UER as a "sweeper enzyme", may be supported by an observations with yeast cells expressing *At*RHM2-N leading to increased biosynthesis of UDP-4K6DG (Oka *et al.*, 2007). These yeast cells showed aggregation and a slow growth phenotype (Oka *et al.*, 2007), suggesting accumulation of UDP-4K6DG may have a negative effect *in vivo*.

In prokaryotes, NDP-rhamnose is biosynthesized from NDP-glc through three sequential enzymes, rfbB, rfbC and rfbD (Reeves *et al.*, 1996; Stevenson *et al.*, 1994), each performing one of the reactions outlined in figure 3.2. Analysis of bacterial genomes shows that these three genes typically cluster together (Yin *et al.*, 2011). This suggests that an ancient eukaryotic ancestor acquired and retained a DNA fragment containing the *rfbB*, *rfbC* and *rfbD* genes which eventually underwent a fusion event to form *RHM*. The presence of this trifunctional RHM in plants and some fungi and the identification of bifunctional 3,5-epimerase/4-keto-reductases-like enzymes in other non-plant organisms (Yin *et al.*, 2011), as well as the identification of other NSEs employing the same biosynthetic mechanism to produce different NDP-sugars, suggests that the first gene fusion event occurred between the 3,5-epimerase and 4-keto-reductase ancestors. While the present day RHM may then have formed by additional fusion with the 4,6-dehydratase, the current plant UER was either

retained from an ancestral 3,5-epimerase/4-keto-reductase or the result of a RHM gene duplication and subsequent partial deletion. Phylogenetic analysis of the C-terminal of plant RHM and UER showed clear separation of the UER and RHM enzymes, suggesting the *UER* gene arose through retention or that the duplication/partial deletion event occurred before the expansion to a multigene family (Fig. S3.10).

Glycosylation is one of the most prevalent and important modifications in specialized metabolism (Bowles *et al.*, 2005; Bowles *et al.*, 2006). In the last two decades, the development of microbial platforms for large-scale production of high-value specialized metabolites has been a major emphasis in academic and industry laboratories (Han *et al.*, 2014; Kim *et al.*, 2012; Pandey *et al.*, 2013). A limiting factor in some microbial production systems is inadequate supply of UDP-sugars, which results in their rapid depletion when GT1 UGTs are overexpressed (De Bruyn *et al.*, 2015). To this end, the metabolite production could potentially be improved with the introduction of additional UDP-sugar biosynthetic activities. The work presented in this chapter provides functional characterization for a suite of both UXSs and RHMs, which could be used for this purpose. Comparison of the *Cxc*UXS's kinetic capabilities to available data in the literature shows that *Cxc*UXS4 could be particularly useful enzyme for metabolic engineering. Currently, kinetic analysis of characterized plant UXS has focused on K_M and not k_{cat}. For soluble UXSs, a range of 400 – 890 μM has been observed for K_M values (Harper and Bar-Peled, 2002; Hayashi and Matsuda, 1988; John *et al.*, 1977; Yin and Kong, 2016; Kuang *et al.*, 2016). Expanding this comparison to non-plant UXS, catalytic efficiencies of between $5.2*10^{-5}$ – 0.0183 $\mu M^{-1}\ s^{-1}$ have been characterized (Duan *et al.*, 2015; Eixelsberger *et al.*, 2012; Gu *et al.*, 2010; Gu *et al.*, 2011; Oka and Jigami, 2006). Based on *Cxc*UXS4's K_M of 147 μM and catalytic efficiency of 0.036 $\mu M^{-1}\ s^{-1}$, it appears to be a strong candidate for incorporation into a microbial MbA production system. While the kinetic parameters of the *Cxc*RHMs could not be compared to previously characterized RHM-dehydratase or -epimerase/reductase domains (Han *et al.*, 2015; Martinez *et al.*, 2012), *Cxc*RHM1 shows the advantageous trait of producing relatively high levels of UDP-Rha and very low levels of the potentially cytotoxic UDP-4K6DG intermediate. As such, future work is warranted to explore how these genes could be used in a potential microbial production system for MbA.

3.5 CONCLUSION

In conclusion, the work presented in this chapter identified and functionally characterized the *C. x crocosmiiflora* nucleotide sugar interconversion enzymes responsible for the biosynthesis of UDP-xylose and UDP-rhamnose, two UDP-sugars critical in the biosynthesis of MbA. Of the enzymes identified as UXSs, *Cxc*UXS2 – *Cxc*UXS5 were found to produce UDP-xylose, and *Cxc*UXS1 produced UDP-4-keto pentose. Site-directed mutagenesis of *Cxc*UXS1 revealed a potential evolutionary route of how the function of *Cxc*UXS1 may have evolved and identified a Glu^{252} near the catalytic Tyr^{249} as involved in defining the UDP-4-keto pentose product. Of the enzymes identified as RHMs, *Cxc*RHM1 – *Cxc*RHM5 catalyzed all three steps in the formation of UDP-Rha, while *Cxc*UER1 was identified as performing the final 3,5-epimerase and 4-keto-reductase reactions needed to produce UDP-Rha. Kinetic and relative activity characterization of these enzymes, identified *Cxc*UXS4 and *Cxc*RHM1 as the most efficient enzymes, making them good candidates for future use in the metabolic engineering towards MbA biosynthesis. Overall, this work contributes to our understanding of the *C. x crocosmiiflora* MbA biosynthetic pathway, and provides proof-of-concept for the use of these gene families for the future characterization of UDP-xylosyl/rhamnosyltransferase in the MbA biosynthetic pathway.

CHAPTER 4: IDENTIFICATION OF CROCOSMIA x CROCOSMIIFLORA UDP-GLYCOSYLTRANSFERASES INVOLVED IN MONTBRETIN A BIOSYNTHESIS

Uridine diphosphate glycosyltransferases (UGTs) of the family 1 glycosyltransferases (GT1) are essential enzymes in the process of glycosylating specialized metabolites. Glycosylation is a common modification in specialized metabolism and affects the physical, chemical and biological properties of the aglycone. Montbretin A (MbA) is a highly glycosylated flavonoid found in *C. x crocosmiiflora* and of interest for its type 2 diabetes therapeutic potential. In this chapter, I explore the *C. x crocosmiiflora* GT1 UGT family with the goal to identify candidate GT1 UGTs involved in the biosynthesis of MbA. Phylogenetic analysis of the *Cxc*UGTs provided insight into the unique pattern of GT1 UGT clustering, showing rampant expansion in group D and an absence of any group H GT1 UGTs. Activities of 14 candidate *Cxc*UGTs, which were identified through association analysis between MbA accumulation and transcript expression, were tested using eight different potential MbA biosynthetic intermediates. While minor activity was observed within assays using myricetin or myricetin-3-rhamnoside as acceptors, the low levels or lack of activity observed suggest that none of these candidate GT1 UGTs are involved in MbA biosynthesis. As this negative result may be due to errors in the underlying hypotheses of the association analysis, future work should focus on identifying specific conditions that affect MbA accumulation levels for use as a model of MbA biosynthetic gene expression. Overall, this work contributes to our understanding of the phylogenetic distribution of GT1 UGTs within vascular plants and facilitates further identification and functional characterization of *Cxc*UGTs involved in MbA biosynthesis.

4.1 INTRODUCTION

Specialized metabolites are an important natural resource with numerous applications such as medicines and nutraceuticals. With an estimated 200,000 – 300,000 different specialized metabolites found in the plant kingdom (Dixon and Strack, 2003; Lawrence, 1964), plants typically employ combinations of large enzyme families to perform the arrays of biosynthetic core reactions and specific modifications. Two of the most common

modifications used in specialized metabolism are esterification and glycosylation. Both modifications can have a large range of effects on a metabolite's properties. Acylation may enhance volatility or biological activity (D'Auria, 2006), while glycosylation typically affects a metabolite's solubility, potential for accumulation, subcellular localization, or biological activity (Gachon *et al.*, 2005; Ghose *et al.*, 2014; Liang *et al.*, 2015; Martinoia *et al.*, 2000). Montbretin A (MbA) is an excellent example of a specialized metabolite, which requires both types of modifications during its biosynthesis. The acylation modification incorporates a caffeic acid moiety, altering a known biological activity of MbA by drastically increasing its ability to inhibit human pancreatic amylase (HPA) through H-bonding and internal π-stacking interactions (Williams *et al.*, 2015). The glycosylation modifications during MbA biosynthesis incorporate a total of five sugar moieties, altering the metabolite's solubility and biological activity through, respectively, increasing its ability to engage in H-bonding with water, as well as more effectively orienting and stabilizing the inhibitory myricetin and caffeic acid moieties in HPA's active site (Williams *et al.*, 2015).

The glycosylation of specialized metabolites is typically performed by uridine diphosphate glycosyltransferases (UGTs) that are members of the family 1 glycosyltransferases (GT1). These enzymes act by using uridine 5′-diphosphate sugars (UDP-sugars) as donor molecules and transfer their sugar moiety to acceptors (Bowles *et al.*, 2005; Bowles *et al.*, 2006). The majority of plant GT1 UGTs contain a conserved 44 amino acid motif known as the plant secondary product glycosylation (PSPG) motif which encompasses the amino acids responsible for binding the UDP-sugar (Campbell *et al.*, 1997; Hughes and Hughes, 1994). Plant GT1 UGTs glycosylate a large diversity of metabolites, and the overall level of sequence identity across the plant GT1 UGT family, which is annotated with several subfamilies, is relatively low (Coutinho *et al.*, 2003; Lim *et al.*, 2003; Vogt and Jones, 2000). DNA sequencing has revealed some of the extent of the UGT family expansion in the plant kingdom. Current estimates from species with characterized genomes suggest most species possess between 100 – 250 UGTs with an average of 0.5% of predicted genes in a given plant genome being UGTs (Caputi *et al.*, 2012).

Based on the structure of MbA, its biosynthesis is thought to involve activities of up to five different UGTs (Fig. 4.1; Fig S4.1). Harvest from field grown plants is currently not a suitable method for production of large quantities of MbA. Thus, employing the genes that

Crocosmia x crocosmiiflora uses in the biosynthesis of MbA towards either an improved *C. x crocosmiiflora* production system with heightened levels of MbA or to engineer a heterologous host to produce MbA are attractive options. A critical step towards this goal is the identification of the MbA biosynthetic *Cxc*UGTs. However, elucidating these *Cxc*UGTs faces four major challenges associated with: (i) the large number of GT1 UGTs typically found in any plant species, (ii) low sequence identity between GT1 UGTs, (iii) lack of clarity on the MbA biosynthetic reaction pathway (Fig. 4.1; Fig S4.1), and (iv) the lack of prior knowledge and data available for *Crocosmia* genes.

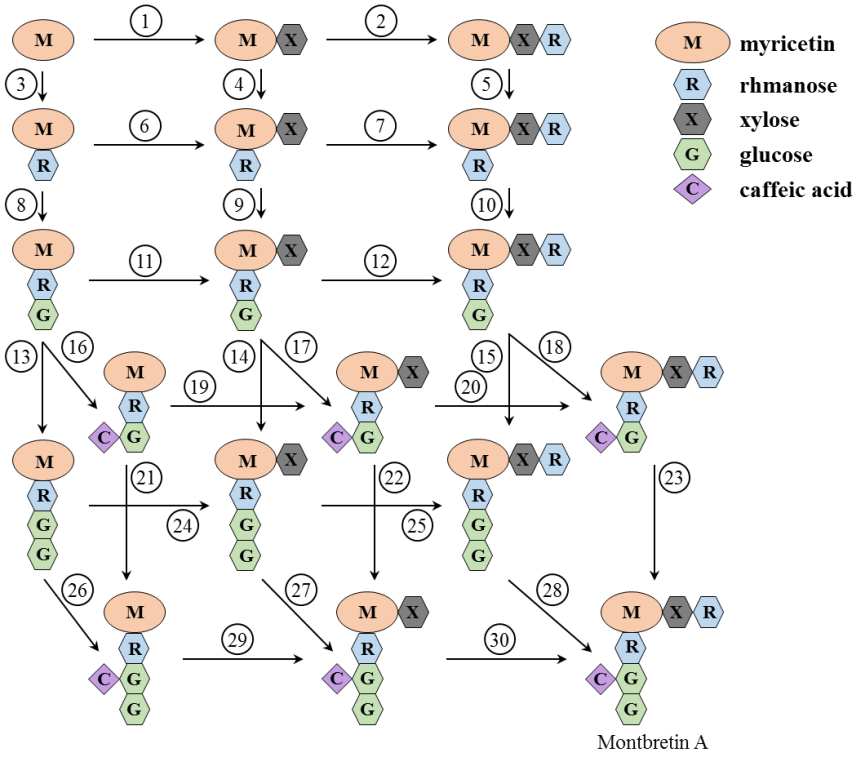


Figure 4.1: Theoretical routes of montbretin A biosynthesis. Considering all potential steps needed to form montbretin A starting with myricetin, and without prior knowledge of the *in planta* intermediates of the pathway, the biosynthetic pathway is represented here by as multi-dimensional matrix. Circled numbers are used as denotations for potential individual steps in the biosynthesis of MbA.

While many approaches can be used in the elucidation of target specialized metabolite biosynthetic genes, advances in sequencing technologies have permitted the integration of deep transcript and targeted metabolite profiles from corresponding tissues to be employed as a method for identifying a relevant list of candidate genes (Facchini *et al.*, 2012; Saito *et al.*, 2008). This approach employs the "guilt-by-association" principle, which proposes that a set

of genes involved in the target metabolite biosynthesis are co-regulated, and thus co-expressed under the control of a shared regulatory system (Yonekura-Sakakibara and Saito, 2009). Several studies have successfully employed this approach as a strategy for identifying genes involved in target metabolite biosynthesis (Augustin *et al.*, 2015; Kilgore *et al.*, 2014; Zerbe *et al.*, 2014; Zerbe *et al.*, 2012). With this approach, a critical factor for successful application is the identification of specific tissues or defined environmental or growth conditions in which biosynthetic activities of the target metabolite differ and correlate to accumulation.

Previous work on *C. x crocosmiiflora* has focused on establishing resources able to support the elucidation of MbA biosynthetic genes. Metabolite profiling and MALDI-imaging have identified that throughout the year, MbA is primarily found in the peripheral tissues of the corm, as well as at minor levels in the other organs (section 2.3.1). A first set of transcriptome assemblies of different organs of the plant was also developed (section 2.3.3). Employing these resources, a guilt-by-association analysis identified 14 putative GT1 UGTs that possess transcript expression patterns highly correlated to MbA accumulation across different *C. x crocosmiiflora* organs (section 2.3.5). To this end, using functional genomic and biochemical approaches this chapter expands on these resources to explore the potential activity of these 14 candidate *Cxc*UGTs in MbA biosynthesis. The findings herein help lay a foundation for a better understanding of MbA biosynthesis in *C. x crocosmiiflora* and the employment of a guilt-by-association approach to identify the MbA biosynthetic GT1 UGTs.

4.2 EXPERIMENTAL

4.2.1 Cloning of *C. x crocosmiiflora* UGTs

Candidate *Cxc*UGT cDNAs were amplified using primers (Table S4.1) designed according to putative GT1 UGT transcripts identified in the transcriptome assemblies with adaptors for cloning into the pASK-IBA37+ vector. If full-length sequences were not available, 5′ or 3′ RACE-PCR were performed with the Marathon cDNA amplification kit (Clonetech, www.clontech.com) per the manufacturer's instructions. Once full-length sequences were obtained, GT1 UGT transcripts were amplified by PCR and cloned into the pASK-IBA37+ expression vector (IBA Life Sciences, www.iba-lifesciences.com) in-frame with an N-terminal His_6-tag. For candidate genes containing *Bsa*I restriction sites, site-directed

base changes were made using primers resulting in a synonymous mutation to remove the site (Table S4.1). Sequences and gene insertion orientation were verified by Sanger sequencing.

4.2.2 Phylogenetic Analysis

For phylogenetic analyses, amino acid sequence alignments were generated using ClustalW (Thompson *et al.*, 1994). Phylogetic analysis was performed using a maximum likelihood algorithm in the MEGA 7.0 (http://www.megasoftware.net) (Kumar *et al.*, 2016) using uniform rate variation among sites, LG substitution model, BIONJ/NJ starting tree, and 1000 bootstrap repetitions. To best establish the GT1 UGT phylogenetic groups in the *Cxc*UGT family, full-length amino acid sequences of 34 previously defined GT1 UGTs were included in the analysis. These non-*Crocosmia* sequences included 15 *Arabidopsis thaliana* GT1 UGTs (*At*UGT72C1, *At*UGT73B1, *At*UGT73C1, *At*UGT75D1, *At*UGT76B1, *At*UGT78D3, *At*UGT82A1, *At*UGT83A1, *At*UGT85A4, *At*UGT86A1, *At*UGT87A1, *At*UGT89B1, *At*UGT90A1, *At*UGT91A1, and *At*UGT92A1), 17 *Zea mays* GT1 UGTs (GRMZM2G021786, GRMZM2G022242, GRMZM2G035755, GRMZM2G041699, GRMZM2G046994, GRMZM2G050748, GRMZM2G058314, GRMZM2G061289, GRMZM2G073376, GRMZM2G120016, GRMZM2G159404, GRMZM2G173315, GRMZM2G175910, GRMZM2G301148, GRMZM2G479038, GRMZM5G834303, and GRMZM5G892627), and 2 *Oryza sativa* GT1 UGTs (*Os*07g30690 and *Os*07g46610). Amino acid sequence alignments were visualized using CLC Bio Main Workbench (https://www.qiagenbioinformatics.com/products/clc-main-workbench) while phylogenetic trees weres visualized using the Interactive Tree of Life software (http://itol.embl.de/) (Letunic and Bork, 2011).

4.2.3 Production of Potential MbA Intermediates

Enzymatic and base-catalyzed reactions performed to produce potential MbA intermediates were performed as previously described (Williams *et al.*, 2015) with minor modifications. The β-glucosidase, naringinase, and β-xylosidase enzymes used herein were provided by Dr. Stephen Withers (UBC, Department of Chemistry).

Deglucosylation reactions employing *Agrobacterium sp.* β-glucosidase were performed in 50 mM sodium phosphate buffer (pH 6.8, 0.1% bovine serum albumin) with 40

μg β-glucosidase per mg substrate. Reactions were incubated at 37°C for 48 hours. Enzymes were precipitated by addition of methanol in a volume equal to 25% of the reaction volume. The resulting solution was filtered by passing through a 0.22 μm hydrophilic polypropylene membrane filter (http://www.pall.com/), evaporated, resolved in 1 mL of H_20 / 25 mg, and subsequently purified as indicated below.

Derhamnosylation reactions employing *Penicillium decumbens* naringinase were performed in 50 mM acetate phosphate buffer (pH 5.6) with 5 μg naringinase per mg substrate. Naringinase was first incubated in buffer (5 mg / mL) at 60°C for 2 hours to inactivate its β-D-glucosidase activity without affecting its α-L-rhamnosidase activity. After partial inactivation, naringinase was added to the substrate and reacted for 48 hours at 22°C. Enzymes were precipitated by addition of methanol in a volume equal to 25% of the reaction volume. The resulting solution was filtered by passing through a 0.22 μm hydrophilic polypropylene membrane filter (http://www.pall.com/), evaporated, resolved in 1 mL of H_20 / 25 mg, and subsequently purified as indicated below.

Dexylosylation reactions employing *Bacillus halodurans* β-xylosidase were performed in 50 mM sodium phosphate (pH 6.8, 50 mM sodium chloride) with 2 μg xylosidase per mg substrate. Reactions were incubated at 37°C for 48 hours. Enzymes were precipitated by addition of methanol in a volume equal to 25% of the reaction volume. The resulting solution was filtered by passing through a 0.22 μm hydrophilic polypropylene membrane filter (http://www.pall.com/), evaporated, resolved in 1 mL of H_20 / 25 mg, and subsequently purified as indicated below.

Removal of the caffeic acid moiety was performed by base catalysis. MbA was dissolved in dry methanol (2 mg / mL) under N_2 gas in a round bottom flask. Sodium methoxide was added until a concentration of ~27.5 mM was reached and left to react for 5 hours. The reaction was quenched by addition of Amberlite IR-120 (H+) to the solution until it possessed a pH ≈ 5. The methanol was evaporated under N_2 gas, and resulting reaction product was dissolved in H_2O. Unwanted products were removed by extraction with equal volumes hexanes, and the aqueous layer was passed through a 0.22 μm hydrophilic polypropylene membrane filter (http://www.pall.com/), evaporated, resolved in 1 mL of H_20 / 25 mg, and subsequently purified as indicated below.

Purification of reaction products was performed by high performance liquid chromatography (HPLC) on a 1260 Infinity Bio-inert Quaternary LC (Agilent). Reaction products were analyzed using an Eclipse XDB-C18 column (Agilent, 9.4 x 250 mm, 5 µm) and purified on a PrepHT XDB-C18 column (21.2 x 100 mm, 5 µm), both being run at room temperature and with flow rates of 1.0 mL min^{-1} and 4.0 mL min^{-1}, respectively. For both columns, the mobile phase used was a combination of two solvents: solvent A (H_20) and solvent B (acetonitrile). The mobile phase run was 5% solvent A by 2 min, 20% solvent A by 10 min, 40% solvent A by 40 min, 95% solvent A by 45 min, 95% solvent A by 50 min, 5% solvent A by 55 minutes, and held for 5 min, giving a total run time of 60 minutes. The effluent was monitored by UV absorption at 254 nm, and 330 nm and peak fractions were collected per the observed chromatograph.

4.2.4 *E. coli*-Based Protein Expression and Purification

Recombinant plasmids were transformed into *E. coli* C43 (www.overexpress.com) containing the pRARE 2 plasmid isolated from Rosetta 2 cells (Novagen) to negate codon bias. Individual colonies were inoculated into 5 mL of Luria Broth containing ampicillin (100 mg * L^{-1}) and grown overnight at 37°C and 200 rpm. 5 mL cultures were used to inoculate 50 mL of Terrific Broth containing ampicillin (100 mg * L^{-1}) and chloramphenicol (50 mg * L^{-1}) and cultured at 22°C and 200 rpm until an OD_{600} = ~0.5 was reached. Cultures were then cooled to 18°C, induced by addition of anhydrous tetracycline (final concentration 0.45 µM), and grown for 16 h at 180 rpm before harvesting.

Cultures were harvested by centrifugation at 4000 *x* g for 20 minutes. Supernatants were removed, and cell pellets were resuspended in 3 mL of ice-cold, 50 mM Tris-HCl extraction buffer (pH 7.5, 10% glycerol, 10 mM $MgCl_2$, 5 mM DTT, 0.2 mg lysozyme per mL, 1/40th Pierce™ Protease Inhibitor Tablet EDTA-free (ThermFischer, www.thermofisher.com) per mL, and 0.1 µL benzonase (EMD Millipore, www.emdmillipore.com/) per mL). After cell ruptured through five cycles of freeze-thawing in liquid nitrogen, lysed cells were clarified by centrifugation. Soluble protein extractions were desalted on Sephadex PD minitrap G-25 columns (GE Healthcare, http://www.gehealthcare.com), pre-equilibrated with a 50 mM Tris-HCl buffer (pH 7.5, 10% glycerol, 1 mM DTT). Protein was confirmed by SDS-PAGE using Coomassie and Western blots. Western blots were performed using His-Tag Antibody HRP

Conjugate Kit (EMD Millipore, www.emdmillipore.com) and visualized with Clarity Western ECL Blotting Substrate (Bio-rad, www.bio-rad.com). Protein concentrations were determined using a bicinchoninic acid (BCA) protein quantification assay kit (Thermo Fisher, www.thermofisher.com) employing a standard curve.

4.2.5 Recombinant UGT Enzyme Assays

To test for enzyme activity, protein assays were performed in triplicate. For reactions requiring UDP-xylose or UDP-rhamnose, nucleotide sugar interconverting enzyme (NSE)-UGT coupled reactions were used (section 3.3.7). For reactions using UDP-glucose (UDP-Glc) as their sugar donor, 100 µL reactions of unpurified protein extracts was combined with 1 mM UDP-Glc and 100 µM of the appropriate acceptor were incubated at 30°C overnight. For reactions using UDP-xylose (UDP-Xyl) as their sugar donor, 100 µL reactions of unpurified protein extracts was combined with 1 mM UDP-glucuronic acid (UDP-GlcA), 1 mM NAD^+, 1 µM *Cxc*UXS4, and 100 µM of the appropriate acceptor were incubated at 30°C overnight. For reactions using UDP-rhamnose (UDP-Rha) as their sugar donor, 100 µL reactions of isolated protein was combined with 1 mM UDP-Glc, 1 mM NAD^+, 1 mM NADPH, 5 µM *Cxc*RHM1, and 100 µM of the appropriate acceptor were incubated at 30°C overnight. Reactions were terminated by incubation at 100°C for 1 minute and the addition of 50 µL of chloroform to precipitate protein. Soluble fractions were separated by centrifugation at 1000 *x* g for 10 minutes and analyzed by LC-MS.

Reaction products from enzyme assays were analyzed by liquid chromatography (LC) (Agilent 1100 Series)/mass spectrometry detector (MSD) Trap (XCTplus) by comparison of retention times and mass spectra with authentic standards or products of previously characterized enzymes. Enzyme products were analyzed on an Agilent ZORBAX SB-C18 column (4.6 mm internal diameter, 50 mm length, 1.8 µM pore size) was used with a temperature of 50°C and flow rate 0.8 mL min^{-1}. The mobile phase used was a combination of two solvents: solvent A (H_20 + 0.2% formic acid) and solvent B (acetonitrile + 0.2% formic acid). The mobile phase run was 95% solvent A by 0.5 min, 80% solvent A by 5 min, 10% solvent A by 7 min, and 95% solvent A by 7.10 min, and held for 2.9 min, giving a total run time 10 min. Diode array detector (DAD) monitored wavelengths at 266 nm and 326 nm. The

mass spectrometer mode was negative electrospray with nebulizer pressure 60 psi, dried gas rate 12 L min^{-1}, dry temp 350ºC, and a m/z scanning range between 50 – 2000.

4.3 RESULTS

4.3.1 Identification and Phylogenetic Analysis of *Crocosmia x crocosmiiflora* UDP-Glycosyltransferases

To gain insight into the size and overall expression of the *Cxc*UGT family, I screened the complete unigene set of the *C. x crocosmiiflora* transcriptome (section 2.3.3) against the Prosite PSPG signature motif (motif #PS00375). This BLASTx search resulted in the identification of 257 putative *Cxc*UGTs. An unrooted phylogenetic tree was constructed with 158 of these putative *Cxc*UGTs, which were at least 300 amino acids in length, along with 15 *Arabidopsis thaliana*, 17 *Zea mays*, and 2 *Oryza sativa* GT1 UGTs (Fig. 4.2) to identify the *Cxc*UGT distribution among the 17 different GT1 UGT phylogenetic groups (Caputi *et al.*, 2012; Li *et al.*, 2001; Li *et al.*, 2014b). In the resulting phylogenetic tree, putative *Cxc*UGTs clustered into 13 of the 14 groups identified in *A. thaliana* (Li *et al.*, 2001), groups A – G and I – N, as well as groups O and P, which were identified through a genome-wide analysis of GT1 UGTs from 12 different species of land plants (Caputi *et al.*, 2012). Of these 15 characterized phylogenetic groups, most *C. x crocosmiiflora* GT1 UGTs belong to group D (36.1%) followed by groups G and P (9.5% each) (Fig. 4.2; Table S4.2).

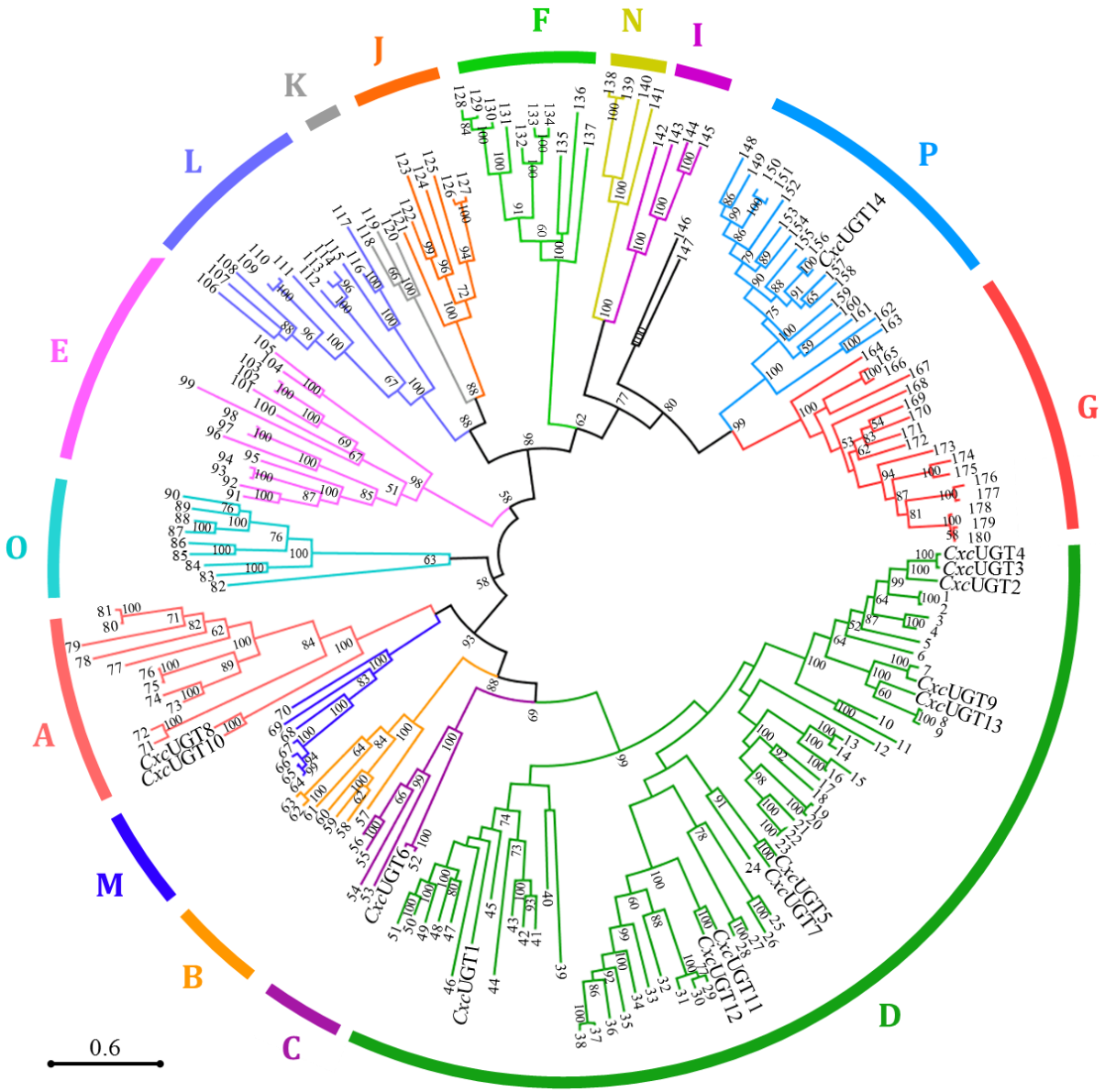


Figure 4.2: Phylogenetic analysis of GT1 UGTs from *C. x crocosmiiflora*. The maximum-likelihood tree was produced using the MEGA 7.0 program (bootstrap value set at 1,000) with sequences of 160 *C. x crocosmiiflora*, 15 *Arabidopsis thaliana*, 17 *Zea mays*, and 2 *Oryza sativa* GT1 UGT. Bootstrap values over 50% are indicated above the nodes. The black bar represents 0.6 amino acid substitutions per site. The GT1 UGTs identified through the Haystack analysis (section 2.3.5) as putatively involved in MbA biosynthesis are indicated on tree. The remaining sequences are numbered 1 – 180 and correspond to the legend found in Table S4.3).

4.3.2 Production of Potential Montbretin A Intermediates

One of the major challenges faced in elucidating genes involved in the MbA biosynthetic pathway is the lack of information on the order of the proposed glycosylation

reactions. With only myricetin and myricetin-3-O-rhamnoside commercially available, two approaches that could be used to establish the pathway are to (i) identify each step in sequential order or (ii) test candidate GT1 UGTs for activity towards a suite of potential MbA intermediates. Previous work by Williams *et al.* (2015) produced a series of MbA "substructures" to study the kinetic and structural roles of MbA moieties in inhibiting HPA (Williams *et al.*, 2015). Employing modified approaches of those used to produce these substructures, five potential MbA intermediates were obtained for use in *in vitro* reactions (Fig 4.3): (i) MbA without the terminal 2-O glucopyranosyl moiety (MbA-G′), (ii) MbA without the terminal 4-O rhamnopyranosyl moiety (MbA-R′), (iii) MbA without the disaccharide rhamnopyranosyl-xylopyranoyl moiety (MbA-XR′), (iv) MbA without the caffeic acid and terminal 4-O rhamnopyranosyl moieties (MbA-CR′), and (v) MbA without the caffeic acid and the disaccharide rhamnopyranosyl-xylopyranoyl moieties (MbA-CXR′).

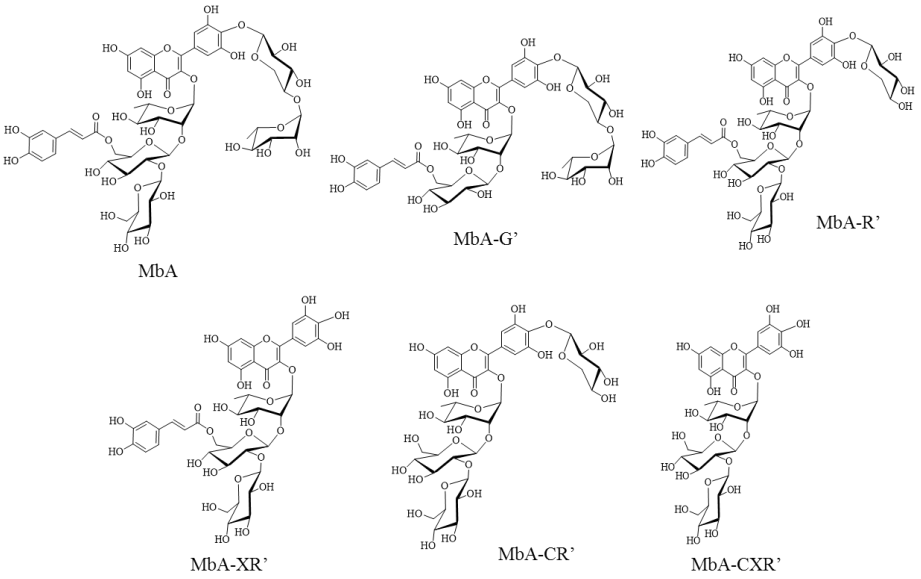


Figure 4.3: Hypothetical montbretin A intermediates produced by enzymatic and chemical degradation of MbA.

MbA-G′ was produced by incubating MbA with *Agrobacterium sp.* β-glucosidase overnight to remove the terminal glucoside moiety of MbA's 3-O-trisaccharide. LC-MS analysis of the reaction product showed a single new peak with a m/z of 1065.5 [M-H](superscript -). Analysis of the purified reaction product showed a retention time and mass spectra that corresponded to a sample of MbA-G′ previously prepared and nuclear magnetic resonance spectroscopy (NMR) verified (Williams *et al.*, 2015) (Fig 4.4).

MbA-R′ was produced by incubating MbA with *Penicillium decumbens* naringinase overnight to remove the terminal rhamnose moiety from the 4′-O-disaccharide. LC-MS analysis of the reaction product showed two new peaks, a major peak with a m/z of 1081.5 [M-H](superscript -) and minor one with m/z of 1065.5 [M-H](superscript -). Analysis of the minor reaction product showed a retention time and mass spectra that corresponded to MbA-G′, suggesting the β-D-glucosidase activity of the naringinase was not completely inactivated by incubation at 60^{o}C and still contained minor glycosylation activity. Analysis of the purified major reaction product after showed a retention time and mass spectra that corresponded to a sample of MbA-R′ previously prepared and NMR verified (Williams *et al.*, 2015) (Fig 4.4).

MbA-XR′ was produced by incubating MbA-R′ with *Bacillus halodurans* β-xylosidase overnight to remove the 4-O-xyloside moiety. LC-MS analysis of the reaction product showed a single new peak with a m/z of 949.3 [M-H](superscript -). While an NMR verified standard was not available, the observed m/z, previously characterized protocol (Williams *et al.*, 2015), and depletion of the MbA-R′ starting material strongly supports the purified product was MbA-XR′ (Fig 4.4).

MbA-CR′ was produced by base-catalyzed removal of the caffeic acid moiety followed by reaction with *P. decumbens* naringinase. MbA-C was produced by incubation with alkaline sodium methoxide for 5 hours. LC-MS analysis of the reaction product showed a single new peak with a m/z of 1065.5 [M-H](superscript -). MbA-C was then incubated with *P. decumbens* naringinase overnight to remove the terminal rhamnose moiety from the 4′-O-disaccharide. LC-MS analysis of the reaction product showed two new peaks, a major peak with a m/z of 949.3 [M-H](superscript -) and minor one with m/z of 919.3 [M-H](superscript -), corresponding to MbA-C without rhamnose or glucose moieties, respectively. While NMR verified standards were not available, the observed m/z, previously characterized protocol (Williams *et al.*, 2015), and depletion of the MbA-C starting material support the purified m/z 949.3 [M-H](superscript -) product as MbA-CR′ (Fig 4.4).

MbA-CXR′ was produced by incubating MbA-CR′ with *B. halodurans* β-xylosidase overnight to remove the 4-O-xyloside moiety. LC-MS analysis of the reaction product showed a single new peak with a m/z of 787.3 [M-H](superscript -). While an NMR verified standard was not available, the observed m/z, previously characterized protocol (Williams *et al.*, 2015),

and depletion of the MbA-CR′ starting material strongly supports the purified product was MbA-XR′ (Fig 4.4).

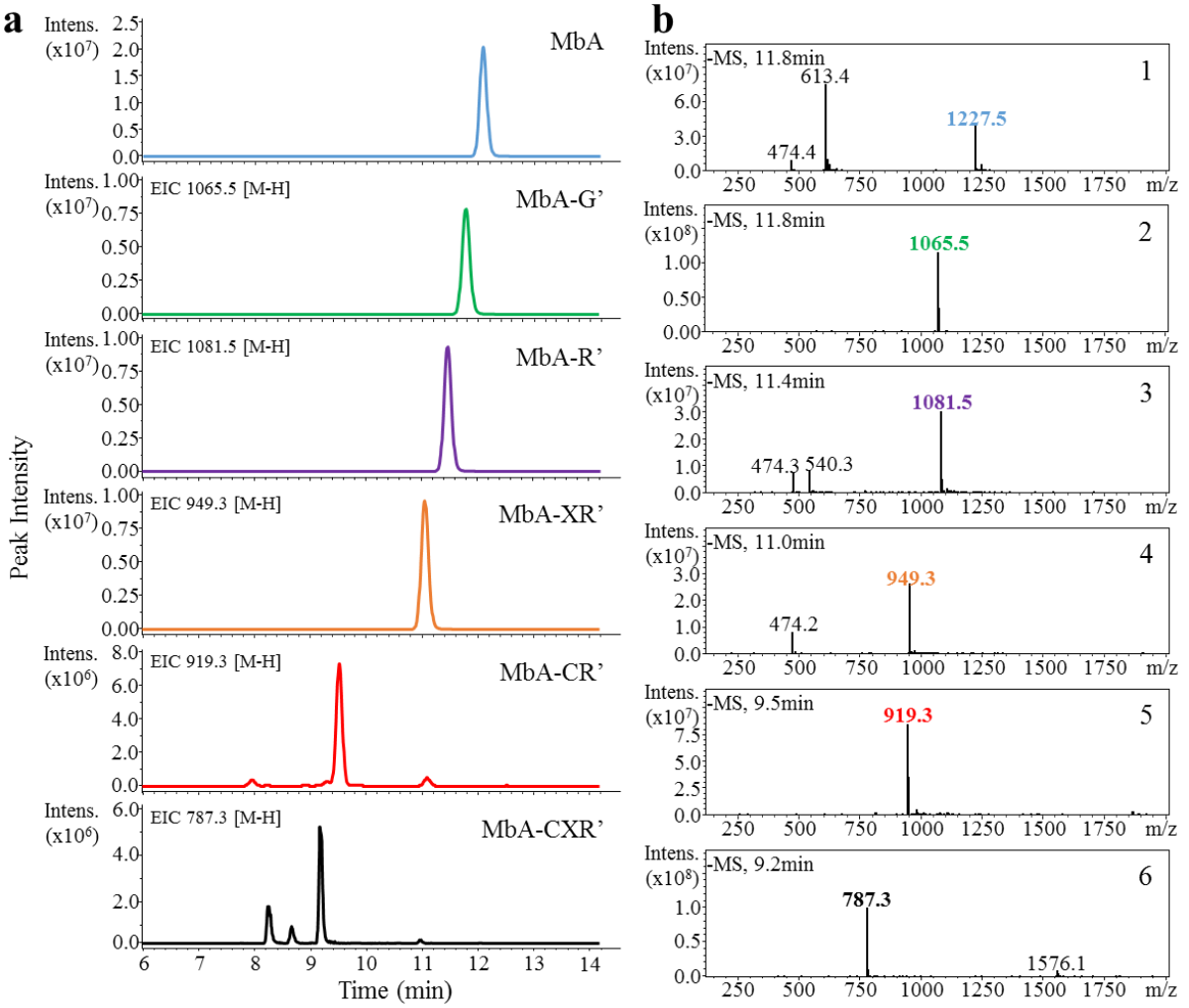


Figure 4.4: Representative regions of extracted ion LC-MS chromatograph and corresponding mass spectra of MbA and hypothetical MbA intermediates produced by degradation of MbA. (a) Extracted ion chromatographs of purified metabolites derived from enzymatic or cleavage reactions of MbA. Blue, green, purple, orange, red, and black traces represent extracted ion chromatographs for m/z of 1227.5 [M-H](superscript -), 1065.5 [M-H](superscript -), 1081.5 [M-H](superscript -), 949.3 [M-H](superscript -), 919.3 [M-H](superscript -), 787.3 [M-H](superscript -), respectively. Based on previously reported NMR analysis, peaks in the blue, green, and purple traces correspond to MbA, MbA-G′, and MbA-R′, respectively (Tarling *et al.*, 2008; Williams *et al.*, 2015). (b) Mass spectra of MbA and hypothetical MbA intermediates produced by MbA degradation. The spectra presented in "1", "2", "3", "4", "5", and "6" are the background subtracted mass spectra for chromatographic peaks corresponding to MbA, MbA-G′, MbA-R′, MbA-XR′, MbA-CR′, and MbA-CXR′ (theoretical molecular weight of each is 1229.05, 1066.92, 1082.92, 950.81, 920.78, and 788.66), respectively.

4.3.3 Testing of Candidate *Cxc*UGTs for Functions in the MbA Biosynthetic Pathway

Building on the characterization of MbA accumulation in *C. x crocosmiiflora* (section 2.3.1), a guilt-by-association analysis was employed to identify transcripts whose relative expression pattern correlated with MbA accumulation across five different organs. Of the 1,967 unigenes whose expression correlated with MbA accumulation, 14 were identified as GT1 UGTs based on the presence of a PSPG motif (section 2.3.5). Phylogenetic analysis showed ten of these putative *Cxc*UGTs were found in group D, two in group A, one in group C, and one in group P (Fig. 4.2). The corresponding cDNAs were amplified by PCR and cloned into the pASK-IBA37+ vector with N-terminal His_6-tag. The resulting clones were designated *Cxc*UGT1 – *Cxc*UGT14. Recombinant proteins were expressed in *E. coli* and Ni^{2+} affinity purified. This resulted in the isolation of His_6-tagged proteins *Cxc*UGT1 – *Cxc*UGT14, which appeared to be soluble and which matched in the western blot with the predicted molar masses of 51.9 kDa – 58.0 kDa (Fig. S4.2).

Using the commercially available myricetin and myricetin-3-O-rhamnoside, as well as the five produced hypothetical MbA intermediates, I tested *Cxc*UGT1 – *Cxc*UGT14 for activity in nine potential MbA biosynthetic steps (Fig. 4.1; reactions 1, 3, 6, 8, 23, 24, 25, 29, and 30). Enzyme assays (n = 3 replicates) using protein extracts of *Cxc*UGTs from *E. coli* were performed against relevant controls, and authentic standards were used where available, to identify the *Cxc*UGT reaction products. For those reactions requiring UDP-Rha as a sugar donor, NSE-UGT coupled reactions were performed by adding 1 mM NAD^+, 1 mM NADPH, 1 mM UDP-Glc, and 5 μM *Cxc*RHM1 to the reaction as described in section 3.3.7. For those reactions requiring UDP-Xyl as a sugar donor, NSE-UGT coupled reactions were performed by adding 1 mM NAD^+, 1 mM UDP-GlcA, and 1 μM *Cxc*UXS4 to the reaction as described in section 3.3.7.

LC-MS analysis of the enzyme assays for *Cxc*UGT1 – *Cxc*UGT14 using myricetin as an acceptor and UDP-rhamnose as a sugar donor showed the product profiles of *Cxc*UGT1, *Cxc*UGT4, *Cxc*UGT7, and *Cxc*UGT8 possessed peaks with a m/z of a myricetin rhamnoside, 463.0 [M-H](superscript -), that were absent in the negative control (Fig 4.5, Fig. S4.3). Comparison of the retention time and mass spectra to the myricetin-3-O-rhamnoside standard suggest the four GT1 UGTs did not rhamnosylate the 3-hydroxyl position, but rather the 5, 7, 3′, 4′, or 5′ positions.

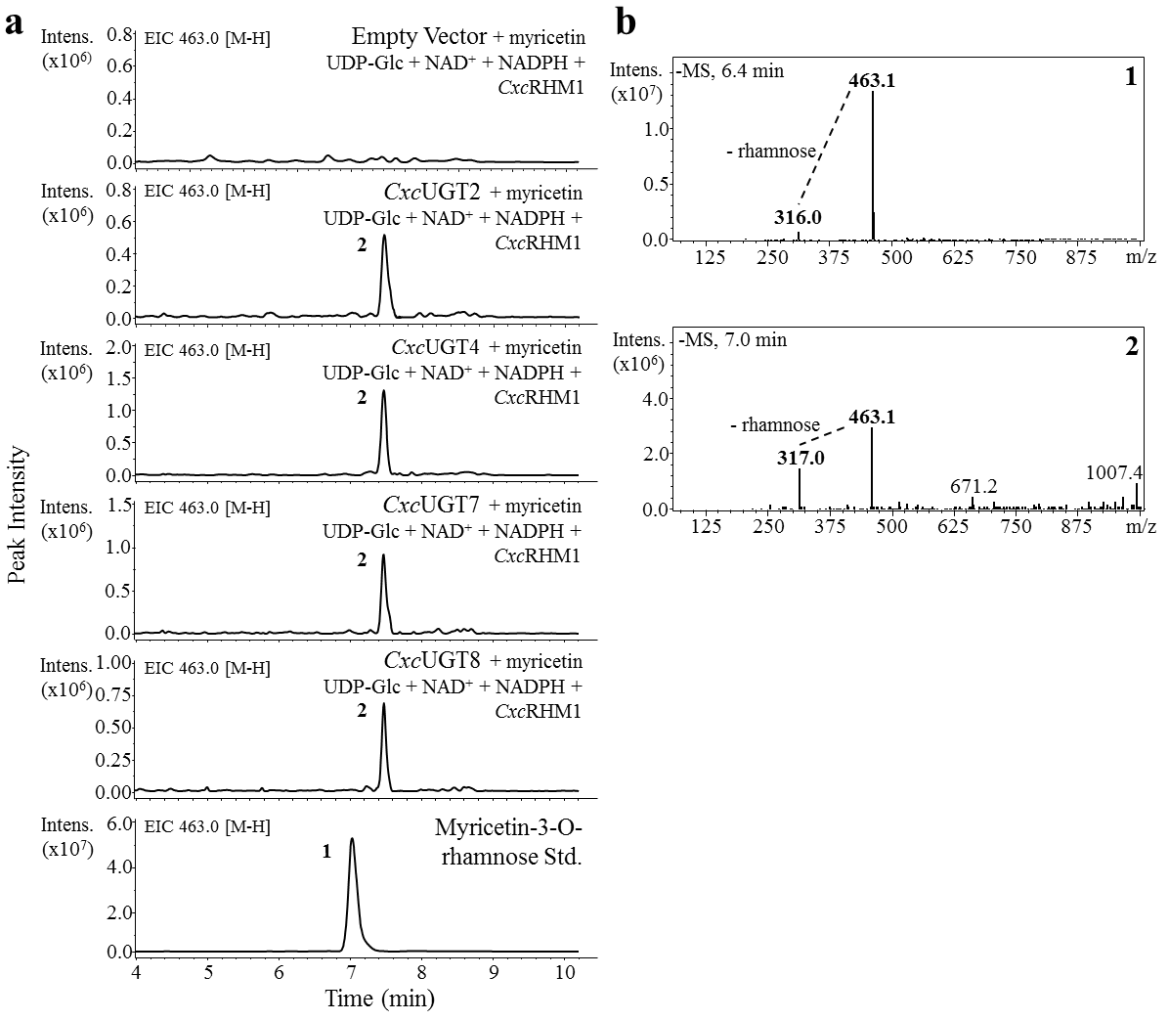


Figure 4.5: Select regions of extracted ion LC-MS chromatograph and corresponding mass spectra for *C. x crocosmiiflora* GT1 UGT enzyme assays. (a) Protein derived from *E. coli* expressing a control vector and *Cxc*UGT1 – *Cxc*UGT14 were incubated overnight with 1 mM UDP-glucose, 1 mM NAD^+, 1 mM NADPH, 5 µM purified *Cxc*RHM1, and 100 µM myricetin and assessed for their ability to form myricetin rhamnoside (theoretical molecular weight of 464.38). Trace shown for each sample is the extracted ion chromatograph for m/z of 463.0 [M-H](superscript -). (b) Mass spectral analysis of myricetin-3-O-rhamnose standard and myricetin rhamnoside enzyme assay products. Numbers next to chromatograph peaks correspond to mass spectra with associated number in the top right-hand corner.

LC-MS analysis of the enzyme assays for *Cxc*UGT1 – *Cxc*UGT14 using myricetin as an acceptor and UDP-xylose as a sugar donor showed the product profiles of *Cxc*UGT2, *Cxc*UGT3, *Cxc*UGT4, *Cxc*UGT5, *Cxc*UGT7, *Cxc*UGT8, *Cxc*UGT11, and *Cxc*UGT12 possessed peaks with a m/z of a myricetin xyloside, 449.0 [M-H](superscript -), that was absent in the negative control (Fig 4.6, Fig. S4.4).

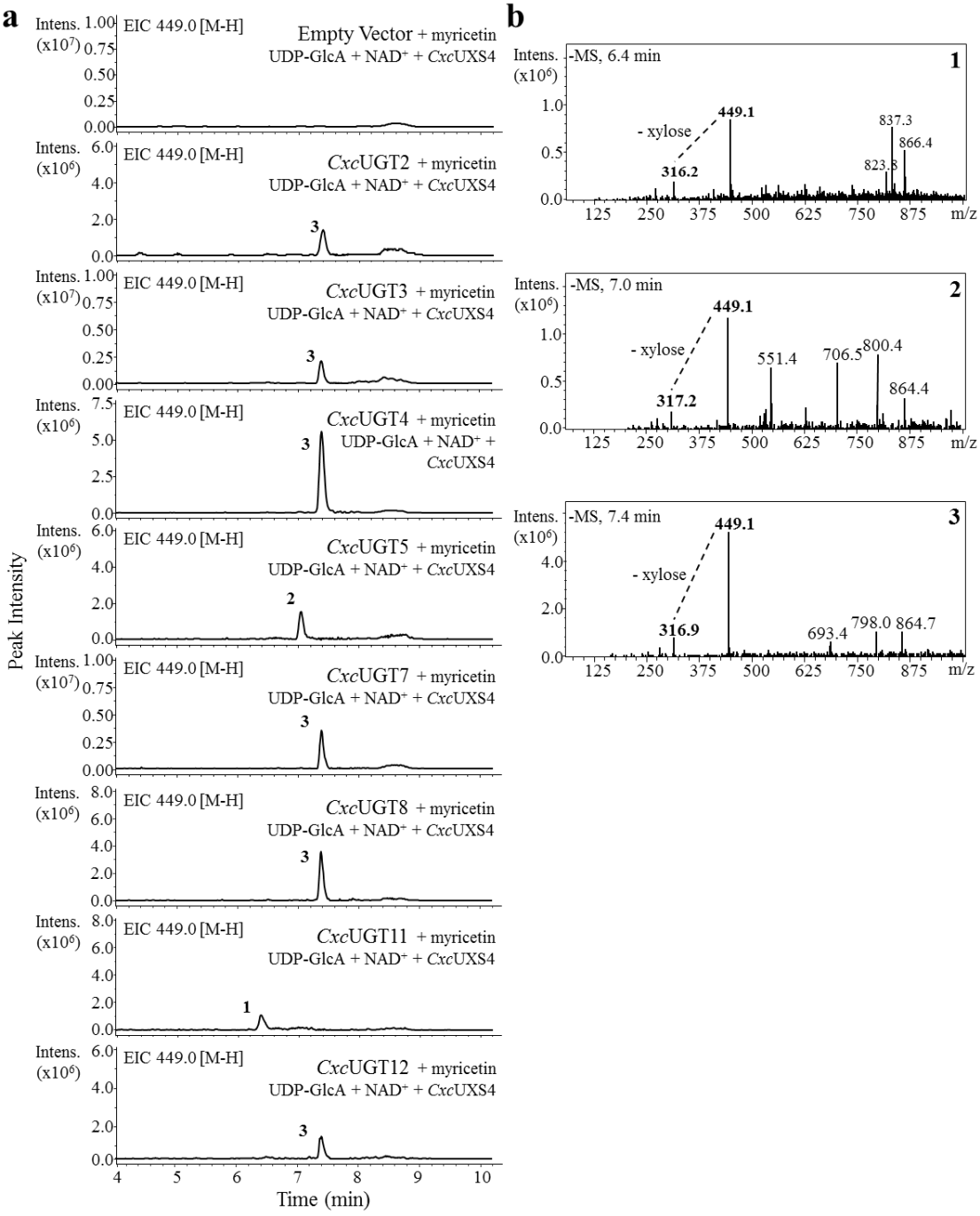


Figure 4.6: Select regions of extracted ion LC-MS chromatograph and corresponding mass spectra for *C. x crocosmiiflora* GT1 UGT enzyme assays. (a) Protein derived from *E. coli* expressing a control vector and *Cxc*UGT1 – *Cxc*UGT14 were incubated overnight with 1 mM UDP-glucuronic acid, 1 mM NAD^+, 1 μM purified *Cxc*UXS4, and 100 μM myricetin and assessed for their ability to form myricetin xyloside (theoretical molecular weight of 450.35). Trace shown for each sample is the extracted ion chromatograph for m/z of 449.0 [M-H](superscript -). (b) Mass spectral analysis of potential myricetin xyloside enzyme assay products. Numbers next to chromatograph peaks correspond to mass spectra with associated number in the top right-hand corner.

LC-MS analysis of the enzyme assays for *Cxc*UGT1 – *Cxc*UGT14 using myricetin-3-O-rhamnoside as an acceptor and UDP-Glc as a sugar donor showed the product profiles of *Cxc*UGT3, *Cxc*UGT4, and *Cxc*UGT6 possessed peaks with the expected m/z of a myricetin-3-rhamnose glucoside, 625.0 [M-H](superscript -), that were absent in the negative control (Fig 4.7, Fig. S4.5).

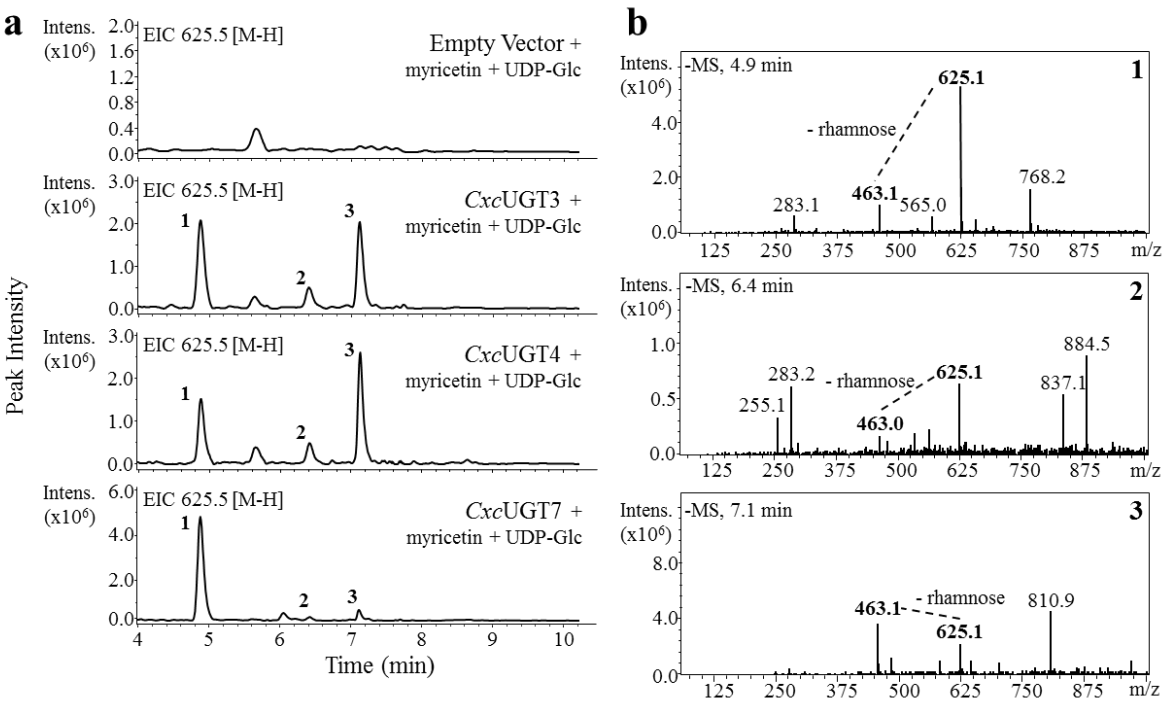


Figure 4.7: Select regions of extracted ion LC-MS chromatograph and corresponding mass spectra for *C. x crocosmiiflora* GT1 UGT enzyme assays. (a) Protein derived from *E. coli* expressing a control vector and *Cxc*UGT1 – *Cxc*UGT14 were incubated overnight with 1 mM UDP-glucose and 100 µM myricetin-3-O-rhamnose and assessed for their ability to form myricetin-3-O-rhamnose glucoside (theoretical molecular weight of 626.49). Trace shown for each sample is the extracted ion chromatograph for m/z of 625.0 [M-H](superscript -). (b) Mass spectral analysis of potential myricetin-xyloside enzyme assay products. Numbers next to chromatograph peaks correspond to mass spectra with associated number in the top right-hand corner.

LC-MS analysis of the enzyme assays for *Cxc*UGT1 – *Cxc*UGT14 using myricetin-3-O-rhamnoside as an acceptor and UDP-Xyl as a sugar donor showed the product profiles of *Cxc*UGT5 and *Cxc*UGT12 possessed peaks with the expected m/z of a myricetin-3-rhamnose xyloside, 595.0 [M-H](superscript -), that were absent in the negative control (Fig 4.8, Fig. S4.6).

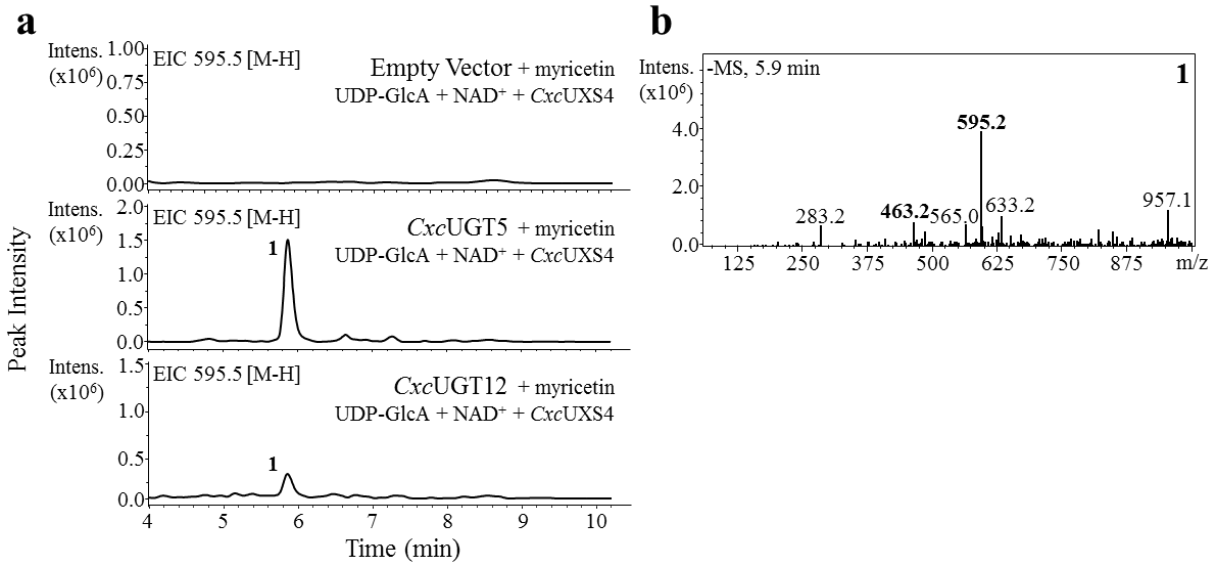


Figure 4.8: Select regions of extracted ion LC-MS chromatograph and corresponding mass spectra for *C. x crocosmiiflora* GT1 UGT enzyme assays. (a) Protein derived from *E. coli* expressing a control vector and *Cxc*UGT1 – *Cxc*UGT14 were incubated overnight with 1 mM UDP-glucuronic acid, 1 mM NAD^+, 1 μM purified *Cxc*UXS4, and 100 μM myricetin-3-O-rhamnose and assessed for their ability to produce a myricetin-3-O-rhamnose xyloside (theoretical molecular weight of 596.49). Traces shown for each sample is the extracted ion chromatograph for m/z of 595.5 [M-H](superscript -). (b) Mass spectral analysis of potential myricetin-3-O-rhamnose xyloside enzyme assay products. Numbers next to chromatograph peaks correspond to mass spectra with associated number in the top right-hand corner.

In assessing the activity of *Cxc*UGT1 – *Cxc*UGT14 towards the five hypothetical MbA intermediates, LC-MS analysis showed that MbA-CXR′ appeared to have degraded, and the peak in the extracted ion chromatograph of 787.5 [M-H](superscript -) was no longer detected. The other four hypothetical intermediates produced by breakdown of MbA appeared to be intact. However, LC-MS analysis of assays with *Cxc*UGT1 – *Cxc*UGT14 using MbA-R′, MbA-G′, MbA-XR′, and MbA-CR′ as acceptors and their corresponding putative sugar donors, UDP-Rha, UDP-Glc, UDP-Xyl, and UDP-Rha, respectively, did not result in any observed activity (Fig. S4.8 – S4.11).

To further assess the potential role of the 14 candidate *Cxc*UGTs in the biosynthesis of MbA, their ability to catalyze eight different combinations of acceptor and sugar donors was analyzed. While some activity was observed when myricetin and myricetin-3-O-rhamnoside were used as acceptors, the conversion detected in overnight assays was very low. Although some GT1 UGTs have been identified through *in vitro* assays as highly specific (Funaki *et al.*, 2015), most have been shown to be able to use multiple acceptors from closely related classes

of metabolites, or multiple UDP-sugar donors, at reduced levels of activity (Kovinich *et al.*, 2010; Masada *et al.*, 2009; Song *et al.*, 2015). Accordingly, the low activity observed in these assays suggest the primary catalytic activity of the 14 candidate GT1 UGTs identified through the guilt-by-association are not the tested potential MbA biosynthetic reactions.

4.4 DISCUSSION

Using transcriptome mining and biochemical approaches, I tested members of the *C. x crocosmiiflora* GT1 UGT family for possible activity in the biosynthesis of MbA. Plant species whose GT1 UGTs have been previously characterized show large families containing up to several hundred GT1 UGTs (Barvkar *et al.*, 2012; Caputi *et al.*, 2012; Huang *et al.*, 2015; Huang *et al.*, 2009; Jaillon *et al.*, 2007; Khorolragchaa *et al.*, 2014; Paterson *et al.*, 2009; Tanaka *et al.*, 2008; Tuskan *et al.*, 2006; Velasco, *et al.*, 2010). The phylogenetic clustering of GT1 UGTs into phylogenetic groups appears to be conserved across the vascular plants with previous studies identifying 17 phylogenetic groups, groups A – Q (Caputi *et al.*, 2012; Li *et al.*, 2001; Li *et al.*, 2014b). Analysis of the *C. x crocosmiiflora* transcriptome shows a similar pattern of large-scale gene expansion with at least 257 putative GT1 UGTs identified in the draft transcriptome and 160 sequences larger than 300 amino acids in length clustering into 15 phylogenetic groups (Fig. 4.2; Table S4.2). Interestingly, when compared to phylogenetic distribution of GT1 UGTs in other species, the *C. x* cr*ocosmiiflora* GT1 UGT family shows an asymmmetric distribution with fewer members in groups E, H, I, K, and L and expansions in groups D, F, N, and P. Surprisingly, the transcriptome did not reveal any *Cxc*UGTs of group H, which is different from other plant species. Characterized group H UGTs of other species have been identified as cytokinin glycosyltransferases (Hou *et al.*, 2004; Kudo *et al.*, 2012; Wang *et al.*, 2011). The apparent absence of *Cxc*UGTs in this group could be a result of the plant material used for RNA isolation perhaps not including developmental stages of active cytokinin metabolism (Sakakibara, 2006), and it may also be possible that group H *Cxc*UGTs may be present among the shorter incomplete transcripts that were not included in the phylogeny. Conversely, group D appears to have undergone a substantial expansion with approximately a third of *Cxc*UGTs clustering in this group. The most recently identified phylogenetic group, group Q, was identified in an analysis of Z. *mays* GT1 UGTs (Li *et al.*, 2014b). Analysis of the *C. x crocosmiiflora* and other monocot GT1 UGT families failed to

identify any members of this group. Interestingly, an expanded phylogenetic analysis of *Cxc*UGTs with the seven previously characterized group Q *Zm*UGTs showed all seven sequences clustered in group D (Fig. S4.12). These results contrast with those previously reported (Li *et al.*, 2014b). Accordingly, based on the current sequence clustering, my results support the identification of 16 phylogenetic groups, A – P, for the GT1 UGT family.

While phylogenetic analysis has previously been proven to be useful for predicting the class of substrate acceptor or UDP-sugar donor used by a given GT1 UGT (Bowles *et al.*, 2005; Cartwright *et al.*, 2008; Lim *et al.*, 2003), more detailed predictions based on sequence phylogenies can be challenging (Hansen *et al.*, 2003; Modolo *et al.*, 2007). The observed difficulty in predicting functions based on sequences alone is likely due to the species-specific variation in the evolution of the GT1 UGT family. Evidence of this species-specific variation can be seen in three observations about plant GT1 UGT families. First, examination of available characterized GT1 UGT families shows expansion of individual GT1 UGT groups appears to have occurred at different rates in different species in a manner that is not correlated to genome size (Table S4.2) (Hellsten *et al.*, 2013; Huang *et al.*, 2009; Jaillon *et al.*, 2007; Lamesch *et al.*, 2012; Li *et al.*, 2014a; Paterson *et al.*, 2009; Schmutz *et al.*, 2010; Schnable *et al.*, 2009; Tanaka *et al.*, 2008; Tuskan *et al.*, 2006; Velasco *et al.*, 2010; Wang *et al.*, 2012). Second, within plant genomes, GT1 UGT co-localization within chromosomes is very common (Caputi *et al.*, 2012). Third, within individual phylogenetic groups, it was found that GT1 UGTs cluster by taxonomy rather than by function (Caputi *et al.*, 2012; Hansen *et al.*, 2003; Modolo *et al.*, 2007; Yonekura-Sakakibara and Hanada, 2011). The combination of these three observations suggest that the expansion of GT1 UGT families likely occurred through gene duplication and subsequent neofunctionalization, resulting in GT1 UGTs utilizing different metabolites as acceptors. Thus, this lineage-specific expansion of GT1 UGTs and subsequent acquisition of new functions makes identification of GT1 UGT function by sequence alone challenging.

Instead of solely relying on sequence relatedness as an approach to identify *Cxc*UGTs as candidates for involvement in MbA biosynthesis, the work presented in this chapter employed a guilt-by-association approach. A critical component of this approach is the identification of different plant tissues, growth stages, or environmental conditions in which active biosynthesis, and thus gene expression, of the target metabolite differ. At the start of

the work presented in this chapter, the lack of data available on MbA biosynthesis in *Crocosmia spp.* presented a challenge for ensuring this candidate gene selection approach would be successful. Based on other work in this thesis that developed resources for *C. x crocosmiiflora*, the guilt-by-association analysis was built on two main hypotheses: (i) that the expression of MbA biosynthetic genes correlates with MbA accumulation and (ii) that MbA biosynthesis genes are actively expressed in the plants used for RNA-sequencing.

While low levels of *in vitro* activity were observed with some of the 14 candidate GT1 UGTs, the low level of activity compared to what would be expected when GT1 UGTs are using their primary acceptor substrate suggests that these *Cxc*UGTs are not involved in any of the tested putative MbA biosynthetic reactions. There are at least three possible explanations for this negative result. First, except for myricetin, none of the other tested substrates are known to be the *in planta* true intermediates in the MbA biosynthetic pathway. Second, the use of a Hig-tag did not help improve the functional expression of the UGTs, resulting in lower levels of activity. Third, and the more likely explanation, at least one of the two hypotheses employed in the guilt-by-association analysis was incorrect. Several explanations for the breakdown of these hypotheses could exist: (i) observed levels of MbA accumulation was the result of prior biosynthesis and subsequent storage, and thus MbA was not being actively produced at the time of tissue sampling; (ii) MbA might be produced outside of the corm and transported into the corm; (iii) MbA might only be produced in specific tissues or cells within the corm and these would have to be isolated to enrich for the relevant transcripts; or (iv) patterns of expression of MbA biosynthetic genes may not correlate with patterns of MbA accumulation. While the guilt-by-association approach can still hold true in further efforts to elucidate the MbA biosynthetic genes, our understanding of and resources for *C. x crocosmiiflora* must be expanded first to better ensure success of this approach. Future work should focus on identifying specific growth stages, environmental conditions, or cellular structures that affect MbA accumulation levels for use as a more accurate model of MbA biosynthetic gene expression. Such an approach is currently being employed by others in the Bohlmann lab, who will continue this research. Future work could also be supported by additional approaches to identify the biosynthetic GT1 UGTs. For example, it may be possible that both GT1 UGTs as well as core genes involved in MbA biosynthesis are colocalized in the genome, as has been seen for some specialized metabolic pathways (Kliebenstein and

Osbourn, 2012; Nützmann and Osbourn, 2014). If this is the case, GT1 UGTs involved in MbA biosynthesis could be identified through sequencing the genomic regions around any GT1 UGT or other core genes characterized as part of MbA biosynthetic pathway.

An alternative approach to identifying a GT1 UGT function would be a reverse genetics approach such as producing a knockout or through gene silencing. However, with a non-model system such as *C. x crocosmiiflora*, such an approach is currently unavailable. While testing a broad range of acceptors and sugar donors for higher levels of activity could be employed to identify activity of a GT1 UGT, their common broad substrate specificity can hinder the identification of their *in vivo* substrate (Achnine *et al.*, 2005). This broad specificity of recombinant GT1 UGTs *in vitro* may not provide insight into the *in planta* activity as substrate availability will also be relevant (Song *et al.*, 2015). A potentially more efficient approach may be to use a physiological aglycone library enriched in *C. x crocosmiiflora*'s naturally occurring aglycones, produced through enzymatic hydrolysis (Bonisch *et al.*, 2014), as a pool of acceptor substrates. However, as the structure of MbA shows, even this approach faces difficulties such as identifying the substrates for GT1 UGTs responsible for secondary and tertiary glycosylations. As nine of the *Cxc*UGTs presented here showed activity towards at least one flavonoid, future work could focus on screening these recombinant GT1 UGT's activity against *C. x crocosmiiflora*-specific aglycone libraries in combination with LC-MS and NMR analysis to characterize the activities of these GT1 UGT's in *C. x crocosmiiflora*.

4.5 CONCLUSION

The work presented in this chapter identified a large set of members of the *C. x crocosmiiflora* GT1 UGT family and tested the activity of 14 candidate GT1 UGTs that were selected through a guilt-by-association analysis for possible roles in MbA biosynthesis. Phylogenetic analysis of the *C. x crocosmiiflora* GT1 UGTs provided insight into the evolution of this family as it is the first member of the order *Asparagales* for which a comprehensive set of GT1 UGT sequences has been reported. Distribution of *Cxc*UGT into gene family phylogenetic groups showed a large expansion in group D and an apparent absence of group H members. While minor levels of activity were observed with myricetin and myricetin-3-O-rhamnoside, these *in vitro* assays suggested that none of the 14 candidate GT1 UGTs are involved in MbA biosynthesis. As these results may be due to errors in the underlying

hypotheses of the guilt-by-association analysis used to identify these candidates, future work towards identifying the MbA biosynthetic genes should focus on identifying specific conditions that affect MbA accumulation levels for use as a model of MbA biosynthetic gene expression.

CHAPTER 5: PLASTICITY AND EVOLUTION OF (+)-3-CARENE SYNTHASE AND (−)-SABINENE SYNTHASE FUNCTIONS OF A SITKA SPRUCE MONOTERPENE SYNTHASE GENE FAMILY ASSOCIATED WITH WEEVIL RESISTANCE

The monoterpene (+)-3-carene is associated with Sitka spruce resistance against the white pine weevil, a major North American forest insect pest of pine and spruce. High and low levels of (+)-3-carene in, respectively, resistant and susceptible Sitka spruce genotypes are due to variation of (+)-3-carene synthase gene copy number, transcript and protein expression levels, enzyme product profiles, and enzyme catalytic efficiency. A family of multiproduct (+)-3-carene synthase-like genes of Sitka spruce includes the three (+)-3-carene synthases, *PsTPS-3car1*, *PsTPS-3car2*, *PsTPS-3car3*, and the (–)-sabinene synthase *PsTPS-sab*. Of these, *PsTPS-3car2* is responsible for the relatively higher levels of (+)-3-carene in weevil-resistant trees. Here, the features of the *Ps*TPS-3car1, *Ps*TPS-3car2, *Ps*TPS-3car3, and *Ps*TPS-sab proteins that determine different product profiles were identified by a series of domain swap and site-directed mutations, supported by structural comparisons. This work identified the amino acid in position 596 as critical for product profiles dominated by (+)-3-carene in *Ps*TPS-3car1, *Ps*TPS-3car2, and *Ps*TPS-3car3, or (–)-sabinene in *Ps*TPS-sab. A leucine in this position promotes the formation of (+)-3-carene, whereas phenylalanine promotes (–)-sabinene. Homology modeling predicts that position 596 directs product profiles through differential stabilization of the reaction intermediate. Kinetic analysis revealed position 596 also plays a role in catalytic efficiency. Mutations of position 596 with different side chain properties resulted in a series of enzymes with different product profiles, further highlighting the inherent plasticity and potential for evolution of alternative product profiles of these monoterpene synthases of conifer defense against pests.

5.1 INTRODUCTION

White pine weevil (*Pissodes strobi*) is one of the most devastating insect pests of spruce (*Picea spp*.) and pine (*Pinus spp.*). Sitka spruce (*Picea sitchensis*), a conifer species in which most genotypes are highly susceptible to weevils (King *et al.*, 2004), is native to the temperate rainforest ecosystem of the North American Pacific coast and is also an economically valuable forest tree in Europe. Susceptibility to weevils caused the nearly complete halt of commercial

Sitka spruce reforestation in the Pacific Northwest. However, successful field trials identified a few highly resistant Sitka spruce genotypes; most notably genotype H898, which has become a focus for research and breeding of conifer resistance to stem boring insects (King *et al.*, 2004).

One of the major defenses of conifers against insects is the chemically complex oleoresin, which includes dozens of different monoterpenes and diterpene resin acids (Keeling and Bohlmann, 2006a; Keeling and Bohlmann, 2006b; Phillips and Croteau, 1999; Zulak *et al.*, 2009). Previous work (Robert *et al.*, 2010) explored the monoterpene and diterpene resin acid profiles of Sitka spruce from different geographic regions of the natural distribution where trees displayed strong, intermediate, or weak resistance. Resistance was positively associated with higher levels of the bicyclic monoterpene (+)-3-carene (Robert *et al.*, 2010). Subsequently, Hall *et al.* (Hall *et al.*, 2011) used a combination of genomic, target specific proteomic, and biochemical approaches to study the basis of variation of (+)-3-carene levels in two contrasting genotypes of Sitka spruce, resistant genotype H898 trees with relatively high levels of (+)-3-carene and susceptible genotype Q903 trees with trace levels of (+)-3-carene. This work identified a small family of (+)-3-carene synthase-like genes in Sitka spruce that contains the three (+)-3-carene synthases *Ps*TPS-3car1, *Ps*TPS-3car2, *Ps*TPS-3car3, and the (–)-sabinene synthase *Ps*TPS-sab. Genotype-specific variations of gene copy number, transcript and protein expression, and catalytic efficiencies of members of this family are responsible for the difference in (+)-3-carene levels (Hall *et al.*, 2011). Specifically, the genomic presence, transcript and protein expression, and enzyme activity of *Ps*TPS-3car2 accounted for much of the high levels of (+)-3-carene in the resistant genotype.

Members of the Sitka spruce (+)-3-carene synthase-like family showed between 82.5 and 95.7% pairwise amino acid sequence identity. These four enzymes are multiproduct enzymes with the same overall product profile of monoterpenes, however, with different relative amounts of individual compounds (Hall *et al.*, 2011). Most notably, *Ps*TPS-3car1, *Ps*TPS-3car2, and *Ps*TPS-3car3 have (+)-3-carene as the predominant product, whereas *Ps*TPS-sab forms (–)-sabinene as the predominant product. All four enzymes produce α-terpinolene as the second most abundant product plus a set of additional minor products (Hall *et al.*, 2011) (Table 5.1, Table S5.1). These similar traits and the particular presence of a (–)-sabinene synthase as a closely related enzyme with a group of (+)-3-carene synthases suggested

a pattern of divergent evolution in which *Ps*TPS-sab arose from a *Ps*TPS-3car ancestor through gene duplication and shift of function (Hall *et al.*, 2011).

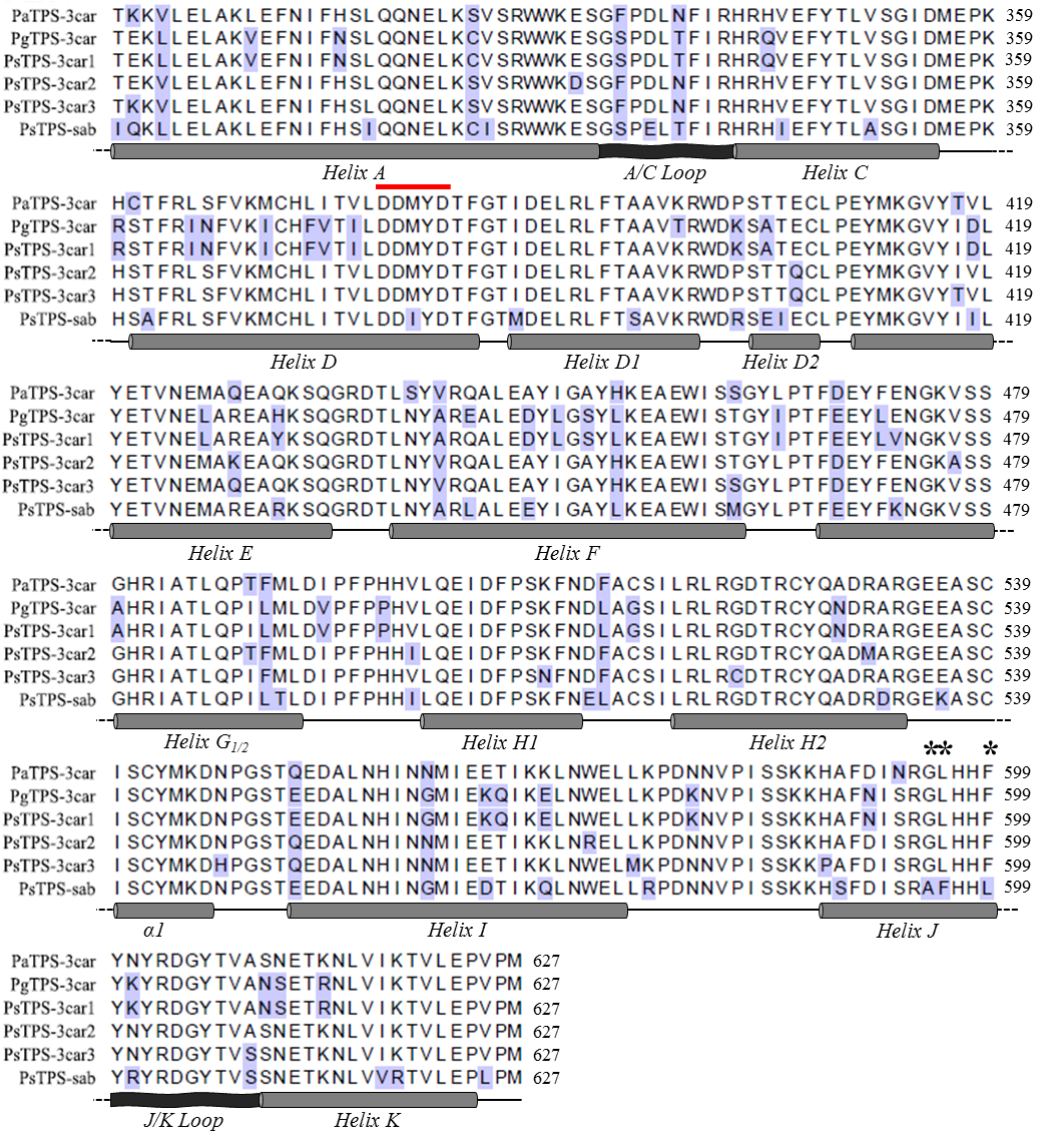


Figure 5.1: Amino acid sequence alignment of the C-terminal α-domain of spruce TPS-3car and TPS-sab enzymes of a family of (+)-3-carene synthase-like monoterpene synthases. The alignment includes protein sequences of (+)-3-carene synthases and (−)-sabinene synthase from Sitka spruce (*P. sitchensis*; *Ps*TPS-3car1, *Ps*TPS-3car2, *Ps*TPS-3car3 and *Ps*TPS-sab (Hall *et al.*, 2011)); as well as (+)-3-carene synthases from Norway spruce (*P. abies*; PaTPS-3car (Fäldt *et al.*, 2003)) and white spruce (*P. glauca*; PgTPS-3car (Hamberger *et al.*, 2009)). Amino acids with highlighted with blue background colour are those different from the consensus. A diagrammatic representation of the secondary structures of the C-terminal domain of the (+)-3-carene synthase-like enzymes is shown with cylinders representing α-helices and ribbons represent loops. The conserved DDxxD motif is identified by the red line. Positions 595, 596, and 599 in helix J are marked with asterisks.

Based on general knowledge of monoterpene synthases (Davis and Croteau, 2000), *Ps*TPS-3car and *Ps*TPS-sab enzymes are thought to employ divalent metal ion-dependent ionization/isomerization/cyclization reaction mechanisms (Fig. 5.2). Initial ionization of the substrate geranyl diphosphate allows the formation of linalyl diphosphate. Attack from the allylic double bond upon reionization of linalyl diphosphate results in the formation of the α-terpinyl cation, an important proposed carbocation intermediate for the formation of various cyclic monoterpenes found in the product profiles of *Ps*TPS-3car and *Ps*TPS-sab enzymes. This intermediate can undergo a series of hydride shifts and/or additional cyclizations until reactions are terminated by deprotonation or addition of a nucleophile. Previous work on angiosperm monoterpene synthases has shown that the product profiles of TPS can be highly affected by specific structural features. Hyatt and Croteau (2005) were able to show that Ser^{485} and Cys^{480} of an *Abies grandis* pinene synthase acted as terminal protein acceptors in the final deprotonation of the pinyl cation to for α- and β-pinene respectively (Hyatt and Croteau, 2005). Kampranis *et al.* (2007) identified Asn^{338} as critical for water capture and its subsequent deprotonation to produce 1,8-cineole instead of sabinene as the primary product (Kampranis *et al.*, 2007). Krause *et al.* (2013) were able to show that stereo-specificity of two *Thymus vulgaris* sabinene hydrate synthases was inter-converted by reciprocal substitution between a pair of isoleucine and asparagine residues (Krause *et al.*, 2013).

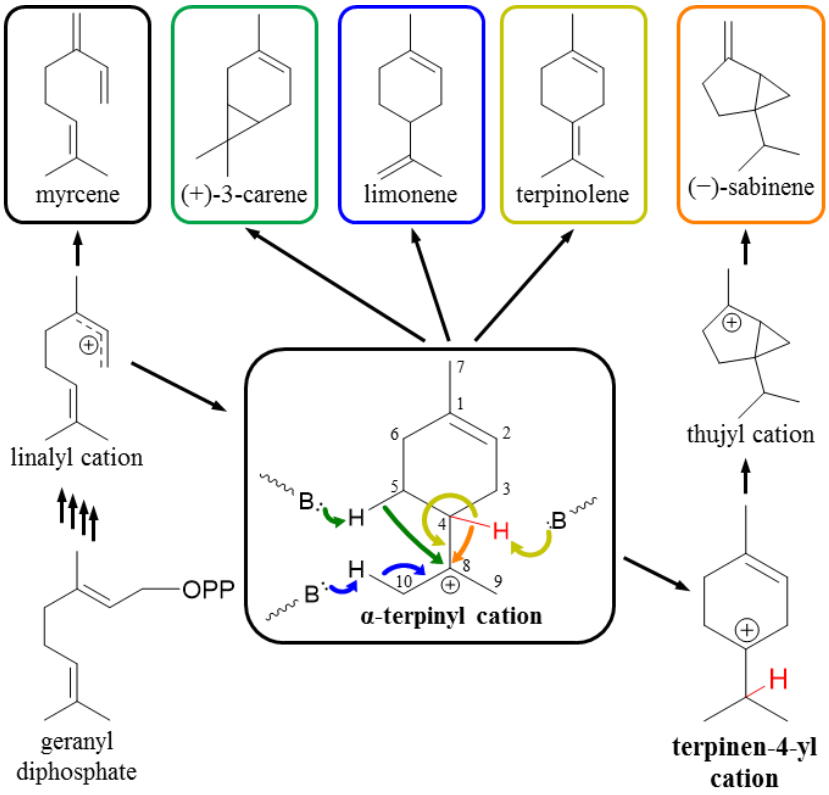


Figure 5.2: Proposed reaction mechanisms explaining monoterpenes of the product profiles of *Ps*TPS-3car and *Ps*TPS-sab enzymes and their variants. Cyclic monoterpene products, including the major products (+)-3-carene, (−)-sabinene, and α-terpinolene, are proposed to be derived from an α-terpinyl cation intermediate. Formation of (−)-sabinene is proposed to involve a terpinen-4-yl cation intermediate. Proposed hydride shifts, cyclizations, and termination reactions by proton loss are indicated with arrows colour coded with the corresponding products.

The high sequence similarity, yet different product profiles of the Sitka spruce *Ps*TPS-3car and *Ps*TPS-sab enzymes, and their different roles in contributing to insect resistance has made them attractive targets for investigating which particular structural features of these enzymes affect their functions. The goal of this chapter was to use domain-swapping and site-directed mutagenesis, guided by sequence comparisons and supported by structural homology modeling, to test which specific domains and amino acids direct *Ps*TPS-3car versus *Ps*TPS-sab product profile and how these domains and amino acids might interact with the reaction intermediates. Our results indicate changes of sequence and functions that may have occurred in the natural evolution of the (+)-3-carene synthase-like family of spruce defense.

5.2 EXPERIMENTAL

5.2.1 Domain Swapping and Site-Directed Mutagenesis

Mutagenesis of the cDNA clones *Ps*TPS-3car1, *Ps*TPS-3car2, *Ps*TPS-3car3, and *Ps*TPS-sab (Hall *et al.*, 2011) was performed using Phusion Hot Start II DNA Polymerase (Thermo Scientific) following the manufacturer's instructions with 25 ng of template DNA per reaction. Primers are listed in Table S5.2. All mutations were verified by Sanger sequencing before expression.

5.2.2 Protein Expression and Purification

Recombinant plasmids were transformed into *E. coli* C41 (www.overexpress.com) containing the pRARE 2 plasmid isolated from Rosetta 2 cells (Novagen) to negate codon bias. Individual colonies were inoculated into 50 mL of Terrific Broth containing kanamycin (50 mg/L) and chloramphenicol (50 mg/L) and cultured at 37°C and 180 rpm until OD_{600} = 1.0. Cultures were then cooled to 16°C, induced by addition of isopropyl β-D-1-thiogalactopyranoside (final concentration 0.1 mM), and grown for 16 h at 180 rpm before harvesting. Recombinant protein was extracted and nickel affinity-purified as previously described (Hall *et al.*, 2011; Keeling *et al.*, 2008). Protein concentrations were determined using a bicinchoninic acid (BCA) protein quantification assay kit (Thermo Fisher, www.thermofisher.com) employing a standard curve and SDS-PAGE with measurement of protein band intensity performed with the program ImageJ (http://rsbweb.nih.gov/ij/). Based on BCA protein assay quantification and SDS-page purity analysis of protein purified from the 50 mL *E. coli* cultures, it was estimated that between approximately 1 and 80 mg of purified *Ps*TPS protein could be isolated from a 1 L *E. coli* culture.

5.2.3 Enzyme Assays

Monoterpene synthase activities were assayed in triplicate as previously described with minor modifications (Hall *et al.*, 2011; Keeling *et al.*, 2008; O'Maille *et al.*, 2004). 500 µL reactions containing 25 mM HEPES, 100 mM KCl, 10 mM $MgCl_2$, 5 mM dithiothreitol, 10% glycerol 61.6 µM GPP (Echelon Biosciences Inc., http://www.echelon-inc.com), and affinity-purified protein extract were overlaid with 500 µL pentane containing 2.5 µM isobutylbenzene as an internal standard and incubated at 30°C for either 1 h (all enzymes derived from *Ps*TPS-

3car1, *Ps*TPS-3car2, and *Ps*TPS-sab) or 4 h (all enzymes derived from *Ps*TPS-3car3). Reaction products were extracted with pentane by vortexing for 30 seconds followed by phase separation by centrifugation at 1000 g for 30 min at 4°C.

To determine enzyme kinetic parameters, assays were performed with nine different concentrations of GPP ranging from 1 μM to 60 μM. *Ps*TPS-3car3 wild-type (WT) was assayed for 20 min at 30°C; all other enzymes were assayed for 10 min at 30°C. Enzyme concentrations in each assay were 19.9 – 26.9 pM for *Ps*TPS-3car2 (WT), 12.9 – 19.4 pM for variant 24, 4.6 – 15.9 pM for variant 25, 10.0 – 11.7 pM for variant 26, 22.3 – 24.3 pM for *Ps*TPS-sab (WT), 60.1 – 62.7 pM for variant 6, 23.9 – 38.0 pM for variant 9, and 60.3 – 60.7 pM for variant 11. Kinetic analysis was performed by non-linear regression using the EXCEL template ANEMONA (Hernandez *et al.*, 1998).

5.2.4 Gas Chromatography/Mass Spectrometry (GC/MS) Analysis

Assay products were identified by GC (Agilent 6890A Series)/MSD (5973N mass selective detector, quadrupole analyzer, electron ionization, 70 eV) by comparison of retention times and mass spectra with authentic standards and by comparison with mass spectral libraries (Wiley7Nist05). Monoterpene synthase assay products were analyzed on a DB-WAX capillary column (J&W 122-7032; 250 μm internal diameter, 30 m length, 0.25μm film thickness) with an initial temperature of 40°C (4 min), increasing by 3°C min^{-1} to 85°C then by 30°C min^{-1} to 250°C (held for 2.5 min), injector temperature was 250°C, flow rate 1.4 ml He min^{-1}, and run time 27.00 min. Compounds were quantified using response factors calculated by comparison to a known concentration of isobutylbenzene.

5.2.5 Homology Modeling and Ligand Docking

Homology models for the (+)-3-carene synthase-like enzymes and their variants were produced using the SWISS-MODEL server (Arnold *et al.*, 2006; Kiefer *et al.*, 2009) and underwent energy minimization using the YASARA force field (Krieger *et al.*, 2009). Models were based on the structure of *Salvia officinalis* (+)-bornyl diphosphate synthase (PDB ID code: 1N22B) containing the substrate analog (4*R*)-7-aza-7,8-dihydrolimonene (Whittington *et al.*, 2002). Ramachandran plots of all models verified high stereochemical quality having greater than 90% of residues in most favoured regions. Energy-minimized ligands for docking

were produced using the PRODRG server (Schuttelkopf and Van Aalten, 2004). Docking studies with the α-terpinyl cation and the protein models were performed using Molegro Virtual Docking. The substrate analog (4*R*)-7-aza-7,8-dihydrolimonene was used as a positional template for docking. Because this analog is inverted in the active site of the *Salvia officinalis* (+)-bornyl diphosphate synthase crystal structure, similarity measurements used in the template docking parameters were relaxed to allow for increased flexibility in the positioning of the α-terpinyl cation. This resulted in two to three of the top five most energetically favorable positions of the α-terpinyl cation oriented in the appropriate direction. Of these, the most energetically favorable position was used. The results were visualized in PyMOL (http://www.pymol.org).

5.3 RESULTS

5.3.1 Exchange of Helix J Region Shifts (−)-Sabinene Synthase Product Profile of *Ps*TPS-sab to a (+)-3-Carene Synthase Profile Resembling *Ps*TPS-3car

We performed a series of domain swaps and site-directed substitutions between *Ps*TPS-sab and *Ps*TPS-3car to explore which regions and specific amino acids of these enzymes affect product profiles. Conifer monoterpene synthases of the TPS-d1 group possess an α-domain structure harboring the class I active site (Hyatt *et al.*, 2007; Kampranis *et al.*, 2007; Whittington *et al.*, 2002). This domain adopts an α-α barrel structure comprised of 14 helices and two loops. To test if product profiles could be altered through mutation of the α-domain as seen in other monoterpene synthases (Croteau, 1987; Hyatt and Croteau, 2005; Katoh *et al.*, 2004), an initial domain swap was performed on *Ps*TPS-sab (WT) enzyme so its helix A-helix K region would be identical to that of *Ps*TPS-3car2 (WT) (Table 5.1). The resulting enzyme (variant 1) showed a product profile nearly identical to *Ps*TPS-3car2 (WT), producing 66.1% (+)-3-carene and 6.9% (–)-sabinene plus additional monoterpenes (Table 5.1, Table S5.1), indicating successful conversion into a *Ps*TPS-3car type (+)-3-carene synthase. To identify the specific regions that caused this change in product profile, four additional domain swaps were performed on *Ps*TPS-sab (WT): helix A-helix E (variant 2), helix F-helix G1/2 (variant 3), helix H1-helix I (variant 4), and helix J-helix K (variant 5). Of these, variants 2, 3, and 4 showed no substantial change in product profile compared with *Ps*TPS-sab (WT); however, variant 5 displayed a product profile containing 44.9% (+)-3-carene and 9.5% (–)-sabinene

(Table 5.1). To narrow down which parts of the helix J-helix K region caused this change, we performed separate substitutions of helix J (variant 6), J/K loop, and helix K on *Ps*TPS-sab (WT). Changes in the J/K loop and helix K regions had no effect on product profile compared with *Ps*TPS-sab (WT). In contrast, variant 6 produced a profile similar to that of *Ps*TPS-3car of 39.7% (+)-3-carene and 9.2% (–)-sabinene (Table 5.1).

In summary, these results indicated that sequence variation in the 11-amino acid long helix J region was responsible for much of the difference of *Ps*TPS-sab and *Ps*TPS-3car product profiles.

Table 5.1: Product profiles of *Ps*TPS-3car1, *Ps*TPS-3car2, *Ps*TPS-3car3, *Ps*TPS-sab and their variants.

Variant #	Enzyme Variant	Percent Product Profile			
		(+)-3-Carene	(–)-Sabinene	Terpinolene	Other
	PsTPS-3car1 (WT)	49.2±0.4	8.7±0.1	24.7±0.2	17.4
	PsTPS-3car2 (WT)	67.5±0.1	6.9±0.2	15.4±0.1	10.2
	PsTPS-3car3 (WT)	46.2±1.9	8.8±0.7	29.7±3.0	15.3
	PsTPS-sab (WT)	1.3±0.04	44.7±1.5	35.9±0.03	18.1
1	PsTPS-sab (Helix A – K)	66.1±0.3	6.9±0.1	16.0±0.1	11.0
2	PsTPS-sab (Helix A – E)	2.3±0.03	45.8±4.1	35.6±0.4	16.3
3	PsTPS-sab (Helix F – $G_{1/2}$)	1.3±0.03	48.9±2.5	33.0±0.4	16.8
4	PsTPS-sab (Helix H_1 – I)	1.3±0.00	48.9±0.6	35.1±0.7	14.7
5	PsTPS-sab (Helix J – K)	44.9±0.6	9.5±0.2	20.0±0.4	25.6
6	PsTPS-sab (Helix J)	39.7±0.5	9.2±0.02	20.8±0.3	30.3
7	PsTPS-sab (S589A)	1.2±0.03	42.9±1.3	35.5±0.3	20.4
8	PsTPS-sab (A595G)	2.4±0.00	39.0±3.7	33.3±0.6	25.3
9	PsTPS-sab (F596L)	28.6±0.4	18.7±0.4	24.9±0.8	27.8
10	PsTPS-sab (L599F)	1.5±0.03	42.8±0.9	36.4±0.4	19.3
11	PsTPS-sab (A595G/F596L/L599F)	42.3±1.4	7.3±0.4	20.1±0.6	30.3
12	PsTPS-sab (Helix A – E + Helix J)	56.2±0.4	7.5±0.05	20.2±0.1	16.1
13	PsTPS-sab (Helix F – $G_{1/2}$ + Helix J)	44.1±2.0	7.9±0.8	16.2±0.9	31.8
14	PsTPS-sab (Helix H_1 – I + Helix J)	41.3±1.5	7.7±0.3	16.7±0.7	34.3
15	PsTPS-sab (Helix A + Helix J)	39.4±0.4	8.0±0.2	17.6±0.3	35.0
16	PsTPS-sab (A/C Loop + Helix J)	35.6±0.4	5.4±0.7	20.5±0.7	38.5
17	PsTPS-sab (Helix C + Helix J)	45.4±1.2	5.2±0.6	20.2±0.8	29.2
18	PsTPS-sab (Helix D + Helix J)	50.8±0.2	4.1±0.5	19.5±0.3	25.6
19	PsTPS-sab (Helix D_1 – D_2 + Helix J)	45.4±0.1	8.2±0.1	19.4±0.5	27.0
20	PsTPS-sab (Helix E + Helix J)	43.8±0.1	8.2±0.1	19.2±0.2	28.8
21	PsTPS-3car1 (Helix J)	1.4±0.1	23.9±0.1	56.4±0.4	18.3
22	PsTPS-3car1 (L596F)	5.0±0.2	20.9±1.0	53.5±0.7	20.6
23	PsTPS-3car1 (G595A/L596F/F599L)	1.8±0.1	23.3±0.7	55.2±0.4	19.7
24	PsTPS-3car2 (Helix J)	4.0±0.1	47.9±0.2	36.4±0.3	11.7
25	PsTPS-3car2 (L596F)	12.3±0.1	37.4±0.3	35.4±0.8	14.9
26	PsTPS-3car2 (G595A/L596F/F599L)	4.7±0.03	47.4±0.2	35.2±0.04	12.7
27	PsTPS-3car3 (Helix J)	8.4±1.0	37.7±2.6	23.8±1.1	30.1
28	PsTPS-3car3 (L596F)	9.2±1.5	20.2±1.1	32.3±2.3	38.3
29	PsTPS-3car3 (G595A/L596F/F599L)	5.4±0.5	29.0±3.3	26.4±0.6	39.2

5.3.2 Mutation of Three Amino Acids of the Helix J Region had Major Effects on Shifting the Product Profile of *Ps*TPS-sab to a Profile Resembling *Ps*TPS-3car

Of the four amino acids in the helix J region that differ between *Ps*TPS-sab and *Ps*TPS-3car2 (Fig. 5.1), individual site-directed substitution in positions 589 (variant 7), 595 (variant 8), and 599 (variant 10) produced no change in product profile relative to *Ps*TPS-sab (WT). However, substitution of the Leu in position 596 (variant 9), an amino acid conserved across all known conifer TPS-3car enzymes, to Phe produced an enzyme with a product profile of 28.6% (+)-3-carene and 18.7% (–)-sabinene (Table 5.1, Fig. 5.3d). The proportion of (+)-3-carene in the product profile of *Ps*TPS-sab variant 9 was less compared with the product profiles of variants 5 and 6, suggesting that at least one of the conserved Ala589, Gly595, and Phe599 of the *Ps*TPS-3car enzymes has a synergistic effect with Leu596 on (+)-3-carene formation. To test this hypothesis, we assessed the product profiles of all six possible *Ps*TPS-sab variants that were produced from combinations of variant 9 with additional substitutions at positions 589, 595, and/or 599. Of these, variant 11 produced the highest levels of (+)-3-carene and a product profile closest to variants 5 and 6 with 42.3% (+)-3-carene and 7.3% (–)-sabinene (Table 5.1).

The observed effects these substitutions had on product profiles identified Phe596 as critical in *Ps*TPS-sab for determining (–)-sabinene as the major product, and Leu596 as critical for (+)-3-carene formation in the mutated *Ps*TPS-sab enzyme with positions Gly595 and Phe599 providing synergistic effects. In turn, it can be proposed that these three amino acids play an important role in (+)-3-carene formation in *Ps*TPS-3car.

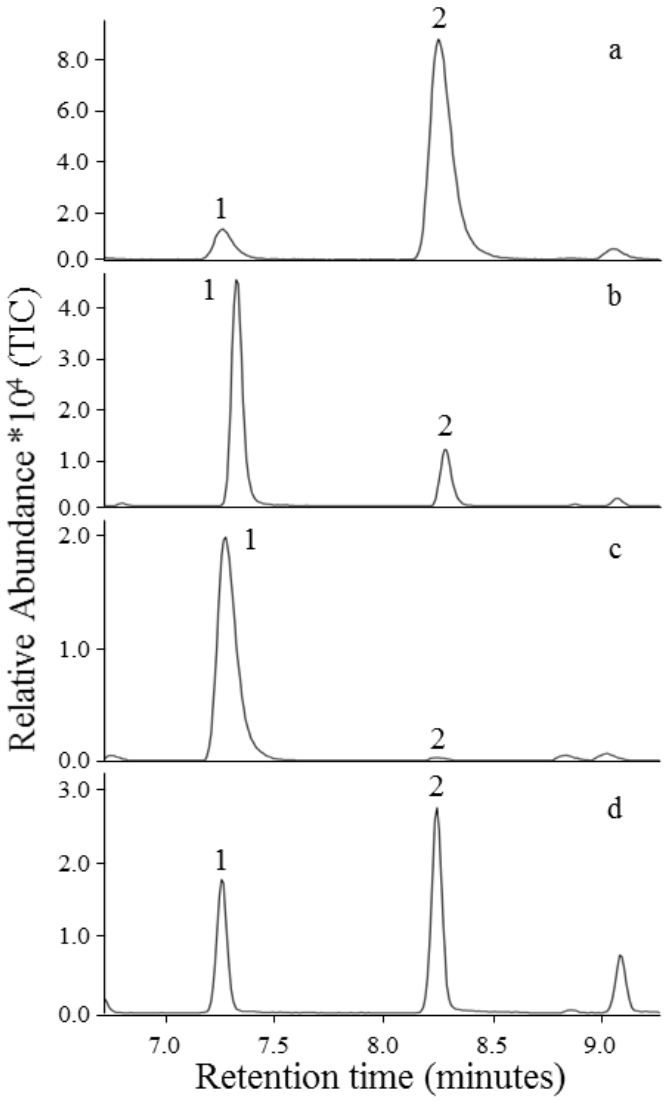


Figure 5.3: Select regions of total ion GCMS traces of products formed by *Ps*TPS-3car and *Ps*TPS-sab and their variants in position 596. Traces **a** and **b** show shifts in the abundance of (–)-sabinene (1) and (+)-3-carene (2) in the product profiles of *Ps*TPS-3car2 (WT) and *Ps*TPS-3car2 (L596F) variant 25, respectively. Traces **c** and **d** show shifts in the abundance of (–)-sabinene (1) and (+)-3-carene (2) in the product profiles of *Ps*TPS-sab (WT) and *Ps*TPS-sab (F596L) variant 9, respectively. Products were confirmed by comparison of mass spectra retention times with those of authentic standards.

5.3.3 Mutations in the Helix A-E Region Synergistically Affect the Shift of (–)-Sabinene Synthase Product Profile to a (+)-3-Carene Synthase Product Profile

The three-helix J amino acid substitutions Ala^{595}-Gly, Phe^{596}-Leu, and Leu^{599}-Phe of *Ps*TPS-sab variant 11 explained the product profile changes observed in variants 5 and 6 relative to *Ps*TPS-sab (WT); however, the (+)-3-carene biosynthesis levels of variant 11 were

only two-thirds of that observed in *Ps*TPS-sab variant 1 and *Ps*TPS-3car2 (WT). To elucidate which additional amino acids promote (+)-3-carene biosynthesis, we performed further domain swaps to variant 6 with the helix A-helix E, helix F-helix G1/2, and helix H-helix I regions of *Ps*TPS-3car2 (variants 12, 13, and 14, respectively). Of the resulting enzymes, variants 13 and 14 showed no increase in (+)-3-carene formation, whereas variant 12 produced a product profile containing 56.2% (+)-3- carene and 7.5% (–)-sabinene. Next, we divided the helix A-helix E region of *Ps*TPS-3car2 into six smaller regions based on individual helices and loops for the design of additional domain swaps in the background of *Ps*TPS-sab variant 6. Of these, the additional swap of helix A (variant 15) and the A/C loop (variant 16) showed no increase in (+)-3-carene formation compared with variant 6. Additional swaps of helix C (variant 17), helix D (variant 18), helix D1-D2 (variant 19), and helix E (variant 20) all showed slight increases in (+)-3-carene resulting in product profiles that contained, respectively, 45.4, 50.8, 45.5, and 43.8% (+)-3-carene and 5.2, 4.1, 8.2, and 8.2% (–)-sabinene. These results suggest that some or all of the 13 amino acids that differ between *Ps*TPS-3car2 (WT) and *Ps*TPS-sab (WT) within the helix C-E region provide additional synergistic effects to (+)-3-carene formation.

5.3.4 Reciprocal Mutations in Positions 595, 596 and 599 Results in Conversion of *Ps*TPS-3car (+)-3-Carene Synthases to (−)-Sabinene Synthases Resembling *Ps*TPS-sab

To substantiate results obtained with substitutions in *Ps*TPS-sab which indicate positions 595, 596 and 599 are critical for determining the predominant (−)-sabinene or (+)-3-carene product profiles of *Ps*TPS-sab and *Ps*TPS-3car, respectively, the reciprocal substitutions corresponding to *Ps*TPS-sab variants 6, 9, and 11 in each of the three different *Ps*TPS-3car enzymes, *Ps*TPS-3car1, *Ps*TPS-3car2, and *Ps*TPS-3car3 were produced (Table 5.1).

These substitutions in the *Ps*TPS-3car1 (WT) background resulted in variants 21, 22, and 23 and reduced the formation of (+)-3-carene to 1.4% (variant 21), 5.0% (variant 22) and 1.8% (variant 23) of the overall product profile compared to 49.2% in *Ps*TPS-3car1 (WT) (Table 5.1). These variant *Ps*TPS-3car1 enzymes produced (−)-sabinene at 23.0%, 20.9%, and 23.3% of total product profile, respectively; compared to 8.7% in *Ps*TPS-3car1 (WT). While α-terpinolene was the major product at 56.4%, 53.5%, and 55.2% of product profile

respectively, these substitutions all caused a large increase in (−)-sabinene product levels. The same three substitutions in *Ps*TPS-3car2 (WT) resulted in variants 24, 25, and 26, with (−)-sabinene as the major product at 47.9%, 37.4%, and 47.4% of total product profile, respectively (Table 5.1, Fig. 5.3). In *Ps*TPS-3car3 (WT), the same three substitutions resulted in variants 27, 28, and 29, again with (−)-sabinene as the major product in the helix J substitution (variant 27; 37.7%) and in the triple amino acid substitution variant 29 (29.0%). (–)-Sabinene was also the second most abundant product in the position 596 variant 28 with 20.2% of product profile respectively. These results demonstrated that Gly^{595}, Leu^{596}, and Phe^{599} are critical in *Ps*TPS-3car for determining (+)-3-carene as the major product, and that Phe^{596} is critical for (−)-sabinene biosynthesis with Ala^{595} and Leu^{599} providing synergistic effects.

5.3.5 Amino Acid in the 596 Position is Important for Functional Plasticity

Reciprocal Phe-Leu substitutions between *Ps*TPS-sab and *Ps*TPS- 3car1, *Ps*TPS-3car2, and *Ps*TPS-3car3 (variants 9, 22, 25, and 28) showed the importance of position 596 in specifying (–)-sabinene or (+)-3-carene as the major product. To further explore the effect of variations in position 596, we tested four additional substitutions in the *Ps*TPS-sab background sequence that introduced side chains with different electrostatic and steric properties (Table 5.2). Substitution of Phe to Glu (variant 30) produced an enzyme with limonene as the major product at 70.9% of product profile. Substitution to His (variant 31) produced an enzyme with a product profile nearly identical to *Ps*TPS-sab (WT) with (–)-sabinene and α-terpinolene being 46.5 and 27.1% of the product profile, respectively. Substitution to Arg (variant 32) produced an enzyme with no detectable activity. Finally, substitution to Gly (variant 33) produced an enzyme displaying a product profile with major products myrcene (27%), α-terpinolene (17.3%), limonene (17.0%), (–)-sabinene (11.2%), and (+)-3-carene (10.0%). The range of changes in product profile seen in these variants highlights the inherent potential for plasticity of this monoterpene synthase and suggests the 596 amino acid plays an important role in directing product outcome and evolution of enzyme function.

Table 5.2: Product profiles of *Ps*TPS-sab variants.

Variant #	Enzyme Variant	Percent Product Profile					
		(+)-3-Carene	(–)-Sabinene	Terpinolene	Myrcene	Limonene	Other
30	PsTPS-sab (F596E)	0.5±0.1	10.2±1.8	5.7±0.5	6.6±0.4	70.9±1.1	6.1
31	PsTPS-sab (F596H)	2.1±0.01	46.5±0.1	27.1±0.3	4.7±0.1	2.9±0.06	16.7
32	PsTPS-sab (F596R)	ND	ND	ND	ND	ND	ND
33	PsTPS-sab (F596G)	10.0±2.6	11.2±2.0	17.3±1.2	27.0±1.5	17.0±1.3	17.5

5.3.6 Homology Models Place Amino Acid 596 in the Active Site of *Ps*TPS-sab and *Ps*TPS-3car

To assess if the amino acid in position 596 interacts with the α-terpinyl cation reaction intermediate or has a catalytic role in directing the predominantly (–)-sabinene or (+)-3-carene product profiles in the *Ps*TPS-sab and *Ps*TPS-3car enzymes, we performed substrate docking experiments with homology models of these proteins and the proposed α-terpinyl cation reaction intermediate. As tertiary structures of plant monoterpene synthases are well conserved, we produced homology models on the structure of (+)-bornyl synthase from S. officinalis (PDB code 1N22) (Whittington *et al.*, 2002) (Fig. 5.4). Due to the large active site cavity volume and consistent with the multiproduct nature of the *Ps*TPS-sab and *Ps*TPS-3car enzymes, several possible ligand positions were obtained in docking experiments with the α-terpinyl cation. To mitigate this problem, docking of the α-terpinyl cation was performed using the position of the substrate analog (4R)-7-aza-7,8-dihydrolimonene within the *S. officinalis* (+)-bornyl diphosphate synthase x-ray crystal structure (PDB code 1N22) as a positional template.

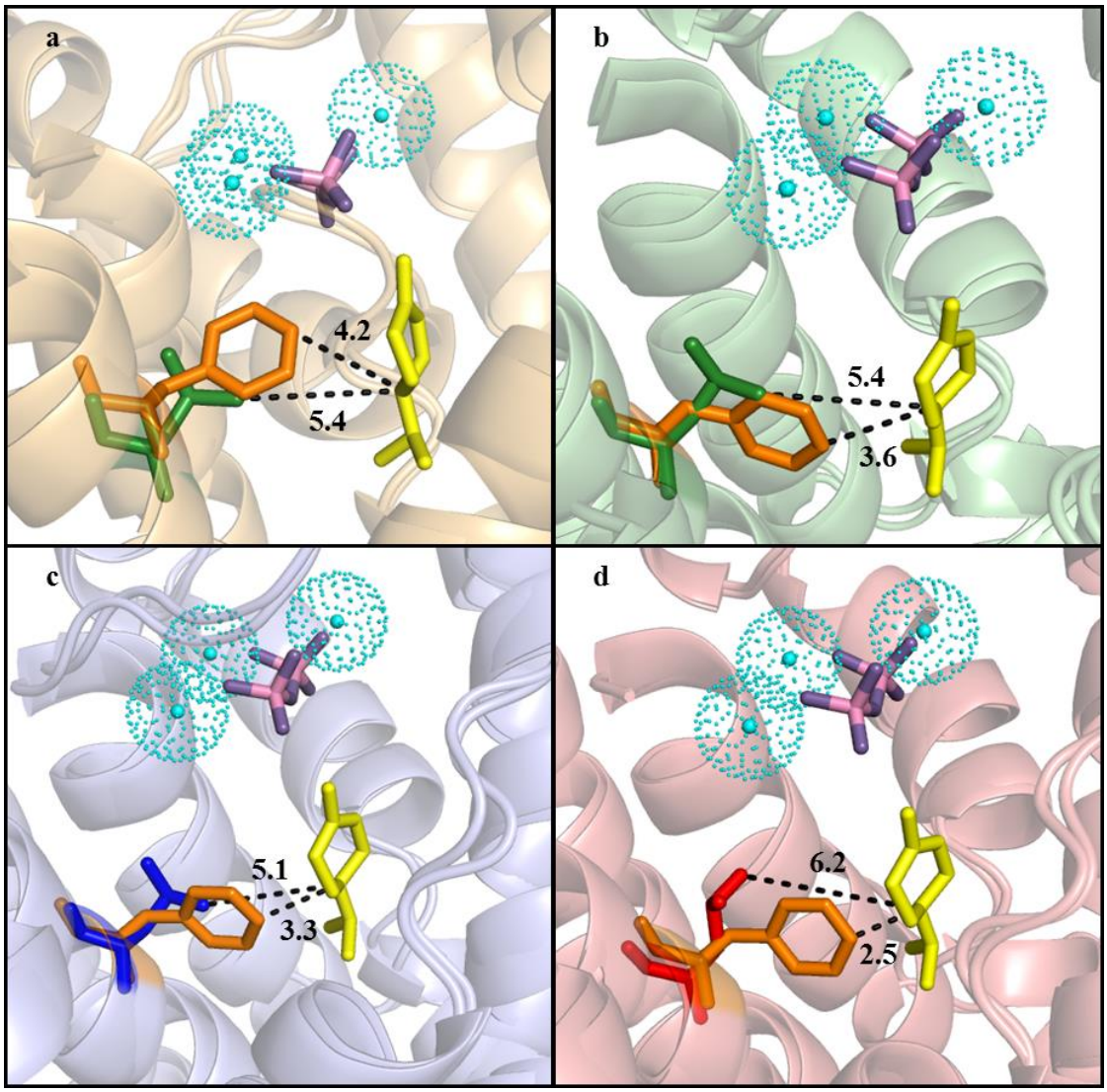


Figure 5.4: Homology models of the active sites of *Ps*TPS-sab (WT), *Ps*TPS-sab (F596L), *Ps*TPS-3car2 (WT), *Ps*TPS-3car2 (L596F), *Ps*TPS-3car1 (WT), *Ps*TPS-3car1 (L596F), *Ps*TPS-3car3 (WT), and *Ps*TPS-3car3 (L596F). Superimposition of the *Ps*TPS-sab (WT) and *Ps*TPS-sab (F596L) enzymes (a), superimposition of the *Ps*TPS-3car2 (WT) and *Ps*TPS-3car2 (L596F) enzymes (b), superimposition of the *Ps*TPS-3car1 (WT) and *Ps*TPS-3car1 (L596F) enzymes (c), and superimposition of the *Ps*TPS-3car3 (WT) and *Ps*TPS-3car3 (L596F) enzymes (d). Helices, loops, and individual amino acids shown in orange denote those found in *Ps*TPS-sab (WT); green denotes those found in *Ps*TPS-3car2 (WT); blue denote those found in *Ps*TPS-3car1 (WT); red denotes those found in *Ps*TPS-3car3 (WT). The Phe or Leu amino acid side chains found in position 596 are shown. The trinuclear magnesium cluster is shown in cyan, and the diphosphate ion is shown in pink and purple. Dotted lines mark the shortest distance between the amino acid side chain in position 596 and the C4 carbon (Fig. 5.2) of the α-terpinyl cation which is shown in yellow.

Models of all four *Ps*TPS-sab and *Ps*TPS-3car enzymes show amino acid 596 positioned in the active site opposite the DDXXD motif and beside the J/K loop believed to

partake in the conformational change that promotes closure of the active site during catalysis (Starks *et al.*, 1997; Starks *et al.*, 1997; Whittington *et al.*, 2002). Of the 36 to 38 amino acids predicted to have side chains within 7 Å of the most energetically favorable conformation of the α-terpinyl cation reaction intermediate, only three amino acids were consistently distinct between *Ps*TPS-3car and *Ps*TPS-sab: amino acids in the 595, 596, and 599 positions. In this position of the reaction intermediate, Leu596 was consistently at a distance of more than 5 Å from the intermediate with no obvious impact on catalysis (Fig. 5.4). In contrast, all models with Phe596 showed this amino acid was consistently within 4 Å from the C4 carbon of the intermediate (Fig. 5.4). Taking the conformational freedom of the carbocation into account, the proximity of the Phe aromatic ring could facilitate steric or van der Waals interactions capable of stabilizing the intermediate in such a manner that promotes (–)-sabinene biosynthesis (Fig. 5.2).

Further exploring the role the 596 amino acid could have on the α-terpinyl cation, additional substrate docking experiments with homology models of variants 30 – 33 and the proposed α-terpinyl cation reaction intermediate were performed. *Ps*TPS-sab (F596E) (variant 30) showed the glutamic acid residue 4.6 Å from the C9 or C10 carbon of the intermediate (Fig. 5.5a). *Ps*TPS-sab (F596H) (variant 31) showed the histidine residue was 4.2 Å from the C4 carbon of the intermediate (Fig. 5.5b). Finally, *Ps*TPS-sab (F596R) and *Ps*TPS-sab (F596G) (variants 32 and 33 respectively) do not show the side chain residues in position in which they would be able to interact with the reaction intermediate.

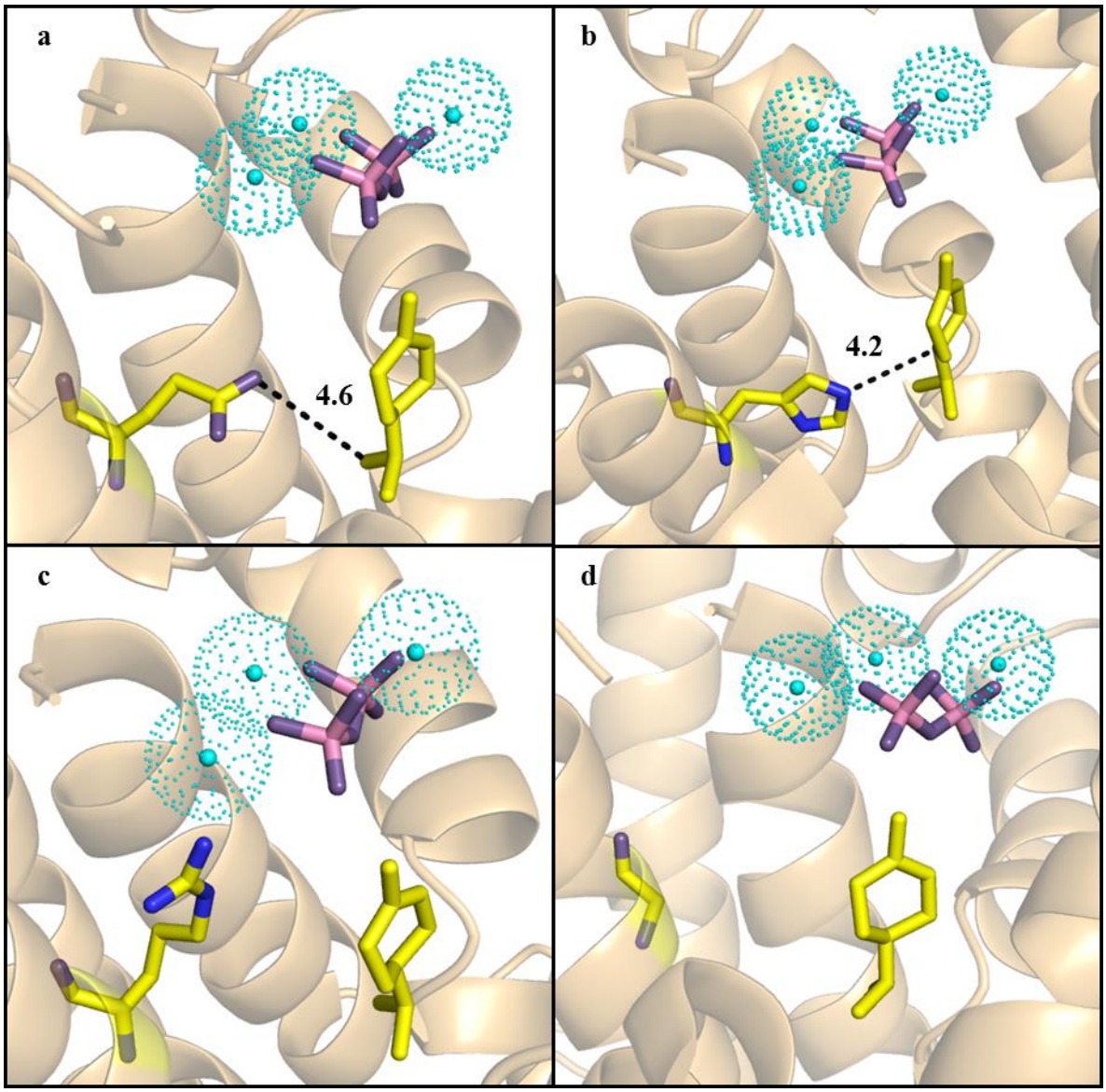


Figure 5.5: Homology models of the active sites of *Ps*TPS-sab (F596E), *Ps*TPS-sab (F596H), *Ps*TPS-sab (F596R), and *Ps*TPS-sab (F596G) active sites. Helices and loops shown in orange colour are those of the *Ps*TPS-sab (WT) background structure. The modified 596 amino acid in each enzyme is shown: Glu in *Ps*TPS-sab (F596E) (a); His in *Ps*TPS-sab (F596H) (b); Arg in *Ps*TPS-sab (F596R) (c); and Gly in *Ps*TPS-sab (F596G) (d). The trinuclear magnesium cluster is shown in cyan, and the diphosphate ion is shown in pink and purple. Where applicable, dotted lines mark the shortest distance between the amino acid side chain in position 596 and the C4 carbon (Fig. 5.2) of the α-terpinyl cation which is shown in yellow.

5.3.7 Kinetic Properties of *Ps*TPS-sab, *Ps*TPS-3car2, and Selected Variants

To investigate if positions that affect contrasting (−)-sabinene or (+)-3-carene product profiles of *Ps*TPS-sab and *Ps*TPS-3car, respectively, also affect differences in other properties

of enzyme activity, kinetic parameters were determined for *Ps*TPS-sab (WT), *Ps*TPS-3car2 (WT), and three variants of each of these enzymes with reciprocal substitutions of either helix J, the 596 amino acid, or the 595, 596, and 599 amino acids (Table 5.3). All of these enzymes which produced (+)-3-carene as the major product, namely *Ps*TPS-3car2 (WT) and the *Ps*TPS-sab variants 6, 9, and 11, had k_{cat} and apparent K_M values in the same order of magnitude. The catalytic efficiencies (k_{cat}/K_M) of these enzymes were similar at 0.37 min^{-1} μM^{-1}, 0.04 min^{-1} μM^{-1}, 0.16 min^{-1} μM^{-1}, and 0.10 min^{-1} μM^{-1}, respectively. Likewise, all enzymes which produced (−)-sabinene as the major product, *Ps*TPS-sab (WT) and the *Ps*TPS-3car2 variants 24, 25, and 26, had apparent K_M values in the same order of magnitude and similar catalytic efficiencies with k_{cat}/K_M values of 0.92 min^{-1} μM^{-1}, 1.09 min^{-1} μM^{-1}, 2.2 min^{-1} μM^{-1}, and 1.22 min^{-1} μM^{-1}, respectively.

Table 5.3: Kinetic profiles of *Ps*TPS-3car2, *Ps*TPS-sab, and their variants.

Variant #	Enzyme Variant	Kinetic Parameters		
		k_{cat} (min^{-1})	K_M (μM)	k_{cat}/K_M (min^{-1} μM^{-1})
	PsTPS-3car2 (WT)	5.86	15.8	0.37
	PsTPS-sab (WT)	32.7	35.7	0.92
6	PsTPS-sab (Helix J)	1.65	38.1	0.04
9	PsTPS-sab (F596L)	4.95	31.6	0.16
11	PsTPS-sab (A595G/F596L/L599F)	2.11	20.2	0.10
24	PsTPS-3car2 (Helix J)	24.9	22.8	1.09
25	PsTPS-3car2 (L596F)	38.6	15.3	2.52
26	PsTPS-sab (G595A/L596F/F599L)	50.3	41.1	1.22

These results demonstrate the single residue substitution in the 596 position (variants 9 and 25) affects both the contrasting product profiles and catalytic efficiencies of enzymes that produce alternatively (−)-sabinene or (+)-3-carene as one of their major products. The single amino acid Leu to Phe substitution in position 596 in the *Ps*TPS-3car2 background (variant 25) increased k_{cat} by more than 6-fold compared to the *Ps*TPS-3car2 (WT). The increased k_{cat} levels of *Ps*TPS-3car2 (L596F) was similar to the k_{cat} of *Ps*TPS-sab (WT), while the opposite effect was observed with the Phe to Leu substitution of the 596 amino acid in the *Ps*TPS-sab background (variant 9).

5.4 DISCUSSION

Using a mutational approach, we investigated the effects that naturally occurring amino acid variations have on functions of monoterpene synthases of the *Ps*TPS-3car/*Ps*TPS-sab family, a group of enzymes that contribute to the phenotypic variation of weevil resistance in Sitka spruce (Hall *et al.*, 2011). Through a progressive series of domain swaps and site-directed mutations, we identified residues at position 596 as critical for enzyme product profile and kinetic properties of the multiproduct *Ps*TPS-3car and *Ps*TPS-sab enzymes. This position appears to be a site of functional plasticity as different substitutions in this position gave rise to enzyme variants with a range of different product profiles. This position had the largest single effect on controlling alternatively (–)-sabinene or (+)-3-carene as major products.

Analysis of the *Ps*TPS-3car (WT) and *Ps*TPS-sab (WT) homology models revealed predicted active sites of similar size and contour. As product selectivity in reactions with multiple outcomes is predominantly determined by energies of the different transition-state structures leading to each product (Weitman and Major, 2010), we suggest the different product profiles of *Ps*TPS-3car and *Ps*TPS-sab are a result of unique active site residues that stabilize a common α-terpinyl cation reaction intermediate in different ways. The changes in (–)-sabinene product levels seen in amino acid substitutions in position 596 of all enzymes (variants 9, 22, 25, and 28) supported this amino acid having a critical role in directing product outcome (Table 5.1; Fig. 5.4). Molecular docking in the class I active site of *Ps*TPS-sab (WT) and *Ps*TPS-3car (L596F) predicted the side chain of Phe596 to be close enough to the α-terpinyl cation for steric blocking or van der Waals forces to mold the energy landscape of the reaction and promote (–)-sabinene formation (Fig. 5.4). Similar to what has been described for other terpene synthases, these effects could help stabilize a carbocation on the C4 carbon, promoting the formation of the terpinen-4-yl cation, through a cation-π interaction (Caruthers *et al.*, 2000; Caruthers *et al.*, 2000; Starks *et al.*, 1997; Thoma *et al.*, 2004; Thoma *et al.*, 2004) or hyperconjugation promoted through a C-H···π interaction (Faraldos *et al.*, 2011; Tantillo, 2010; Tantillo, 2010). Phe596 could also affect the conformation of the cyclohexene ring of the α-terpinyl cation by promoting an equatorial position for the C8 carbon, preventing the 5,8-cyclization needed to form (+)-3-carene (Weitman and Major, 2010).

Additional substitutions at positions 595 and 599 in the background of the *Ps*TPS-sab (F596L), variant 9, and in the backgrounds of the three *Ps*TPS-3car (L596F), variants 22, 25,

and 28, resulted in an increased formation of the predominant monoterpene. The proximity of these two synergistic positions to the 596 position suggests they either contribute to optimal side chain orientation of the residue in the 596 position or alter the active site contour toward stabilizing the conformation of the α-terpinyl cation reaction intermediate within the active site as opposed to directly interacting with the carbocation intermediate. The additional 13-amino acid substitutions within helix C, D, D1-D2, and E (variants 17, 18, 19, and 20, respectively) cause a subtle increase in (+)-3-carene biosynthesis levels. As none of these amino acids were found to be positioned in the active site cavity, these substitutions likely influence product specificity by contributing to the overall barrel structure and/or contour of the class I active site as has been demonstrated in other terpene synthases (Keeling *et al.*, 2008). Quantum chemical calculations have shown that changes in the carbocation structure in response to the distribution of energy in its vibrational nodes could play a substantial role in terpene synthase reaction selectivity (Hong and Tantillo, 2014), suggesting that Phe596 or these synergistic amino acids might play a role in modulating the reaction energy landscape of the intermediate to promote dynamic tendencies that lead to the formation of (–)-sabinene as opposed to directly stabilizing the intermediate.

The distinct product profiles of *Ps*TPS-sab (F596E) (variant 30), *Ps*TPS-sab (F596H) (variant 31), *Ps*TPS-sab (F596R) (variant 32), and *Ps*TPS-sab (F596G) (variant 33) also support the conclusion of the amino acid in position 596 playing a role in modulating carbocation stabilization. Substitution to a glutamic acid (variant 30) introduces a negative charge poised to interact with the reaction intermediate. The observed high levels of limonene formation (Table 5.2) and homology modeling suggest this residue plays a role in promoting the deprotonation of the C9 or C10 carbons of the α-terpinyl cation to terminate the reaction (Fig. 5.2; Fig. 5.5a). Substitution to a histidine (variant 31) resulted in a product profile nearly identical to *Ps*TPS-sab (WT). Protein modeling positions the histidine residue similarly to the phenylalanine in *Ps*TPS-sab (WT) (Fig. 5.4.a; Fig. 5.5b), suggesting the specific steric and aromatic effects histidine has on the reaction intermediate are similar to those of phenylalanine and promote (–)-sabinene biosynthesis. Substitution to an arginine (variant 32) introduces a positive charge and increases the steric constraints on the intermediate (Fig. 5.5c). It is not clear how exactly this affects enzyme functioning, but the absence of enzyme activity suggests it disrupted either the trinuclear magnesium ion configuration or reaction intermediate

formation. Perhaps most interesting, substitution to a glycine (variant 33) shows that the absence of a side chain in the 596 amino acid position results in a very broad monoterpene profile with no major predominant product. In addition to removing a side chain that could direct carbocation stabilization, the glycine residue likely allows the reaction intermediate a larger degree of freedom by expanding the active site pocket (Fig. 5.5d). The result is less specificity in carbocation stabilization and a more general product profile. These results further highlighted the importance of the amino acid in the 596 position promoting specific stabilization of the reaction intermediate to produce a unique product profile.

As this study centered around those residues of *Ps*TPS-car and *Ps*TPS-sab that appear to influence the fate of product profile once the α-terpinyl cation is formed, it is likely that differences in the kinetic properties are due to final reaction steps of product formation and subsequent product release from the active site rather than geranyl diphosphate ionization and the formation of the α-terpinyl cation. Although the formation of (–)-sabinene requires intramolecular proton transfer and cyclization, followed by deprotonation, those enzymes with a phenylalanine in the 596 position had significantly higher catalytic rates and catalytic efficiencies than the enzymes that produced higher levels of (+)-3-carene synthase, which only requires a 5,8-ring closure to form (+)-3-carene and terminate the reaction. Although it is possible that the amino acids in the active site are poised to act on the intermediate leading to (–)-sabinene, our current structural/functional understanding of these enzymes is insufficient to explain the difference in kinetic ability.

Previous work proposed that gene duplications and divergent functional evolution lead to the predominance of either (+)-3-carene or (–)-sabinene as alternative primary products of closely related members of the Sitka spruce (+)-3-carene synthase-like family, the four members of which generally share the same products in varying quantities (Hall *et al.*, 2011). The mutational work presented here supports this hypothesis and provides mechanistic insights into how functional variations in this family may have evolved. The higher sequence identity of *Ps*TPS- 3car2 and *Ps*TPS-3car3 to each other (97.2% nucleotide identity) and their higher sequence relatedness to *Ps*TPS-sab (91.8 and 92.6% nucleotide identity, respectively), compared with the sequence relatedness between *Ps*TPS-3car1 and *Ps*TPS-sab (89.3% nucleotide identity) support a phylogeny of this gene family according to which *Ps*TPS-3car1 has diverged the most from a common ancestor of these four genes (Hall *et al.*, 2011). The

present work supports this pattern, given the relative ease with which *Ps*TPS-3car2 and *Ps*TPS-3car3 could be converted to (–)-sabinene synthases compared with *Ps*TPS-3car1. Mutagenesis studies here show that as little as one base pair substitution could convert a more ancestral (+)-3-carene synthase to a (–)-sabinene synthase highlighting how *Ps*TPS-sab gene function could have evolved. Together, these results provide evidence of active site plasticity and underscore that subtle alteration of the active site contour, such as shifting a backbone atom by a fraction of an Angstrom or the addition or removal of a methyl group can have a large effect on enzyme activity. In the context of the evolution of a family of defense- and resistance-related conifer TPS-d genes, we highlight how gene duplications and conspicuously minor sequence variation may lead to diversification of terpenoid profiles as is seen with the complex mixtures of oleoresin monoterpenes in general, and with the intraspecific variation of (+)-3-carene profiles in Sitka spruce in particular.

5.5 CONCLUSION

In conclusion, we provide a mechanistic underpinning for apparent patterns of the functional evolution of the Sitka spruce (+)-3-carene synthase/(–)-sabinene synthase gene family associated with white pine weevil resistance in Sitka spruce. With as little as one amino acid substitution, we observed large changes in both product profile and kinetic capabilities of the enzymes. These results underscore the large catalytic plasticity of conifer monoterpene synthases as a major factor in the expansive evolutionary diversification of the TPS-d family, which in the case of *Ps*TPS-3car and *Ps*TPS-sab explains the formation of either (+)-3-carene or (–)-sabinene as the major product with a single residue switch.

CHAPTER 6: CONCLUSION

6.1 BRIEF SUMMARY OF WORK

To survive and thrive in the face of many different biotic and abiotic stresses, plants have evolved a complex system of specialized metabolites. For thousands of years, humans have used these metabolites with new functions being regularly identified. The onset of the "-omics age" has enabled us with the capability to collect large amounts of biological data, which can provide insights as to how biological systems function. With the goal of harnessing specialized metabolism for sustained human use, "-omics"-enabled research typically flows through three phases: (i) development of resources to provide a foundation for studying the specialized metabolite system, (ii) characterization of the genes involved in the specialized metabolite system, and (iii) utilizing the functions of the specialized metabolite system for human applications. With an overall emphasis on improving our understanding of specialized metabolism, the body of work in this thesis presents four research chapters, chapter 2 – chapter 5, each focusing on one of the above phases.

In chapter 2, I provided an example of the development of foundational resources needed for studying a specialized metabolite system. With the goal of studying MbA biosynthesis in *C. x crocosmiiflora*, I focused on the development of a series of biological, metabolite profiling, and transcriptome resources. Temporal and spatial analyses provided the first insights into sites of MbA accumulation and potential sites of MbA biosynthesis. These results identified that *in planta*, MbA appears to primarily accumulate in corms, specifically in the cortex exterior to the central vascular cylinder, at statistically similar levels throughout the year. Transcriptome sequencing and high-level gene annotation resulted in the first *C. x crocosmiiflora* transcriptome resource. While *in silico* annotation proved challenging due to a lack of closely related, functionally characterized reference sequences, candidate genes for all parts of the MbA biosynthetic pathway were identified through a combination of sequence homology and comparisons of digital gene expression and metabolite accumulation patterns. Collectively, these results laid a foundation for future functional characterization of MbA biosynthesis in *C. x crocosmiiflora*.

In chapter 3 and 4, I provide examples of characterizing genes involved in the specialized metabolite system. With the goal of elucidating the genes involved in MbA biosynthesis, chapter 3 focused on functionally characterizing the nucleotide interconversion enzymes primarily responsible for the biosynthesis of UDP-xylose and UDP-rhamnose, two components needed in the biosynthesis of MbA. Within the UDP-xylose synthases (UXS) family, I functionally characterized four UXS and one putative UDP-4-keto pentose synthase (UDP-4KPS). While site-directed mutagenesis of putative UDP-4KPS presented a potential molecular and evolutionary mechanism as to why UDP-4-keto pentose was the observed predominant enzymatic product, additional work is needed to be confident in determining if the observed activity is the same *in planta*, or if the observation is an artifact of *in vitro* analysis. Within the UDP-rhamnose synthases (RHM) family, I functionally characterized five RHM and one as a 3,5-epimerase/4-keto-reductase. Through kinetic and relative activity analysis, the most efficient *Cxc*UXS and *Cxc*RHM enzymes were identified. In addition to providing examples of how these gene families could be employed for the future characterization of UDP-xylosyl/rhamnosyltransferase, this body of work further contributes to our understanding of the *C. x crocosmiiflora* MbA biosynthetic pathway.

Continuing the work of elucidating genes involved in MbA biosynthesis, chapter 4 focuses on identifying those GT1 UGTs involved in the late biosynthetic pathway. Exploration of the *C. x crocosmiiflora* transcriptome identified 257 GT1 UGTs and subsequent phylogenetic analysis provided insight into the unique pattern of GT1 UGT clustering. Employing an association analysis between transcript expression and MbA accumulation, 14 *Cxc*UGTs were identified as candidates for involvement in MbA biosynthesis. In exploring the activity of these *Cxc*UGTs against eight potential MbA biosynthetic reactions, only minor activity was observed in four reaction. This low level or lack of activity suggests that none of the 14 candidate *Cxc*UGTs are involved in MbA biosynthesis. As this negative result may be due to errors in the underlying hypotheses of the association analysis, future work should focus on identifying specific conditions that affect MbA accumulation levels for use as a model of MbA biosynthetic gene expression. Collectively, the work in this chapter contributes to our understanding of the phylogenetic distribution of GT1 UGTs within vascular plants and facilitates further identification and functional characterization of *Cxc*UGTs involved in MbA biosynthesis.

In chapter 5, I provided an example of exploring a specialized metabolite system to gain new biological insights. With the goal of showcasing how enzyme plasticity can serve as a mechanistic underpinning for the evolutionary divergence seen in a large gene family of specialized metabolism, I explored a series of domain swaps and site-directed mutations in a set of *Picea sitchensis* monoterpene synthases. With as few as one amino acid substitution in a group of (+)-3-carene and sabinene synthases, I was able to alter both product profile and enzyme kinetics substantially. The diverse product profiles coupled with homology modeling helped underscore how subtle alterations to the active site contour can have large effects on enzyme activity through proposed new interactions with the reaction intermediate. Collectively, these results increase our understanding of how this family could have evolved and provide information on how catalytic plasticity can drive expansive evolutionary diversification of large gene families.

Collectively, the research presented in this thesis provides an example of three phases critical to improving our knowledge of specialized metabolism. While this work represents meaningful advances in each of the biological systems explored, there is still much work that can be done across the two systems presented.

6.2 CONCLUDING REMARKS AND FUTURE DIRECTIONS

As we continue to identify new uses for specialized plant metabolites, it is important to continue broad research in this field and avoid a focus towards the singular. Continuing the types of research presented in this thesis enables us to further establish resources for plant systems and increase our ability to study *non-model* plants, identify genes with unique and not-yet-identified functions, expand our knowledge of gene families, understand how genes within a specialized metabolite biosynthetic pathway interact, and understand the role specialized metabolites play in the biology of the producing plant. The result is an increase in our understanding of specialized metabolism and improved use of these valuable metabolites.

6.2.1 Resource Development for MbA Pathway Discovery and Production

Except for some initial phytochemical work (Asada *et al.*, 1989; Asada *et al.*, 1990; Asada *et al.*, 1988; Asada *et al.*, 1994), this thesis presents the first substantial work towards establishing *C. x crocosmiiflora* as a system for genomics-based discovery of MbA

biosynthesis. While substantial progress has been made in developing resources for elucidating the MbA biosynthetic pathway, the next stages of resource development should focus on improving our understanding of effecters of differential levels of MbA biosynthesis and collectively employing them to help ensure the strongest suite of candidate genes is produced.

One approach could be to explore the use of potential inducers, i.e. elicitors, of MbA biosynthesis in *C. x crocosmiiflora*. Multiple examples are available which show this approach inducing specialized metabolite biosynthesis *in planta* (Li *et al.*, 2009; Smith and Banks, 1986; Verpoorte *et al.*, 2000). If elicitor treatment of plants results in increased accumulation of MbA, it is likely that elicitation would also induce the transcriptional up-regulation of MbA biosynthesis. Should a successful method for elicitation be identified, it would greatly assist with the identification of MbA biosynthesis genes through differential gene expression analysis of contrasting induced and non-induced plants and could also be employed to enhance MbA production in a horticultural system.

The use of plant cell cultures as a resource for specialized metabolite production has received renewed interest (Kolewe *et al.*, 2008; Obembe *et al.*, 2011). Although cell cultures are often limited by low product yields of specialized metabolites, a *C. x crocosmiiflora* derived cell culture would have the advantage of containing the needed protein machinery for MbA biosynthesis. The establishment of callus cultures from *C. x crocosmiiflora* meristem and corms has been completed as a side project of this thesis (Fig. S6.1). While callus lines appeared to be stable, no MbA was identified in any of the samples. Elicitation was attempted with methyl jasmonate, salicylic acid, copper sulfate, growth at 4^{o}C, and growth under UV-C light, but all failed to induce MbA. Identifying what limits or prevents the biosynthesis of MbA in these cell cultures could inform future manipulation strategies aimed at eliciting MbA biosynthesis. Cell cultures able to produce MbA could potentially serve as a production system or be employed to facilitate identification of MbA biosynthetic genes through a differential gene expression analysis.

The genus Crocosmia comprises nine species, each containing a number of varieties selected for different horticultural traits (Barnard et al., 2007; Goldblatt et al., 2004). Metabolite analysis of corms from multiple C. x crocosmiiflora varieties and other Crocosmia species shows large variations in MbA levels (Williams *et al.*, 2015). In addition, metabolite

analysis (using method found in the section 2.2.2) of corms of Watsonia densiflora and Gladiolus grandifloras, members of the same sub-family as Crocosmia, Crocoideae, showed that W. densiflora corms accumulate MbA, though at levels substantially lower than found in C. x crocosmiiflora (Table 6.1).

Table 6.1. Montbretin A accumulation levels in *Watsonia densiflora, C. x crocosmiiflora,* and *Gladiolus grandifloras*. MbA levels were analyzed using the same protocols found in section 2.2.2.

Species	Average MbA /g of FW Tissue (mg/g)
Watsonia densiflora	0.061 ± 0.001
C. x crocosmiiflora	1.62 ± 0.098
Gladiolus grandifloras	0 ± 0

Future work could explore the distribution of MbA biosynthesis across species closely and distantly related to *C. x crocosmiiflora*. Exploring the broader range of natural variations across the *Crocosmia* genus may lead to varieties possessing naturally high levels of MbA, making it an attractive target for plant breeding focused on MbA yields. To my knowledge, past efforts of horticultural selection and improvement of *Crocosmia* has been focused on ornamental traits without consideration of MbA as a valuable chemical trait. As the *Watsonia*, *Crocosmia*, and *Gladiolus* genera represent the three different subclades of the *Crocoideae* subfamily (Goldblatt *et al.*, 2008) (Fig. 6.1), it would be of interest to explore the genetic evolutionary pattern which allowed the biosynthesis of MbA, since variation within it could be employed for gene discovery by analysis of differential transcriptome patterns.

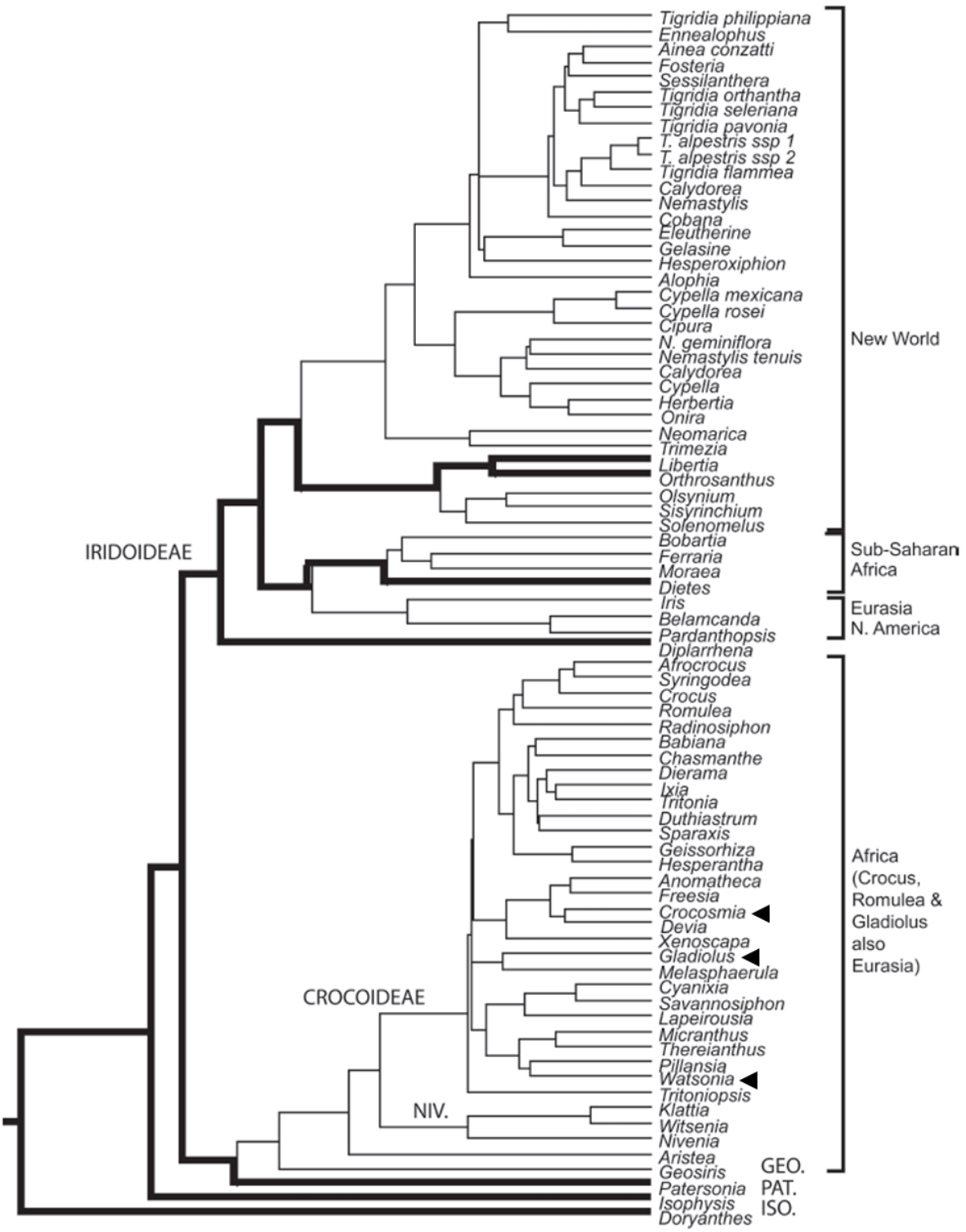


Figure 6.1: Chronogram for the phylogeny of the *Iridaceae* family adapted from Goldblatt et al. (2008). Subfamilies are indicated in capitals (NIV = Nivenioideae; GEO = Geosiridoideae; PAT = Patersonioideae; ISO = Isophysidoideae). Bold lines highlight the lineages of Australasian origin. Arrowheads indicate the three genus *Crocosmia*, *Gladiolus*, and *Watsonia*.

6.2.2 Specialized Metabolite Pathway Characterization

In general, the elucidation of a biosynthetic pathway can be a challenging process and is particularly difficult when working in a system for which no relevant resources and knowledge were previously developed. In this case, approaches for gene discovery are based on *a priori* assumptions about the biological system and these assumptions may or may not be verified. For the work of this thesis, the underlying assumption has been that the site of MbA accumulation coincides with biosynthetic gene expression in the corms. As this has not yet been proven, it would be valuable to pursue additional streams of research aimed at understanding MbA biosynthesis in *C. x crocosmiiflora*.

A significant challenge in establishing the biosynthetic pathway for MbA is the lack of knowledge as to how, and in what sequence, the complex MbA molecule is assembled from its individual building blocks. Moreover, establishing the pathway is further complicated by the difficulties commonly faced when working with GT1 UGTs, which are members of very large gene families. Correctly predicting specific functions based on sequence alone has proven difficult due to the relatively low sequence homology between GT1 UGTs performing similar reactions and the fact that enzymes with high similarity have been found to perform different reactions (Bowles *et al.*, 2005; Bowles *et al.*, 2006). In addition, studies have indicated GT1 UGTs often show broader substrate promiscuity *in vitro* than they do *in vivo* (Bowles *et al.*, 2005). This can present complications with drawing conclusions for functions of GT1 UGTs *in planta* based on *in vitro* assays. To help overcome these problems and continue research on the foundation of this thesis, a two-step *in planta* approach could be taken. First, work should continue to produce a suite of postulated MbA intermediates through degradation of MbA using techniques outlined in Williams *et al.* (2015). While multiple postulated intermediates have been produced, purification has proven difficult due to their similar structures. Second, pending their ability to be taken up, these postulated intermediates should be fed to *Nicotiana benthamiana* plants transiently expressing GT1 UGTs of interest.

In this thesis, I employed a transcriptome-based approach to identify candidate MbA biosynthetic genes. While the decrease in sequencing costs has resulted in this approach being widely employed, protein-based approaches have historically been used to identify specific GT1 UGTs (Hall and De Luca, 2007; Hall *et al.*, 2011; Hall *et al.*, 2012). To help improve the

chances of identifying the MbA biosynthetic GT1 UGTs, *C. x crocosmiiflora* protein extracts should be fractionated and activity against the aforementioned suite of potential intermediates tested. Should any fractions show positive activity, they should be submitted for peptide sequencing and subsequently compared against available transcriptome data to establish a short list of high-potential candidate genes likely involved in MbA biosynthesis.

While natural harvest of MbA from genetically improved *C. x crocosmiiflora* could be employed for large scale production, engineered microbial systems have also been shown to be good options for specialized metabolite biosynthesis due to their fast growth rates, ability to accept genetic modifications, and inexpensive carbon sources (Chang and Keasling, 2006). Sequences for all the *C. x crocosmiiflora* genes involved in the early biosynthetic pathway have been identified in this thesis. As such, these genes could be used to develop a microbial strain capable of producing the precursor molecules of MbA. Such a system could then be used as a platform to study *C. x crocosmiiflora* GT1 UGTs.

6.2.3 Exploring Large Gene Families Involved in Specialized Metabolism

The diversity of specialized metabolites found in the plant kingdom is, in part, the result of some large gene families encoding a diversity of enzymes that act on a relatively small set of core metabolite skeletal structures. By expanding our understanding of such gene families, we can draw new insights into the family's functional evolution, catalytic activity, or specific role(s) *in planta*. The research presented in this thesis on members of the TPS gene family highlights approaches that could be applied to other gene families as well.

Multiple factors have been proposed to play a role in controlling the product profile selectivity of terpene synthesis. These include reactant preorganization (Aaron *et al.*, 2010; Hong and Tantillo, 2011; Noel *et al.*, 2010), geometric constraints imposed by the enzyme active site (Hong and Tantillo, 2009; Sigala *et al.*, 2008; Vedula *et al.*, 2005), selective oriented intermolecular interactions with intermediates and transition state structures (Hong and Tantillo, 2013), and inherent reactivity of carbocations generated from the reactant (Hong and Tantillo, 2010; Pemberton and Tantillo, 2014). In recent years, work exploring the inherent dynamic preferences of carbocation intermediates as they transverse their potential energy landscapes through quasiclassical dynamic calculations has become a topic of interest (Hong and Tantillo, 2010; Hong and Tantillo, 2014; Pemberton and Tantillo, 2014). As a continuation

of the work presented in this thesis, six of the enzymes, *Ps*TPS-sab (WT), *Ps*TPS-sab (F596L), *Ps*TPS-sab (F596E), *Ps*TPS-sab (F596G), *Ps*TPS-sab (F596R), and *Ps*TPS-sab (F596H), which contain only one amino acid difference yet five drastically different product profiles, could be used to further explore the mechanism(s) that terpene synthases employ to selectively control their product profiles and biosynthetic capabilities.

MbA is a specialized metabolite that is highly glycosylated. If the hypothesis of MbA serving as a defense metabolite holds true, results reported by Williams *et al.* (2015) show that the core unit of the two aromatic moieties linked via glucosyl rhamnose disaccharide (mini-MbA) is still able to function as a strong α-amylase inhibitor (Williams *et al.*, 2015). As further glycosylations to mini-MbA improve its inhibitory capabilities, it is possible that evolutionary pressures promoted the further glycosylation, resulting in the powerful inhibitor. This evolution could have taken the form of GT1 UGTs being repurposed through neo-functionalization to perform the additional glycosylations to mini-MbA, or gene duplication and subsequent neo- or sub-functionalization of GT1 UGTs already acting to produce mini-MbA. If either holds true, studying the functional evolution of the GT1 UGTs involved in the biosynthesis of MbA would provide interesting insight into the enzymatic evolution involved in producing the complex, specialized metabolite. Once the GT1 UGTs involved in MbA biosynthesis are identified, the research approach presented in this portion of the thesis could be employed to explore the series of evolutionary changes on an amino acid scale that were needed during the functional evolution of the GT1 UGTs involved in MbA biosynthesis.

BIBLIOGRAPHY

Aaron, J. A., Lin, X., Cane, D. E. and Christianson, D. W. (2010) Structure of epi-isozizaene synthase from *Streptomyces coelicolor* A3 (2), a platform for new terpenoid cyclization templates. *Biochem.,* **49**(8), 1787-1797.

Achnine, L., Huhman, D. V., Farag, M. A., Sumner, L. W., Blount, J. W. and Dixon, R. A. (2005) Genomics-based selection and functional characterization of triterpene glycosyltransferases from the model legume *Medicago truncatula*. *Plant J.,* **41**(6), 875-887.

Adamidi, C., Wang, Y., Gruen, D., Mastrobuoni, G., You, X., Tolle, D., Dodt, M., Mackowiak, S. D., Gogol-Doering, A., Oenal, P., Rybak, A., Ross, E., Sanchez Alvarado, A., Kempa, S., Dieterich, C., Rajewsky, N. and Chen, W. (2011) *De novo* assembly and validation of planaria transcriptome by massive parallel sequencing and shotgun proteomics. *Genome Res.,* **21**(7), 1193-1200.

Aharoni, A., Gaidukov, L., Khersonsky, O., Gould, S. M., Roodveldt, C. and Tawfik, D. S. (2005) The 'evolvability' of promiscuous protein functions. *Nat. Genet.,* **37**(1), 73-76.

Aharoni, A., Gaidukov, L., Yagur, S., Toker, L., Silman, I. and Tawfik, D. S. (2004) Directed evolution of mammalian paraoxonases PON1 and PON3 for bacterial expression and catalytic specialization. *Proc. Natl. Acad. Sci. U. S. A.,* **101**(2), 482-487.

Ajay, M., Gilani, A. H. and Mustafa, M. R. (2003) Effects of flavonoids on vascular smooth muscle of the isolated rat thoracic aorta. *Life Sci.,* **74**(5), 603-612.

Ajikumar, P. K., Xiao, W-H., Tyo1, K. E. J., Wang, Y., Simeon, F., Leonard, E., Mucha, O., Phon, T. H., Pfeifer, B., Stephanopoulos, G. (2010). Isoprenoid Pathway Optimization for Taxol Precursor Overproduction in *Escherichia coli*. *Science,* **330**(6000), 70-74.

Alasalvar, C., Grigor, J. M., Zhang, D., Quantick, P. C. and Shahidi, F. (2001) Comparison of volatiles, phenolics, sugars, antioxidant vitamins, and sensory quality of different colored carrot varieties. *J. Agric. Food Chem.,* **49**(3), 1410-1416.

Alfaro, R. I., Borden, J. H., King, J. N., Tomlin, E. S., McIntosh, R. L. and Bohlmann, J. (2002) Mechanisms of resistance in conifers against shoot infesting insects. In: Mechanisms and Deployment of Resistance in Trees to Insects. Netherlands: Springer.

Ali, R. and Abbas, H. (2003) Response of salt stressed barley seedlings to phenylurea. *Plant Soil Environ.,* **49**(4), 158-162.

Andersen, R., Woods, K., Withers, S., Brayer, G. and Tarling, A. C. (2009) Alpha-amylase inhibitors: the montbretins and uses thereof. *US Patent #8431541*.

Aoki, K., Muraoka, T., Ito, Y., Togashi, Y. and Terauchi, Y. (2010) Comparison of adverse gastrointestinal effects of acarbose and miglitol in healthy men: a crossover study. *Intern. Med.,* **49**(12), 1085-1087.

Arnold, K., Bordoli, L., Kopp, J. and Schwede, T. (2006) The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics,* **22**(2), 195-201.

Asada, Y., Ueoka, T. and Furuya, T. (1990) Novel acylated saponins from montbretia (*Crocosmia x crocosmiiflora*). II. The structures of crocosmiosides C, D, E, F, G and I. *Chem. Pharm. Bull.,* **38**(1), 142-149.

Asada, Y., Ueoka, T. and Furuya, T. (1989) Novel acylated saponins from montbretia (*Crocosmia x crocosmiiflora*). Isolation of saponins and the structures of crocosmiosides A, B, and H. *Chem. Pharm. Bull.,* **37**(8), 2139-2146.

Asada, Y., Hirayama, Y. and Furuya, T. (1988) Acylated flavonols from *Crocosmia x crocosmiiflora*. *Phytochemistry.,* **27**(5), 1497-1501.

Asada, Y., Ikeno, M. and Furuya, T. (1994) Acylated saponins, masonosides A–C, from the corms of *Crocosmia masoniorum*. *Phytochemistry,* **35**(3), 757-764.

Asmann, Y. W., Hossain, A., Necela, B. M., Middha, S., Kalari, K. R., Sun, Z., Chai, H. S., Williamson, D. W., Radisky, D., Schroth, G. P., Kocher, J. P., Perez, E. A. and Thompson, E. A. (2011) A novel bioinformatics pipeline for identification and characterization of fusion transcripts in breast cancer and normal cell lines. *Nucleic Acids Res.,* **39**(15), e100.

Atalay, M., Gordillo, G., Roy, S., Rovin, B., Bagchi, D., Bagchi, M. and Sen, C. K. (2003) Anti-angiogenic property of edible berry in a model of hemangioma. *FEBS Lett.,* **544**(1-3), 252-257.

Attia, M., Kim, S. and Ro, D. (2012) Molecular cloning and characterization of (+)-epi-α-bisabolol synthase, catalyzing the first step in the biosynthesis of the natural sweetener, hernandulcin, in *Lippia dulcis*. *Arch. Biochem. Biophys.,* **527**(1), 37-44.

Augustin, M. M., Ruzicka, D. R., Shukla, A. K., Augustin, J. M., Starks, C. M., O'Neil-Johnson, M., McKain, M. R., Evans, B. S., Barrett, M. D., Smithson, A., Wong, G. K., Deyholos, M. K., Edger, P. P., Pires, J. C., Leebens-Mack, J., Mann, D. A. and Kutchan, T. M. (2015) Elucidating steroid alkaloid biosynthesis in *Veratrum californicum*: production of verazine in Sf9 cells. *Plant J.,* **82**(6), 991-1003.

Baba, S. A., Mohiuddin, T., Basu, S., Swarnkar, M. K., Malik, A. H., Wani, Z. A., Abbas, N., Singh, A. K. and Ashraf, N. (2015) Comprehensive transcriptome analysis of crocus sativus for discovery and expression of genes involved in apocarotenoid biosynthesis. *BMC Genomics,* 16, 698.

Bakker, H., Routier, F., Oelmann, S., Jordi, W., Lommen, A., Gerardy-Schahn, R. and Bosch, D. (2005) Molecular cloning of two *Arabidopsis* UDP-galactose transporters by complementation of a deficient Chinese hamster ovary cell line. *Glycobiology*, **15**(2), 193-201.

Balasundram, N., Sundram, K. and Samman, S. (2006) Phenolic compounds in plants and agri-industrial by-products: antioxidant activity, occurrence, and potential uses. *Food Chem.*, **99**(1), 191-203.

Baldwin, T. C., Handford, M. G., Yuseff, M. I., Orellana, A. and Dupree, P. (2001) Identification and characterization of GONST1, a golgi-localized GDP-mannose transporter in *Arabidopsis*. *Plant Cell*, **13**(10), 2283-2295.

Ballerini, E. S., Mockaitis, K. and Arnold, M. L. (2013) Transcriptome sequencing and phylogenetic analysis of floral and leaf MIKC C MADS-box and R2R3 MYB transcription factors from the monocot *Iris fulva*. *Gene*, **531**(2), 337-346.

Bao, X., Singletary, G. W., Wetterberg, D. J., Nair, R., Dhugga, K. S., Liebergesell, M. and Selinger, D. A. (2014) UDP-xylose synthases (UXS) polynucleotides, polypeptides and uses thereof. *US Patent 8735652*.

Barnard, L., Rodd, T. and Bryant, G. (2007) The plant finder: the right plants for every garden. Canada: Firefly Books.

Bar-Peled, M. and O'Neill, M. A. (2011) Plant nucleotide sugar formation, interconversion, and salvage by sugar recycling*. *Annu. Rev. Plant Biol.*, **62**, 127-155.

Barvkar, V. T., Pardeshi, V. C., Kale, S. M., Kadoo, N. Y. and Gupta, V. S. (2012) Phylogenomic analysis of UDP glycosyltransferase 1 multigene family in *Linum usitatissimum* identified genes with varied expression patterns. *BMC Genomics*, **13**, 175.

Bertea, C., Freije, J., Van der Woude, H., Verstappen, F., Perk, L., Marquez, V., De Kraker, J., Posthumus, M., Jansen, B., De Groot, A., Franssen, M. C. R. and Bouwmeester, H. J. (2005) Identification of intermediates and enzymes involved in the early steps of artemisinin biosynthesis in *Artemisia annua*. *Planta Med.*, **71**(1), 40-47.

Binder, B. Y., Peebles, C. A., Shanks, J. V. and San, K. (2009) The effects of UV-B stress on the production of terpenoid indole alkaloids in *Catharanthus roseus* hairy roots. *Biotechnol. Prog.*, **25**(3), 861-865.

Bindschedler, L. V., Wheatley, E., Gay, E., Cole, J., Cottage, A. and Bolwell, G. P. (2005) Characterisation and expression of the pathway from UDP-glucose to UDP-xylose in differentiating tobacco tissue. *Plant Mol. Biol.*, **57**(2), 285-301.

Boghigian, B. A., Zhang, H., Pfeifer, B. A. (2011). Multi-factorial engineering of heterologous polyketide production in *Escherichia coli* reveals complex pathway interactions. *Biotechnol. Bioeng.,* **108**(6), 1360–1371.

Bohlmann, J., Steele, C. L. and Croteau, R. (1997) Monoterpene synthases from grand fir (*Abies grandis*) cDNA isolation, characterization, and functional expression of mycrene synthase, (−)-(4S)-limonene synthase, and (−)-(1S, 5S)-pinene synthase. *J. Biol. Chem.,* **272**(35), 21784-21792.

Bolger, A. M., Lohse, M. and Usadel, B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics,* **30**(15), 2114-2120.

Bongue-Bartelsman, M. and Phillips, D. (1995) Nitrogen stress regulates gene expression of enzymes in the flavonoid biosynthetic pathway of tomato. *Plant Physiol. Biochem.,* **33**(5), 539-546.

Bonisch, F., Frotscher, J., Stanitzek, S., Ruhl, E., Wust, M., Bitz, O. and Schwab, W. (2014) Activity-based profiling of a physiologic aglycone library reveals sugar acceptor promiscuity of family 1 UDP-glucosyltransferases from grape. *Plant Physiol.,* **166**(1), 23-39.

Borevitz, J. O. and Ecker, J. R. (2004) Plant genomics: the third wave. *Annu. Rev. Genomics Hum. Genet.,* **5**, 443-477.

Bowles, D., Isayenkova, J., Lim, E. K. and Poppenberger, B. (2005) Glycosyltransferases: managers of small molecules. *Curr. Opin. Plant Biol.,* **8**(3), 254-263.

Bowles, D., Lim, E., Poppenberger, B. and Vaistij, F. E. (2006) Glycosyltransferases of lipophilic small molecules. *Annu. Rev. Plant Biol.,* **57**, 567-597.

Brazier-Hicks, M., Offen, W. A., Gershater, M. C., Revett, T. J., Lim, E., Bowles, D. J., Davies, G. J. and Edwards, R. (2007) Characterization and engineering of the bifunctional N-and O-glucosyltransferase involved in xenobiotic metabolism in plants. *Proc. Nat. Acad. Sci.,* **104**(51), 20238-20243.

Brod, S. (2000) Unregulated inflammation shortens human functional longevity. *Inflammation Res.,* **49**(11), 561-570.

Brown, G.D. (2010) The biosynthesis of artemisinin (Qinghaosu) and the phytochemistry of *Artemisia annua* L.(Qinghao). *Molecules,* **15**(11), 7603-7698.

Cai, J., Liu, X., Vanneste, K., Proost, S., Tsai, W. C., Liu, K. W., Chen, L. J., He, Y., Xu, Q., Bian, C., Zheng, Z., Sun, F., Liu, W., Hsiao, Y. Y., Pan, Z., Hsu, C. C., Yang, Y., Hsu, Y., Chuang, Y., Dievart, A., Dufayard, J. F., Xu, X., Wang, J., Wang, J., Xiao, X., Zhao, X., Du, R., Zhang, G., Wang, M., Su, Y., Xie, G., Liu, Li, L., Huang, L.,

Luo, Y., Chen, H., Van de Peer, Y. and Liu, Z. (2015) The genome sequence of the orchid *Phalaenopsis equestris*. *Nat. Genet.*, **47**(1), 65-72.

Campbell, J., Davies, G., Bulone, V. and Henrissat, B. (1997) A classification of nucleotide-diphospho-sugar glycosyltransferases based on amino acid sequence similarities. *Biochem. J.*, **326**(3), 929-939.

Cane, D. E., Yang, G., Xue, Q. and Shim, J. H. (1995) Trichodiene synthase. Substrate specificity and inhibition. *Biochem.*, **34**(8), 2471-2479.

Cane, D. E., Saito, A., Croteau, R., Shaskus, J. and Felton, M. (1982) Enzymic cyclization of geranyl pyrophosphate to bornyl pyrophosphate. Role of the pyrophosphate moiety. *J. Am. Chem. Soc.*, **104**(21), 5831-5833.

Caputi, L., Lim, E. K. and Bowles, D. J. (2008) Discovery of new biocatalysts for the glycosylation of terpenoid scaffolds. *Chem. Eur. J.*, **14**(22), 6656-6662.

Caputi, L., Malnoy, M., Goremykin, V., Nikiforova, S. and Martens, S. (2012) A genome-wide phylogenetic reconstruction of family 1 UDP-glycosyltransferases revealed the expansion of the family during the adaptation of plants to life on land. *Plant J.*, **69**(6), 1030-1042.

Carpita, N. C. and Gibeaut, D. M. (1993) Structural models of primary cell walls in flowering plants: consistency of molecular structure with the physical properties of the walls during growth. *Plant J.*, **3**(1), 1-30.

Carrón, R., Sanz, E., Puebla, P., Martín, M. L., San Román, L. and Guerrero, M. F. (2010) Mechanisms of relaxation induced by flavonoid ayanin in isolated aorta rings from wistar rats. *Colombia Medica*, **41**(1), 10-16.

Cartwright, A. M., Lim, E. K., Kleanthous, C. and Bowles, D. J. (2008) A kinetic analysis of regiospecific glucosylation by two glycosyltransferases of *Arabidopsis thaliana*: domain swapping to introduce new activities. *J. Biol. Chem.*, **283**(23), 15724-15731.

Caruthers, J. M., Kang, I., Rynkiewicz, M. J., Cane, D. E. and Christianson, D. W. (2000) Crystal structure determination of aristolochene synthase from the blue cheese mold, *Penicillium roqueforti*. *J. Biol. Chem.*, **275**(33), 25533-25539.

Chang, M. C. Y. and Keasling, J. D. (2006) Production of isoprenoid pharmaceuticals by engineered microbes. *Nat. Chem. Biol.*, **2**(12), 674-681.

Chen, F., Tholl, D., Bohlmann, J. and Pichersky, E. (2011) The family of terpene synthases in plants: a mid-size family of genes for specialized metabolism that is highly diversified throughout the kingdom. *Plant J.*, **66**(1), 212-229.

Chen, S., Yang, P., Jiang, F., Wei, Y., Ma, Z. and Kang, L. (2010) *De novo* analysis of transcriptome dynamics in the migratory locust during the development of phase traits. *PLoS One*, **5**(12), e15633.

Cheng, A. X., Lou, Y. G., Mao, Y. B., Lu, S., Wang, L. J. and Chen, X. Y. (2007) Plant terpenoids: biosynthesis and ecological functions. *J. of Integr. Plant Biol.,* **49**(2), 179-186.

Cholewa, E. and Griffith, M. (2004) The unusual vascular structure of the corm of *Eriophorum vaginatum*: implications for efficient retranslocation of nutrients. *J. Exp. Bot.,* **55**(397), 731-741.

Christen, A. A., Gibson, D. M. and Bland, J. (1991) Production of taxol or taxol-like compounds in cell culture. *US Patent 5019504A.*

Christianson, D. W. (2006) Structural biology and chemistry of the terpenoid cyclases. *Chem. Rev.,* **106**(8), 3412.

Colby, S. M., Alonso, W. R., Katahira, E. J., McGarvey, D. J. and Croteau, R. (1993) 4S-limonene synthase from the oil glands of spearmint (*Mentha spicata*). cDNA isolation, characterization, and bacterial expression of the catalytically active monoterpene cyclase. *J. Biol. Chem.,* **268**(31), 23016-23024.

Conesa, A., Gotz, S., Garcia-Gomez, J. M., Terol, J., Talon, M. and Robles, M. (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics,* **21**(18), 3674-3676.

Coutinho, P. M., Deleury, E., Davies, G. J. and Henrissat, B. (2003) An evolving hierarchical family classification for glycosyltransferases. *J. Mol. Biol.,* **328**(2), 307-317.

Cragg, G. M. (1998) Paclitaxel (taxol®): a success story with valuable lessons for natural product drug discovery and development. *Med. Res. Rev.,* **18**(5), 315-331.

Crandall, J. P., Knowler, W. C., Kahn, S. E., Marrero, D., Florez, J. C., Bray, G. A., Haffner, S. M., Hoskin, M. and Nathan, D. M. (2008) The prevention of type 2 diabetes. *Nat. Clin. Pract. Endocrinol. Metab.,* **4**(7), 382-393.

Croteau, R. and Felton, M. (1981) Conversion of [1-^{3} H_2, G-^{14}C]geranyl pyrophosphate to cyclic monoterpenes without loss of tritium. *Arch. Biochem. Biophys.,* **207**(2), 460-464.

Croteau, R. and Karp, F. (1979) Biosynthesis of monoterpenes: preliminary characterization of bornyl pyrophosphate synthetase from sage (*Salvia officinalis*) and demonstration that geranyl pyrophosphate is the preferred substrate for cyclization. *Arch. Biochem. Biophys.,* **198**(2), 512-522.

Croteau, R., Felton, M. and Ronald, R. C. (1980) Biosynthesis of monoterpenes: preliminary characterization of l-endo-fenchol synthetase from fennel (*Foeniculum vulgare*) and

evidence that no free intermediate is involved in the cyclization of geranyl pyrophosphate to the rearranged product. *Arch. Biochem. Biophys.,* **200**(2), 534-546.

Croteau, R. B., Shaskus, J. J., Renstrom, B., Felton, N. M., Cane, D. E., Saito, A. and Chang, C. (1985a) Mechanism of the pyrophosphate migration in the enzymic cyclization of geranyl and linalyl pyrophosphates to (+)-and (−)-bornyl pyrophosphates. *Biochem.,* **24**(25), 7077-7085.

Croteau, R. (1987) Biosynthesis and catabolism of monoterpenoids. *Chem. Rev.,* **87**(5), 929-954.

Croteau, R., Felton, N. M. and Wheeler, C. J. (1985b) Stereochemistry at C-1 of geranyl pyrophosphate and neryl pyrophosphate in the cyclization to (+)- and (−)-bornyl pyrophosphate. *J. Biol. Chem.,* **260**(10), 5956-5962.

Croteau, R., Satterwhite, D. M., Wheeler, C. J. and Felton, N. M. (1989) Biosynthesis of monoterpenes. Stereochemistry of the enzymatic cyclizations of geranyl pyrophosphate to (+)-alpha-pinene and (−)-beta-pinene. *J. Biol. Chem.,* **264**(4), 2075-2080.

Cushnie, T. T. and Lamb, A. J. (2005) Antimicrobial activity of flavonoids. *Int. J. Antimicrob. Agents,* **26**(5), 343-356.

Cuvelier, M. E., Richard, H. and Berset, C. (1996) Antioxidative activity and phenolic composition of pilot-plant and commercial extracts of sage and rosemary. *J. Am. Oil Chem. Soc.,* **73**(5), 645-652.

D'Auria, J. C. (2006) Acyltransferases in plants: a good time to be BAHD. *Curr. Opin. Plant Biol.,* **9**(3), 331-340.

D'Auria, J. C., Reichelt, M., Luck, K., Svatoš, A. and Gershenzon, J. (2007) Identification and characterization of the BAHD acyltransferase malonyl CoA: anthocyanidin 5-O-glucoside-6″-O-malonyltransferase (*At*5MAT) in *Arabidopsis thaliana*. *FEBS Lett.,* **581**(5), 872-878.

Davis, E. M. and Croteau, R. (2000) Cyclization enzymes in the biosynthesis of monoterpenes, sesquiterpenes, and diterpenes. *Top. Curr. Chem.,* **209**(1), 53-95.

De Bruyn, F., Maertens, J., Beauprez, J., Soetaert, W. and De Mey, M. (2015) Biotechnological advances in UDP-sugar based glycosylation of small molecules. *Biotechnol. Adv.,* **33**(2), 288-302.

DeJong, J. M., Liu, Y., Bollon, A. P., Long, R. M., Jennewein, S., Williams, D. and Croteau, R. B. (2006) Genetic engineering of taxol biosynthetic genes in *Saccharomyces cerevisiae*. *Biotechnol. Bioeng.,* **93**(2), 212-224.

Diaz, M. N., Frei, B., Vita, J. A. and Keaney Jr, J. F. (1997) Antioxidants and atherosclerotic heart disease. *N. Engl. J. Med.,* **337**(6), 408-416.

Dixon, R. A. and Strack, D. (2003) Phytochemistry meets genome analysis, and beyond. *Phytochem.,* **62**(6), 815-816.

Dominy, N. J., Vogel, E. R., Yeakel, J. D., Constantino, P. and Lucas, P. W. (2008) Mechanical properties of plant underground storage organs and implications for dietary models of early hominins. *Evol. Biol.,* **35**(3), 159-175.

Driouich, A., Faye, L. and Staehelin, A. (1993) The plant Golgi apparatus: a factory for complex polysaccharides and glycoproteins. *Trends Biochem. Sci.,* **18**(6), 210-214.

Du, Q., Pan, W., Tian, J., Li, B. and Zhang, D. (2013) The UDP-glucuronate decarboxylase gene family in *Populus*: structure, expression, and association genetics. *PLoS One,* **8**(4), e60880.

Duan, X. C., Lu, A. M., Gu, B., Cai, Z. P., Ma, H. Y., Wei, S., Laborda, P., Liu, L. and Voglmeir, J. (2015) Functional characterization of the UDP-xylose biosynthesis pathway in *Rhodothermus marinus*. *Appl. Microbiol. Biotechnol.* **99**, 9463-9472.

Duax, W. L., Ghosh, D. and Pletnev, V. (2000) Steroid dehydrogenase structures, mechanism of action, and disease. *Vitam. Horm.,* **58**(1), 121-148.

Durand, C., Vicre-Gibouin, M., Follet-Gueye, M. L., Duponchel, L., Moreau, M., Lerouge, P. and Driouich, A. (2009) The organization pattern of root border-like cells of *Arabidopsis* is dependent on cell wall homogalacturonan. *Plant Physiol.,* **150**(3), 1411-1421.

Ebert, B., Rautengarten, C., Guo, X., Xiong, G., Stonebloom, S., Smith-Moritz, A. M., Herter, T., Chan, L. J., Adams, P. D., Petzold, C. J., Pauly, M., Willats, W. G., Heazlewood, J. L. and Scheller, H. V. (2015) Identification and characterization of a Golgi-localized UDP-xylose transporter family from *Arabidopsis*. *Plant Cell,* **27**(4), 1218-1227.

Eixelsberger, T., Sykora, S., Egger, S., Brunsteiner, M., Kavanagh, K. L., Oppermann, U., Brecker, L. and Nidetzky, B. (2012) Structure and mechanism of human UDP-xylose synthase: evidence for a promoting role of sugar ring distortion in a three-step catalytic conversion of UDP-glucuronic acid. *J. Biol. Chem.,* **287**(37), 31349-31358.

Evert, R. F. (2006) Esau's plant anatomy, meristems, cells, and tissues of the plant body: their structure, function, and development. USA: John Wiley & Sons.

Facchini, P. J., Bohlmann, J., Covello, P. S., De Luca, V., Mahadevan, R., Page, J. E., Ro, D. K., Sensen, C. W., Storms, R. and Martin, V. J. J. (2012) Synthetic biosystems for the production of high-value plant metabolites. *Trends Biotechnol.,* **30**(3), 127-131.

Fäldt, J., Martin, D., Miller, B., Rawat, S. and Bohlmann, J. (2003) Traumatic resin defense in Norway spruce (*Picea abies*): methyl jasmonate-induced terpene synthase gene expression, and cDNA cloning and functional characterization of (+)-3-carene synthase. *Plant Mol. Biol.,* **51**(1), 119-133.

Faraldos, J. A., Antonczak, A. K., González, V., Fullerton, R., Tippmann, E. M. and Allemann, R. K. (2011) Probing eudesmane cation-π interactions in catalysis by aristolochene synthase with non-canonical amino acids. *J. Am. Chem. Soc.,* **133**(35), 13906-13909.

Fay, M. F., Rudalt, P. J., Sullivan, S., Stobart, K. L., de Bruijn, A. Y., Reeves, G., Qamaruz-Zaman, F., Hong, W. P., Joseph, J., Hahn, W. J., Conran, J. G. and Chase, M. W. (2000) Phylogenetic studies asparagales based on four plastid DNA regions. In: Monocots: systematics and Evolution. Australia: CSIRO.

Feldmeyer, B., Wheat, C. W., Krezdorn, N., Rotter, B. and Pfenninger, M. (2011) Short read Illumina data for the *de novo* assembly of a non-model snail species transcriptome (*Radix balthica*, *Basommatophora*, *Pulmonata*), and a comparison of assembler performance. *BMC Genomics,* **12**, 317.

Ferrer, J., Austin, M., Stewart, C. and Noel, J. (2008) Structure and function of enzymes involved in the biosynthesis of phenylpropanoids. *Plant Physiol. Biochem.,* **46**(3), 356-370.

Fischbach, R., Kossmann, B., Panten, H., Steinbrecher, R., Heller, W., Seidlitz, H., Sandermann, H., Hertkorn, N. and Schnitzler, J. (1999) Seasonal accumulation of ultraviolet-B screening pigments in needles of Norway spruce (*Picea abies* (L.) Karst.). *Plant, Cell Environ.,* **22**(1), 27-37.

Fischer, D., Stich, K., Britsch, L. and Grisebach, H. (1988) Purification and characterization of (+)-dihydroflavonol (3-hydroxyflavanone) 4-reductase from flowers of *Dahlia variabilis*. *Arch. Biochem. Biophys.,* **264**(1), 40-47.

Forkmann, G., Heller, W. and Grisebach, H. (1980) Anthocyanin biosynthesis in flowers of *Matthiola incana* flavanone 3-and flavonoid 3′-hydroxylases. *Zeitschrift Für Naturforschung C,* **35**(9-10), 691-695.

Fowler, M. J. (2008) Microvascular and macrovascular complications of diabetes. *Clin. Diabetes,* **26**(2), 77-82.

Frankel, E., Waterhouse, A. and Kinsella, J. E. (1993) Inhibition of human LDL oxidation by resveratrol. *The Lancet,* **341**(8852), 1103-1104.

Frei, H., Lüthy, J., Brauchli, J., Zweifel, U., Würgler, F. E. and Schlatter, C. (1992) Structure/activity relationships of the genotoxic potencies of sixteen pyrrolizidine

alkaloids assayed for the induction of somatic mutation and recombination in wing cells of *Drosophila melanogaster*. *Chem. Biol. Interact.*, **83**(1), 1-22.

Fu, N., Wang, Q. and Shen, H. (2013) *De novo* assembly, gene annotation and marker development using Illumina paired-end transcriptome sequences in celery (*Apium graveolens* L.). *PLoS One*, **8**(2), e57686.

Funaki, A., Waki, T., Noguchi, A., Kawai, Y., Yamashita, S., Takahashi, S. and Nakayama, T. (2015) Identification of a highly specific isoflavone 7-O-glucosyltransferase in the soybean (*Glycine max* (L.) *merr.*). *Plant Cell Physiol.*, **56**(8), 1512-1520.

Gachon, C. M., Langlois-Meurinne, M. and Saindrenan, P. (2005) Plant secondary metabolism glycosyltransferases: the emerging functional analysis. *Trends Plant Sci.*, **10**(11), 542-549.

Gerlt, J. A., Babbitt, P. C. and Rayment, I. (2005) Divergent evolution in the enolase superfamily: the interplay of mechanism and specificity. *Arch. Biochem. Biophys.*, **433**(1), 59-70.

Gershenzon, J. and Kreis, W. (1999) Biochemistry of terpenoids: monoterpenes, sesquiterpenes, diterpenes, sterols, cardiac glycosides and steroid saponins. *Biochem. Plant Sec. Metabol.*, **2**, 222-299.

Ghose, K., Selvaraj, K., McCallum, J., Kirby, C. W., Sweeney-Nixon, M., Cloutier, S. J., Deyholos, M., Datla, R. and Fofana, B. (2014) Identification and functional characterization of a flax UDP-glycosyltransferase glucosylating secoisolariciresinol (SECO) into secoisolariciresinol monoglucoside (SMG) and diglucoside (SDG). *BMC Plant Biol.*, **14**, 82.

Gibeaut, D. M. and Carpita, N. C. (1994) Biosynthesis of plant cell wall polysaccharides. *Faseb J.*, **8**(12), 904-915.

Gibon, Y., Usadel, B., Blaesing, O. E., Kamlage, B., Hoehne, M., Trethewey, R. and Stitt, M. (2006) Integration of metabolite with transcript and enzyme activity profiling during diurnal cycles in *Arabidopsis* rosettes. *Genome Biol.*, **7**(8), 1-23.

Goff, S. A., Ricke, D., Lan, T. H., Presting, G., Wang, R., Dunn, M., Glazebrook, J., Sessions, A., Oeller, P., Varma, H., Hadley, D., Hutchison, D., Martin, C., Katagiri, F., Lange, B. M., Moughamer, T., Xia, Y., Budworth, P., Zhong, J., Miguel, T., Paszkowski, U., Zhang, S., Colbert, M., Sun, W. L., Chen, L., Cooper, B., Park, S., Wood, T. C., Mao, L., Quail, P., Wing, R., Dean, R., Yu, Y., Zharkikh, A., Shen, R., Sahasrabudhe, S., Thomas, A., Cannings, R., Gutin, A., Pruss, D., Reid, J., Tavtigian, S., Mitchell, J., Eldredge, G., Scholl, T., Miller, R. M., Bhatnagar, S., Adey, N., Rubano, T., Tusneem, N., Robinson, R., Feldhaus, J., Macalma, T., Oliphant, A. and

Briggs, S. (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. japonica). *Science,* **296**(5565), 92-100.

Goldblatt, P. (1971) Cytological and morphological studies in the southern African Iridaceae. *Jl. S. Afr. Bot.,* **37**(1), 317-460.

Goldblatt, P., Manning, J. and Dunlop, G. (2004) *Crocosmia* and *Chasmanthe*. Royal Horticultural Society plant collector guide. UK: Timber Press.

Goldblatt, P., Walbot, V. and Zimmer, E. A. (1984) Estimation of genome size (C-value) in Iridaceae by cytophotometry. *Ann. Mo. Bot. Gard.*, **71**(1), 176-180.

Goldblatt, P., Rodriguez, A., Powell, M., Davies, J. T., Manning, J. C., Van der Bank, M. and Savolainen, V. (2008) Iridaceae "out of Australasia"? Phylogeny, biogeography, and divergence time based on plastid DNA sequences. *Syst. Bot.,* **33**(3), 495-508.

Góngora-Castillo, E. and Buell, C. R. (2013) Bioinformatics challenges in *de novo* transcriptome assembly using short read sequences in the absence of a reference genome sequence. *Nat. Prod. Rep.,* **30**(4), 490-500.

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B. W., Nusbaum, C., Lindblad-Toh, K., Friedman N. and Regev, A. (2011) Full-length transcriptome assembly from RNA-seq data without a reference genome. *Nat. Biotechnol.,* **29**(7), 644-652.

Greenhagen, B. T., O'Maille, P. E., Noel, J. P. and Chappell, J. (2006) Identifying and manipulating structural determinates linking catalytic specificities in terpene synthases. *Proc. Natl. Acad. Sci. U. S. A.,* **103**(26), 9826-9831.

Grégoire, J., Baisier, M., Drumont, A., Dahlsten, D., Meyer, H. and Francke, W. (1991) Volatile compounds in the larval frass of *Dendroctonus valens* and *Dendroctonus micans* (coleoptera: Scolytidae) in relation to oviposition by the predator, *Rhizophagus grandis* (coleoptera: Rhizophagidae). *J. Chem. Ecol.*, **17**(10), 2003-2019.

Grégoire, J., Couillien, D., Krebber, R., König, W. A., Meyer, H. and Francke, W. (1992) Orientation of *Rhizophagus grandis* (coleoptera: Rhizophagidae) to oxygenated monoterpenes in a species-specific predator-prey relationship. *Chemoecology,* **3**(1), 14-18.

Gross, G. G. and Zenk, M. H. (1974) Isolation and properties of hydroxycinnamate: CoA ligase from lignifying tissue of *Forsthia*. *Eur. J. Biochem.,* **42**(2), 453-459.

Gu, Y., Zhu, C., Iwamoto, H. and Chen, J. (2005) Genistein inhibits invasive potential of human hepatocellular carcinoma by altering cell cycle, apoptosis, and angiogenesis. *World J. Gastroenterol.,* **11**(41), 6512.

Gu, X., Glushka, J., Yin, Y., Xu, Y., Denny, T., Smith, J., Jiang, Y. and Bar-Peled, M. (2010) Identification of a bifunctional UDP-4-keto-pentose/UDP-xylose synthase in the plant pathogenic bacterium *Ralstonia solanacearum* strain GMI1000, a distinct member of the 4,6-dehydratase and decarboxylase family. *J. Biol. Chem.*, **285**(12), 9030-9040.

Gu, X., Lee, S. G. and Bar-Peled, M. (2011) Biosynthesis of UDP-xylose and UDP-arabinose in *Sinorhizobium meliloti* 1021: first characterization of a bacterial UDP-xylose synthase, and UDP-xylose 4-epimerase. *Microbiology*, **157**, 260-269.

Gurr, E. (1965) The rational use of dyes in biology. London: Leonard Hill.

Guyett, P., Glushka, J., Gu, X. and Bar-Peled, M. (2009) Real-time NMR monitoring of intermediates and labile products of the bifunctional enzyme UDP-apiose/UDP-xylose synthase. *Carbohydr. Res.*, **344**(9), 1072-1078.

Ha, M., Kwak, J. H., Kim, Y. and Zee, O. P. (2012) Direct analysis for the distribution of toxic glycoalkaloids in potato tuber tissue using matrix-assisted laser desorption/ionization mass spectrometric imaging. *Food Chem.*, **133**(4), 1155-1162.

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., Couger, M. B., Eccles, D., Li, B., Lieber, M., MacManes, M. D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C. N., Henschel, R., LeDuc, R. D., Friedman N. and Regev, A. (2013) *De novo* transcript sequence reconstruction from RNA-seq using the trinity platform for reference generation and analysis. *Nat. Prot.*, **8**(8), 1494-1512.

Hakamatsuka, T., Mori, K., Ishida, S., Ebizuka, Y. and Sankawa, U. (1998) Purification of 2-hydroxyisoflavanone dehydratase from the cell cultures of *Pueraria lobata* in honour of professor GH neil towers 75th birthday. *Phytochem.*, **49**(2), 497-505.

Hall, D. and De Luca, V. (2007) Mesocarp localization of a bi-functional resveratrol/hydroxycinnamic acid glucosyltransferase of Concord grape (*Vitis labrusca*). *Plant J.*, **49**(4), 579-591.

Hall, D., Yuan, X. X., Murata, J. and De Luca, V. (2012) Molecular cloning and biochemical characterization of the UDP-glucose: flavonoid 3-O-glucosyltransferase from Concord grape (*Vitis labrusca*). *Phytochemistry*, **74**(1), 90-99.

Hall, D. E., Robert, J. A., Keeling, C. I., Domanski, D., Quesada, A. L., Jancsik, S., Kuzyk, M. A., Hamberger, B., Borchers, C. H. and Bohlmann, J. (2011) An integrated genomic, proteomic and biochemical analysis of (+)-3-carene biosynthesis in Sitka spruce (*Picea sitchensis*) genotypes that are resistant or susceptible to white pine weevil. *Plant J.*, **65**(6), 936-948.

Hamberger, B., Hall, D., Yuen, M., Oddy, C., Hamberger, B., Keeling, C. I., Ritland, C., Ritland, K. and Bohlmann, J. (2009) Targeted isolation, sequence assembly and characterization of two white spruce (*Picea glauca*) BAC clones for terpenoid synthase and cytochrome P450 genes involved in conifer defence reveal insights into a conifer genome. *BMC Plant Biol.,* **9**, 106.

Han, X., Qian, L., Zhang, L. and Liu, X. (2015) Structural and biochemical insights into nucleotide-rhamnose synthase/epimerase-reductase from *Arabidopsis thaliana*. *Biochim. Biophys. Acta.,* **1854**(10), 1476-1486.

Han, S. H., Kim, B. G., Yoon, J. A., Chong, Y. and Ahn, J. H. (2014) Synthesis of flavonoid O-pentosides by *Escherichia coli* through engineering of nucleotide sugar pathways and glycosyltransferase. *Appl. Environ. Microbiol.,* **80**(9), 2754-2762.

Handford, M., Sicilia, F., Brandizzi, F., Chung, J. and Dupree, P. (2004) *Arabidopsis thaliana* expresses multiple Golgi-localised nucleotide-sugar transporters related to GONST1. *Mol. Genet. and Genomics,* **272**(4), 397-410.

Hansen, K. S., Kristensen, C., Tattersall, D. B., Jones, P. R., Olsen, C. E., Bak, S. and Møller, B. L. (2003) The *in vitro* substrate regiospecificity of recombinant UGT85B1, the cyanohydrin glucosyltransferase from *Sorghum bicolor*. *Phytochem.,* **64**(1), 143-151.

Harper, A. D. and Bar-Peled, M. (2002) Biosynthesis of UDP-xylose. Cloning and characterization of a novel *Arabidopsis* gene family, UXS, encoding soluble and putative membrane-bound UDP-glucuronic acid decarboxylase isoforms. *Plant Physiol.,* **130**(4), 2188-2198.

Hayashi, T. and Matsuda, K. (1981) Biosynthesis of xyloglucan in suspension-cultured soybean cells. Occurrence and some properties of xyloglucan 4-b-D-glucosyltransferaseand 6-b-D-xylosyltransferase. *J. Biol. Chem.* **256**, 11117-11122.

Hefner, J., Rubenstein, S. M., Ketchum, R. E., Gibson, D. M., Williams, R. M. and Croteau, R. (1996) Cytochrome P450-catalyzed hydroxylation of taxa-4(5), 11(12)-diene to taxa-4(20), 11(12)-dien-5a-o1: the first oxygenation step in taxol biosynthesis. *Chem. Biol.,* **3**(6), 479-489.

Heller, W. and Hahlbrock, K. (1980) Highly purified "flavanone synthase" from parsley catalyzes the formation of naringenin chalcone. *Arch. Biochem. Biophys.,* **200**(2), 617-619.

Hellsten, U., Wright, K. M., Jenkins, J., Shu, S., Yuan, Y., Wessler, S. R., Schmutz, J., Willis, J. H. and Rokhsar, D. S. (2013) Fine-scale variation in meiotic recombination in *Mimulus* inferred from population shotgun sequencing. *Proc. Natl. Acad. Sci. U. S. A.,* **110**(48), 19478-19482.

Hernandez, A. and Ruiz, M. T. (1998) An EXCEL template for calculation of enzyme kinetic parameters by non-linear regression. *Bioinformatics*, **14**, 227–228.

Hilker, M., Kobs, C., Varama, M. and Schrank, K. (2002) Insect egg deposition induces *Pinus sylvestris* to attract egg parasitoids. *J. Exp. Biol.*, **205**(Pt 4), 455-461.

Himejima, M., Hobson, K. R., Otsuka, T., Wood, D. L. and Kubo, I. (1992) Antimicrobial terpenes from oleoresin of ponderosa pine tree *Pinus ponderosa*: a defense mechanism against microbial invasion. *J. Chem. Ecol.*, **18**(10), 1809-1818.

Hiromoto, T., Honjo, E., Noda, N., Tamada, T., Kazuma, K., Suzuki, M., Blaber, M. and Kuroki, R. (2015) Structural basis for acceptor-substrate recognition of UDP-glucose: anthocyanidin 3-O-glucosyltransferase from *Clitoria ternatea*. *Protein Sci.*, **24**(3), 395-407.

Hofmann, K. and Stoffel, W. (1993) TMbase - a database of membrane spanning protein segments. *Biol. Chem. Hoppe Seyler*, **374**(1), 166.

Holton, R. A., Kim, H. B., Somoza, C., Liang, F., Biediger, R. J., Boatman, P. D., Shindo, M., Smith, C. C., Kim, S., Nadizadeh, H., Suzuki, Y., Tao, C., Vu, P., Tang, S., Zhang, P., Murthi, K. K., Gentile, L. N. and Liu, J. H. (1994) First total synthesis of taxol. 2. Completion of the C and D rings. *J. Am. Chem. Soc.*, **116**(4), 1599-1600.

Hong, B., Kim, J., Kim, N., Kim, B., Chong, Y. and Ahn, J. (2007) Characterization of uridine-diphosphate dependent flavonoid glucosyltransferase from *Oryza sativa*. *BMB Reports*, **40**(6), 870-874.

Hong, Y. J. and Tantillo, D. J. (2013) C–H··· π interactions as modulators of carbocation structure - implications for terpene biosynthesis. *Chem. Sci.*, **4**(6), 2512-2518.

Hong, Y. J. and Tantillo, D. J. (2011) The taxadiene-forming carbocation cascade. *J. Am. Chem. Soc.*, **133**(45), 18249-18256.

Hong, Y. J. and Tantillo, D. J. (2010) Quantum chemical dissection of the classic terpinyl/pinyl/bornyl/camphyl cation conundrum - the role of pyrophosphate in manipulating pathways to monoterpenes. *Org. Biomol. Chem.*, **8**(20), 4589-4600.

Hong, Y. J. and Tantillo, D. J. (2009) Consequences of conformational preorganization in sesquiterpene biosynthesis: theoretical studies on the formation of the bisabolene, curcumene, acoradiene, zizaene, cedrene, duprezianene, and sesquithuriferol sesquiterpenes. *J. Am. Chem. Soc.*, **131**(23), 7999-8015.

Hong, Y. J. and Tantillo, D. J. (2014) Biosynthetic consequences of multiple sequential post-transition-state bifurcations. *Nat. Chem.*, **6**(2), 104-111.

Hou, B., Lim, E. K., Higgins, G. S. and Bowles, D. J. (2004) N-glucosylation of cytokinins by glycosyltransferases of *Arabidopsis thaliana*. *J. Biol. Chem.*, **279**(46), 47822-47832.

Huang, J., Pang, C., Fan, S., Song, M., Yu, J., Wei, H., Ma, Q., Li, L., Zhang, C. and Yu, S. (2015) Genome-wide analysis of the family 1 glycosyltransferases in cotton. *Mol. Genet. Genomics*, **290**(5), 1805-1818.

Huang, S., Li, R., Zhang, Z., Li, L., Gu, X., Fan, W., Lucas, W. J., Wang, X., Xie, B., Ni, P., Ren, Y., Zhu, H., Li, J., Lin, K., Jin, W., Fei, Z., Li, G., Staub, J., Kilian, A., van der Vossen, E. A. G., Wu, Y., Guo, J., He, J., Jia, Z., Ren, Y., Tian, G., Lu, Y., Ruan, J., Qian, W., Wang, M., Huang, Q., Li, B., Xuan, Z., Cao, J., Zhigang, A., Zhang, J., Cai, Q., Bai, Y., Zhao, B., Han, Y., Li, Y., Li, X., Wang, S., Shi, Q., Liu, S., Cho, W. K., Kim, J. Y., Xu, Y., Heller-Uszynska, K., Miao, H., Cheng, Z., Zhang, S., Wu, J., Yang, Y., Kang, H., Li, M., Liang, H., Ren, X., Shi, Z., Wen, M., Jian, M., Yang, M., Zhang, G., Yang, Z., Chen, R., Liu, S., Li, J., Ma, L., Liu, H., Zhou, Y., Zhao, J., Fang, X., Li, G., Fang, L., Li, Y., Liu, D., Zheng, H., Zhang, Y., Qin, N., Li, Z., Yang, G., Yang, S., Bolund, L., Kristiansen, K., Zheng, H., Li, S., Zhang, S., Yang, H., Wang, J., Sun, R., Zhang, B., Jiang, S., Wang, J., Du1, Y. and Li, S. (2009) The genome of the cucumber, *Cucumis sativus* L. *Nat. Genet.*, **41**(12), 1275-1281.

Hughes, J. and Hughes, M. A. (1994) Multiple secondary plant product UDP-glucose glucosyltransferase genes expressed in cassava (*Manihot esculenta* crantz) cotyledons. *DNA Sequence*, **5**(1), 41-49.

Hyatt, D. C. and Croteau, R. (2005) Mutational analysis of a monoterpene synthase reaction: altered catalysis through directed mutagenesis of (α)-pinene synthase from *Abies grandis*. *Arch. Biochem. Biophys.*, **439**(2), 222-233.

Hyatt, D. C., Youn, B., Zhao, Y., Santhamma, B., Coates, R. M., Croteau, R. B. and Kang, C. H. (2007) Structure of limonene synthase, a simple model for terpenoid cyclase catalysis. *Proc. Natl. Acad. Sci. U.S.A.*, **104**(13), 5360-5365.

Jaillon, O., Aury, J. M, Noel, B., Policriti, A., Clepet, C., Casagrande, A., Choisne, N., Aubourg, S., Vitulo, N., Jubin, C., Vezzi, A., Legeai, F., Hugueney, P., Dasilva, C., Horner, D., Mica, E., Jublot, D., Poulain, J., Bruyère, C., Billault, A., Segurens, B., Gouyvenoux, M., Ugarte, E., Cattonaro, F., Anthouard, V., Vico1, V., Del Fabbro, C., Alaux, M., Di Gaspero, G., Dumas, V., Felice, N., Paillard, S., Juman, I., Moroldo, M., Scalabrin, S., Canaguier, A., Le Clainche, I., Malacrida, G., Durand, E., Pesole, G., Laucou, V., Chatelet, P., Merdinoglu, D., Delledonne, M., Pezzotti, M., Lecharny, A., Scarpelli, C., Artiguenave, F., Pè, M. E., Valle, G., Morgante, M., Caboche, M., Adam-Blondon, A. F., Weissenbach, J., Quétier, F. and Wincker, P. (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature*, **449**(7161), 463-467.

Jain, M., Srivastava, P. L., Verma, M., Ghangal, R. and Garg, R. (2016) *De novo* transcriptome assembly and comprehensive expression profiling in *Crocus sativus* to gain insights into apocarotenoid biosynthesis. *Sci. Rep.,* **6**, 22456.

James, L. C. and Tawfik, D. S. (2003) Conformational diversity and protein evolution - a 60-year-old hypothesis revisited. *Trends Biochem. Sci.,* **28**(7), 361-368.

Jennewein, S., Wildung, M. R., Chau, M., Walker, K. and Croteau, R. (2004) Random sequencing of an induced *Taxus* cell cDNA library for identification of clones involved in taxol biosynthesis. *Proc. Natl. Acad. Sci. U. S. A.,* **101**(24), 9149-9154.

Jensen, W. A. (1962) Botanical histochemistry. USA: Freedman.

Jensen, J. K., Sorensen, S. O., Harholt, J., Geshi, N., Sakuragi, Y., Moller, I., Zandleven, J., Bernal, A. J., Jensen, N. B., Sorensen, C., Pauly, M., Beldman, G., Willats, W. G. and Scheller, H. V. (2008) Identification of a xylogalacturonan xylosyltransferase involved in pectin biosynthesis in *Arabidopsis*. *Plant Cell,* **20**(5), 1289-1302.

Jez, J. M. and Noel, J. P. (2002) Reaction mechanism of chalcone isomerase. pH dependence, diffusion control, and product binding differences. *J. Biol. Chem.,* **277**(2), 1361-1369.

John, K. V., Schutzbach, J. S. and Ankel, H. (1977) Separation and allosteric properties of two forms of UDP-glucuronate carboxy-lyase. *J. Biol. Chem.,* **252**, 8013-8017.

Kampranis, S. C., Ioannidis, D., Purvis, A., Mahrez, W., Ninga, E., Katerelos, N. A., Anssour, S., Dunwell, J. M., Degenhardt, J., Makris, A. M., Goodenough, P. W. and Johnson, C. B. (2007) Rational conversion of substrate and product specificity in a salvia monoterpene synthase: structural insights into the evolution of terpene synthase function. *Plant Cell,* **19**(6), 1994-2005.

Kamsteeg, J., Brederode, J. V. and Nigtevecht, G. V. (1978) The formation of UDP-L-rhamnose from UDP-D-glucose by an enzyme preparation of red campion (*Silene dioica* (L) clairv) leaves. *FEBS Lett.,* **91**(2), 281-284.

Katoh, S., Hyatt, D. and Croteau, R. (2004) Altering product outcome in *Abies grandis* (−)-limonene synthase and (−)-limonene/(−)-α-pinene synthase by domain swapping and directed mutagenesis. *Arch. Biochem. Biophys.,* **425**(1), 65-76.

Keeling, C. I. and Bohlmann, J. (2006a) Diterpene resin acids in conifers. *Phytochemistry,* **67**(22), 2415-2423.

Keeling, C. I., Weisshaar, S., Ralph, S. G., Jancsik, S., Hamberger, B., Dullat, H. K. and Bohlmann, J. (2011) Transcriptome mining, functional characterization, and phylogeny of a large terpene synthase gene family in spruce (*Picea spp.*). *BMC Plant Biol.,* **11**, 43.

Keeling, C. I. and Bohlmann, J. (2006b) Genes, enzymes and chemicals of terpenoid diversity in the constitutive and induced defence of conifers against insects and pathogens. *New Phytol.,* **170**(4), 657-675.

Keeling, C. I., Weisshaar, S., Lin, R. P. C. and Bohlmann, J. (2008) Functional plasticity of paralogous diterpene synthases involved in conifer defense. *Proc. Nat. Acad. Sci.,* **105**(3), 1085-1090.

Khersonsky, O., Roodveldt, C. and Tawfik, D. S. (2006) Enzyme promiscuity: evolutionary and mechanistic aspects. *Curr. Opin. Chem. Biol.,* **10**(5), 498-508.

Khorolragchaa, A., Kim, Y., Rahimi, S., Sukweenadhi, J., Jang, M. and Yang, D. (2014) Grouping and characterization of putative glycosyltransferase genes from *Panax ginseng* Meyer. *Gene,* **536**(1), 186-192.

Kiefer, F., Arnold, K., Künzli, M., Bordoli, L. and Schwede, T. (2009) The SWISS-MODEL repository and associated resources. *Nucleic Acids Res.,* **37**(suppl 1), D387-D392.

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**(7), e103223.

Kim, B., Jung, W. D. and Ahn, J. (2013) Cloning and characterization of a putative UDP-rhamnose synthase 1 from *Populus euramericana Guinier*. *J. Plant Biol.,* **56**(1), 7-12.

Kim, B., Kim, H. J. and Ahn, J. (2012) Production of bioactive flavonol rhamnosides by expression of plant genes in *Escherichia coli*. *J. Agric. Food Chem.,* **60**(44), 11143-11148.

King, J. N., Alfaro, R. I. and Cartwright, C. (2004) Genetic resistance of Sitka spruce (*Picea sitchensis*) populations to the white pine weevil (*Pissodes strobi*): distribution of resistance. *Forestry,* **77**(4), 269-278.

King, J. N. and Alfaro, R. I. (2009) Developing sitka spruce populations for resistance to the white pine weevil: Summary of research and breeding program. *Technical Report - Ministry of Forests and Range, Forest Science Program, British Columbia,* (050).

Kliebenstein, D. J. and Osbourn, A. (2012) Making new molecules - evolution of pathways for novel metabolites in plants. *Curr. Opin. Plant Biol.,* **15**(4), 415-423.

Knaggs, A. R. (2003) The biosynthesis of shikimate metabolites. *Nat. Prod. Rep.,* **20**(1), 119-136.

Ko, J. H., Kim, B. G., Kim, J. H., Kim, H., Lim, C. E., Lim, J., Lee, C., Lim, Y. and Ahn, J. (2008) Four glucosyltransferases from rice: cDNA cloning, expression, and characterization. *J. Plant Physiol.,* **165**(4), 435-444.

Kojima, M. and Takeuchi, W. (1989) Detection and characterization of p-coumaric acid hydroxylase in mung bean, *Vigna mungo*, seedlings. *J. Biochem.,* **105**(2), 265-270.

Kolattukudy, P. E. (1980) Biopolyester membranes of plants: cutin and suberin. *Science,* **208**(4447), 990-1000.

Kolewe, M. E., Gaurav, V. and Roberts, S. C. (2008) Pharmaceutically active natural product synthesis and supply via plant cell culture technology. *Mol. Pharm.,* **5**(2), 243-256.

Kollner, T. G., Schnee, C., Gershenzon, J. and Degenhardt, J. (2004) The variability of sesquiterpenes emitted from two *Zea mays* cultivars is controlled by allelic variation of two terpene synthase genes encoding stereoselective multiple product enzymes. *Plant Cell,* **16**(5), 1115-1131.

Koonin, E. V., Fedorova, N. D., Jackson, J. D., Jacobs, A. R., Krylov, D. M., Makarova, K. S., Mazumder, R., Mekhedov, S. L., Nikolskaya, A. N., Rao, B. S., Rogozin, I. B., Smirnov, S., Sorokin, A. V., Sverdlov, A. V., Vasudevan, S., Wolf, Y. I., Yin J. J. and Natale, D. A. (2004) A comprehensive evolutionary classification of proteins encoded in complete eukaryotic genomes. *Genome Biol.,* **5**(2), R7.

Kopper, B. J., Illman, B. L., Kersten, P. J., Klepzig, K. D. and Raffa, K. F. (2005) Effects of diterpene acids on components of a conifer bark beetle–fungal interaction: tolerance by *Ips pini* and sensitivity by its associate *Ophiostoma ips*. *Environ. Entomol.,* **34**(2), 486-493.

Kostelijk, P. (1984) Crocosmia in gardens. *The Plantsman,* **5**, 246-253.

Kotake, T., Yamaguchi, D., Ohzono, H., Hojo, S., Kaneko, S., Ishida, H. K. and Tsumuraya, Y. (2004) UDP-sugar pyrophosphorylase with broad substrate specificity toward various monosaccharide 1-phosphates from pea sprouts. *J. Biol. Chem.,* **279**(44), 45728-45736.

Kovács, Z., Simon-Sarkadi, L., Szűcs, A. and Kocsy, G. (2010) Differential effects of cold, osmotic stress and abscisic acid on polyamine accumulation in wheat. *Amino Acids,* **38**(2), 623-631.

Kovinich, N., Saleem, A., Arnason, J. T. and Miki, B. (2010) Functional characterization of a UDP-glucose: flavonoid 3-O-glucosyltransferase from the seed coat of black soybean (*Glycine max* (L.) *merr.*). *Phytochem.,* **71**(11), 1253-1263.

Krause, S. T., Köllner, T. G., Asbach, J. and Degenhardt, J. (2013) Stereochemical mechanism of two sabinene hydrate synthases forming antipodal monoterpenes in thyme (*Thymus vulgaris*). *Arch. Biochem. Biophys.,* **529**(2), 112-121.

Krieger, E., Joo, K., Lee, J., Lee, J., Raman, S., Thompson, J., Tyka, M., Baker, D. and Karplus, K. (2009) Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: four approaches that performed well in CASP8. *Proteins: Struc., Funct., and Bioinf.,* **77**(S9), 114-122.

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. *J. Mol. Biol.,* **305**(3), 567-580.

Kuang, B., Zhao, X., Zhou, C., Zeng, W., Ren, J., Ebert, B., Beahan, C. T., Deng, X., Zeng, Q., Zhou, G., Doblin, M. S., Heazlewood, J. L., Bacic, A., Chen, X. and Wu., A-M. (2016) Role of UDP-Glucuronic Acid Decarboxylase in Xylan Biosynthesis in Arabidopsis. *Mol. Plant,* **9**(8), 1119-1131.

Kudo, T., Makita, N., Kojima, M., Tokunaga, H. and Sakakibara, H. (2012) Cytokinin activity of cis-zeatin and phenotypic alterations induced by overexpression of putative cis-zeatin-O-glucosyltransferase in rice. *Plant Physiol.,* **160**(1), 319-331.

Kumar, S., Stecher, G. and Tamura, K. (2016) MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.,* **33**(7), 1870-1874.

Lairson, L., Henrissat, B., Davies, G. and Withers, S. (2008) Glycosyltransferases: structures, functions, and mechanisms. *Biochem.,* **77**(1), 521-555.

Lamesch, P., Berardini, T. Z., Li, D., Swarbreck, D., Wilks, C., Sasidharan, R., Muller, R., Dreher, K., Alexander, D. L., Garcia-Hernandez, M., Karthikeyan, A. S., Lee, C. H., Nelson, W. D., Ploetz, L., Singh, S., Wensel, A. and Huala, E. (2012) The *Arabidopsis* information resource (TAIR): Improved gene annotation and new tools. *Nucleic Acids Res.,* **40**(D1), D1202-D1210.

Lang, V., Usadel, B. and Obermeyer, G. (2015) *De novo* sequencing and analysis of the lily pollen transcriptome: an open access data source for an orphan plant species. *Plant Mol. Biol.,* **87**(1), 69-80.

Lanot, A., Hodge, D., Jackson, R. G., George, G. L., Elias, L., Lim, E. K., Vaistij, F. E. and Bowles, D. J. (2006) The glucosyltransferase UGT72E2 is responsible for monolignol 4-O-glucoside production in *Arabidopsis thaliana*. *Plant J.,* **48**(2), 286-295.

Lao, J., Oikawa, A., Bromley, J. R., McInerney, P., Suttangkakul, A., Smith-Moritz, A. M., Plahar, H., Chiu, T., González Fernández-Niño, S. M., Berit Ebert, Yang, F., Christiansen, K. M., Hansen, S. F., Stonebloom, S., Adams, P. D., Ronald, P. C., Hillson, N. J., Hadi, M. Z., Vega-Sánchez, M. E., Loqué, D., Scheller, H. V. and

Heazlewood, J. L. (2014) The plant glycosyltransferase clone collection for functional genomics. *Plant J.,* **79**(3), 517-529.

Lawrence, G. H. M. (1964) Taxonomy of vascular plants. USA: Macmillan.

Leonarda, E., Ajikumara, P. K., Thayerb, K., Xiaoa, W-H., Moa, J. D., Tidorb, B., Stephanopoulosa, G., and Prathera, K. L. J. (2010). Combining metabolic and protein engineering of a terpenoid biosynthetic pathway for overproduction and selectivity control. *Proc. Natl. Acad. Sci.,* **107**(31), 13654–13659

Lesburg, C. A., Zhai, G., Cane, D. E. and Christianson, D. W. (1997) Crystal structure of pentalenene synthase: mechanistic insights on terpenoid cyclization reactions in biology. *Science,* **277**(5333), 1820-1824.

Letunic, I. and Bork, P. (2011) Interactive tree of life v2: online annotation and display of phylogenetic trees made easy. *Nucleic Acids Res.,* **39**(suppl_2), W475-W478.

Li, B., Knudsen, C., Hansen, N. K., Jørgensen, K., Kannangara, R., Bak, S., Takos, A., Rook, F., Hansen, S. H., Møller, B. L., Janfelt, C. and Bjarnholt, N. (2013) Visualizing metabolite distribution and enzymatic conversion in plant tissues by desorption electrospray ionization mass spectrometry imaging. *Plant J.,* **74**(6), 1059-1071.

Li, F., Fan, G., Wang, K., Sun, F., Yuan, Y., Song, G., Li, Q., Ma, Z., Lu, C., Zou, C., Chen, W., Liang, X., Shang, H., Liu, W., Shi, C., Xiao, G., Gou, C., Ye, W., Xu, X., Zhang, X., Wei, H., Li, Z., Zhang, G., Wang, J., Liu, K., Kohel, R. J., Percy, R. G., Yu, J. Z., Zhu, Y. X., Wang, J. and Yu, S. (2014a) Genome sequence of the cultivated cotton *Gossypium arboreum*. *Nat. Genet.,* **46**(6), 567-572.

Li, L., Modolo, L. V., Escamilla-Trevino, L. L., Achnine, L., Dixon, R. A. and Wang, X. (2007) Crystal structure of *Medicago truncatula* UGT85H2-insights into the structural basis of a multifunctional (iso) flavonoid glycosyltransferase. *J. Mol. Biol.,* **370**(5), 951-963.

Li, Y., Baldauf, S., Lim, E. K. and Bowles, D. J. (2001) Phylogenetic analysis of the UDP-glycosyltransferase multigene family of *Arabidopsis thaliana*. *J. Biol. Chem.,* **276**(6), 4338-4343.

Li, Y., Li, P., Wang, Y., Dong, R., Yu, H. and Hou, B. (2014b) Genome-wide identification and phylogenetic analysis of family-1 UDP glycosyltransferases in maize (*Zea mays*). *Planta,* **239**(6), 1265-1279.

Li, Y., Tao, W. and Cheng, L. (2009) Paclitaxel production using co-culture of *Taxus* suspension cells and paclitaxel-producing endophytic fungi in a co-bioreactor. *Appl. Microbiol. Biotechnol.,* **83**(2), 233-239.

Li, W. and Godzik, A. (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, **22**(13), 1658-1659.

Liang, D., Liu, J., Wu, H., Wang, B., Zhu, H. and Qiao, J. (2015) Glycosyltransferases: mechanisms and applications in natural product development. *Chem. Soc. Rev.*, **44**, 8350-8374.

Lichtenthaler, H. K. (1998) The plants' 1-deoxy-D-xylulose-5-phosphate pathway for biosynthesis of isoprenoids. *Eur. J. Lipid Sci. Technol.*, **100**(4-5), 128-138.

Lim, C. G., Fowler, Z. L., Hueller, T., Schaffer, S., and Koffas, M. A. G. (2011). High-Yield Resveratrol Production in Engineered *Escherichia coli*. *Appl. Environ. Microbiol.*, **77**(10), 3451-3460.

Lim, E. K., Ashford, D. A., Hou, B., Jackson, R. G. and Bowles, D. J. (2004) *Arabidopsis* glycosyltransferases as biocatalysts in fermentation for regioselective synthesis of diverse quercetin glucosides. *Biotechnol. Bioeng.*, **87**(5), 623-631.

Lim, E. K., Doucet, C. J., Li, Y., Elias, L., Worrall, D., Spencer, S. P., Ross, J. and Bowles, D. J. (2002) The activity of *Arabidopsis* glycosyltransferases toward salicylic acid, 4-hydroxybenzoic acid, and other benzoates. *J. Biol. Chem.*, **277**(1), 586-592.

Lim, E. K., Baldauf, S., Li, Y., Elias, L., Worrall, D., Spencer, S. P., Jackson, R. G., Taguchi, G., Ross, J. and Bowles, D. J. (2003) Evolution of substrate recognition across a multigene family of glycosyltransferases in *Arabidopsis*. *Glycobiology*, **13**(3), 139-145.

Liscombe, D. K., Ziegler, J., Schmidt, J., Ammer, C. and Facchini, P. J. (2009) Targeted metabolite and transcript profiling for elucidating enzyme function: isolation of novel n-methyltransferases from three benzylisoquinoline alkaloid-producing species. *Plant J.*, **60**(4), 729-743.

Liu, H. and Thorson, J. S. (1994) Pathways and mechanisms in the biogenesis of novel deoxysugars by bacteria. *Annu. Rev. Microbiol.*, **48**(1), 223-256.

Liu, T., Zhu, S., Tang, Q., Chen, P., Yu, Y. and Tang, S. (2013) *De novo* assembly and characterization of transcriptome using Illumina paired-end sequencing and identification of CesA gene in ramie (*Boehmeria nivea L. gaud*). *BMC Genomics*, **14**, 125.

Liu, Z., Carpenter, S. B., Bourgeois, W. J., Yu, Y., Constantin, R. J., Falcon, M. J. and Adams, J. C. (1998) Variations in the secondary metabolite camptothecin in relation to tissue age and season in *Camptotheca acuminata*. *Tree Physiol.*, **18**(4), 265-270.

Logacheva, M. D., Kasianov, A. S., Vinogradov, D. V., Samigullin, T. H., Gelfand, M. S., Makeev, V. J. and Penin, A. A. (2011) *De novo* sequencing and characterization of floral transcriptome in two species of buckwheat (*Fagopyrum*). *BMC Genomics*, **12**, 30.

Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M. and Henrissat, B. (2014) The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.,* **42**, 490-495.

Long, R. M., Lagisetti, C., Coates, R. M. and Croteau, R. B. (2008) Specificity of the N-benzoyl transferase responsible for the last step of taxol biosynthesis. *Arch. Biochem. Biophys.,* **477**(2), 384-389.

Lukačin, R., Wellmann, F., Britsch, L., Martens, S. and Matern, U. (2003) Flavonol synthase from *Citrus unshiu* is a bifunctional dioxygenase. *Phytochem.,* **62**(3), 287-292.

Ma, F., Cholewa, E., Mohamed, T., Peterson, C. A. and Gijzen, M. (2004) Cracks in the palisade cuticle of soybean seed coats correlate with their permeability to water. *Ann. Bot.,* **94**(2), 213-228.

Mageroy, M. H., Parent, G., Germanos, G., Giguere, I., Delvas, N., Maaroufi, H., Bauce, É, Bohlmann, J. and Mackay, J. J. (2015) Expression of the β-glucosidase gene *Pgβglu*-1 underpins natural resistance of white spruce against spruce budworm. *Plant J.,* **81**(1), 68-80.

Maher, C. A., Palanisamy, N., Brenner, J. C., Cao, X., Kalyana-Sundaram, S., Luo, S., Khrebtukova, I., Barrette, T. R., Grasso, C., Yu, J., Lonigro, R. J., Schroth, G., Kumar-Sinha, C. and Chinnaiyan, A. M. (2009) Chimeric transcript discovery by paired-end transcriptome sequencing. *Proc. Natl. Acad. Sci. U. S. A.,* **106**(30), 12353-12358.

Malik, S., Cusidó, R. M., Mirjalili, M. H., Moyano, E., Palazón, J. and Bonfill, M. (2011) Production of the anticancer drug taxol in *Taxus baccata* suspension cultures: A review. *Process Biochem.,* **46**(1), 23-34.

Manning, J. and Goldblatt, P. (2008) The iris family: natural history and classification. USA: Timber

Martens, S., Forkmann, G., Matern, U. and Lukačin, R. (2001) Cloning of parsley flavone synthase I. *Phytochem.,* **58**(1), 43-46.

Martin, D. M., Gershenzon, J. and Bohlmann, J. (2003) Induction of volatile terpene biosynthesis and diurnal emission by methyl jasmonate in foliage of Norway spruce. *Plant Physiol.,* **132**(3), 1586-1599.

Martin, D. M., Fäldt, J. and Bohlmann, J. (2004) Functional characterization of nine Norway spruce TPS genes and evolution of gymnosperm terpene synthases of the TPS-d subfamily. *Plant Physiol.,* **135**(4), 1908-1927.

Martinez, V., Ingwers, M., Smith, J., Glushka, J., Yang, T. and Bar-Peled, M. (2012) Biosynthesis of UDP-4-keto-6-deoxyglucose and UDP-rhamnose in pathogenic fungi *Magnaporthe grisea* and *Botryotinia fuckeliana*. *J. Biol. Chem.,* **287**(2), 879-892.

Martinoia, E., Klein, M., Geisler, M., Sánchez-Fernández, R. and Rea, P. (2000) Vacuolar transport of secondary metabolites and xenobiotics. *Annual Plant Reviews,* **5**, 221-253.

Masada, S., Terasaka, K., Oguchi, Y., Okazaki, S., Mizushima, T. and Mizukami, H. (2009) Functional and structural characterization of a flavonoid glucoside 1,6-glucosyltransferase from *Catharanthus roseus*. *Plant Cell Physiol.,* **50**(8), 1401-1415.

Masuda, K., Funayama, S., Komiyama, K., Umezawa, I. and Ito, K. (1987) Constituents of *Tritonia crocosmaeflora*, II. tricrozarin B, an antitumor naphthazarin derivative. *J. Nat. Prod.,* **50**(5), 958-960.

Mathers, C. D. and Loncar, D. (2006) Projections of global mortality and burden of disease from 2002 to 2030. *PLos Med,* **3**(11), e442.

Mattocks A. R. (1986) Chemistry and Toxicology of Pyrrolizidine Alkaloids. UK: Academic.

Meng, H., Wang, Y., Hua, Q., Zhang, S. and Wang, X. (2011) *In silico* analysis and experimental improvement of taxadiene heterologous biosynthesis in *Escherichia coli*. *Biotechnol. Bioprocess Eng.,* **16**(2), 205-215.

Menhard, B. and Zenk, M. H. (1999) Special publication purification and characterization of acetyl coenzyme a: 10-hydroxytaxane O-acetyltransferase from cell suspension cultures of *Taxus chinensis*. *Phytochem.,* **50**(5), 763-774.

Menting, J., Scopes, R. K. and Stevenson, T. W. (1994) Characterization of flavonoid 3',5'-hydroxylase in microsomal membrane fraction of *Petunia hybrida* flowers. *Plant Physiol.,* **106**(2), 633-642.

Mercke, P., Bengtsson, M., Bouwmeester, H. J., Posthumus, M. A. and Brodelius, P. E. (2000) Molecular cloning, expression, and characterization of amorpha-4, 11-diene synthase, a key enzyme of artemisinin biosynthesis in *Artemisia annua* L. *Arch. Biochem. Biophys.,* **381**(2), 173-180.

Michael, T. P., Mockler, T. C., Breton, G., McEntee, C., Byer, A., Trout, J. D., Hazen, S. P., Shen, R., Priest, H. D., Sullivan, C. M., Givan, S. A., Yanovsky, M., Hong, F., Kay, S. A. and Chory, J. (2008) Network discovery pipeline elucidates conserved time-of-day–specific cis-regulatory modules. *PLoS Genet.,* **4**(2), e14.

Middleton Jr., E. (1998) Effect of plant flavonoids on immune and inflammatory cell function. *Adv. Exp. Med. Biol*., **439**(1), 175-182.

Miller, B., Madilao, L. L., Ralph, S. and Bohlmann, J. (2005) Insect-induced conifer defense. White pine weevil and methyl jasmonate induce traumatic resinosis, *de novo* formed volatile emissions, and accumulation of terpenoid synthase and putative octadecanoid pathway transcripts in Sitka spruce. *Plant Physiol.,* **137**(1), 369-382.

Mockler, T. C., Michael, T. P., Priest, H. D., Shen, R., Sullivan, C. M., Givan, S. A., McEntee, C., Kay, S. A. and Chory, J. (2007) The DIURNAL project: DIURNAL and circadian expression profiling, model-based pattern matching, and promoter analysis. *Cold Spring Harb. Symp. Quant. Biol.,* **72**(1), 353-363.

Modolo, L. V., Li, L., Pan, H., Blount, J. W., Dixon, R. A. and Wang, X. (2009) Crystal structures of glycosyltransferase UGT78G1 reveal the molecular basis for glycosylation and deglycosylation of (iso) flavonoids. *J. Mol. Biol.,* **392**(5), 1292-1302.

Modolo, L. V., Blount, J. W., Achnine, L., Naoumkina, M. A., Wang, X. and Dixon, R. A. (2007) A functional genomics approach to (iso) flavonoid glycosylation in the model legume *Medicago truncatula*. *Plant Mol. Biol.,* **64**(5), 499-518.

Moraga, Á. R., Mozos, A. T., Ahrazem, O. and Gómez-Gómez, L. (2009) Cloning and characterization of a glucosyltransferase from *Crocus sativus* stigmas involved in flavonoid glucosylation. *BMC Plant Biol.,* **9**, 109.

Mora-Pale, M., Sanchez-Rodriguez, S. P., Linhardt, R. J., Dordick, J. S., and Koffas M. A. G. (2013). Metabolic engineering and in vitro biosynthesis of phytochemicals and non-natural analogues. Plant Sci., **210**, 10-24.

Moustafa, E. and Wong, E. (1967) Purification and properties of chalcone-flavanone isomerase from soya bean seed. *Phytochem.,* **6**(5), 625-632.

Mudalkar, S., Golla, R., Ghatty, S. and Reddy, A. R. (2014) *De novo* transcriptome analysis of an imminent biofuel crop, *Camelina sativa L.* using Illumina GAIIX sequencing platform and identification of SSR markers. *Plant Mol. Biol.,* **84**(1), 159-171.

Mumm, R. and Hilker, M. (2005) The significance of background odour for an egg parasitoid to detect plants with host eggs. *Chem. Senses,* **30**(4), 337-343.

Nagamoto, N., Noguchi, H., Itokawa, A., Nakata, K., Namba, K., Nishimura, H., Matsui, M. and Mizuno, M. (1988) Antitumor constituents from bulbs of *Crocosmia x crocosmiiflora*. *Planta Med.,* **54**(04), 305-307.

Nakagawa, A., Minami, H., Kim, J-S., Koyanagi, T., Katayama, T., Sato, F., Kumagai, H. (2010) A bacterial platform for fermentative production of plant alkaloids. *Nat. Commun.,* **2**:326, 1-8.

Nakayama, T., Yonekura-Sakakibara, K., Sato, T., Kikuchi, S., Fukui, Y., Fukuchi-Mizutani, M., Ueda, T., Nakao, M., Tanaka, Y., Kusumi, T. and Nishino, T. (2000) Aureusidin synthase: A polyphenol oxidase homolog responsible for flower coloration. *Science,* **290**(5494), 1163-1166.

Nathan, D. M., Buse, J. B., Davidson, M. B., Ferrannini, E., Holman, R. R., Sherwin, R., Zinman, B., American Diabetes Association and European Association for Study of Diabetes (2009) Medical management of hyperglycemia in type 2 diabetes: a consensus algorithm for the initiation and adjustment of therapy. *Diabetes Care,* **32**(1), 193-203.

Newman, J. D., Marshall, J., Chang, M., Nowroozi, F., Paradise, E., Pitera, D., Newman, K. L. and Keasling, J. D. (2006) High-level production of amorpha-4, 11-diene in a two-phase partitioning bioreactor of metabolically engineered *Escherichia coli*. *Biotechnol. Bioeng.,* **95**(4), 684-691.

Nicolaou, K., Yang, Z., Liu, J. J., Ueno, H., Nantermet, P., Guy, R., Claiborne, C., Renaud, J., Couladouros, E., Paulvannan, K. and Sorensen, E. J. (1994) Total synthesis of taxol. *Nature,* **367**(6464), 630-634.

Noel, J. P., Dellas, N., Faraldos, J. A., Zhao, M., Hess Jr, B. A., Smentek, L., Coates, R. M. and O'Maille, P. E. (2010) Structural elucidation of cisoid and transoid cyclization pathways of a sesquiterpene synthase using 2-fluorofarnesyl diphosphates. *ACS Chem. Biol.,* **5**(4), 377-392.

Norambuena, L., Nilo, R., Handford, M., Reyes, F., Marchant, L., Meisel, L. and Orellana, A. (2005) *At*UTr2 is an *Arabidopsis thaliana* nucleotide sugar transporter located in the Golgi apparatus capable of transporting UDP-galactose. *Planta,* **222**(3), 521-529.

Norambuena, L., Marchant, L., Berninsone, P., Hirschberg, C. B., Silva, H. and Orellana, A. (2002) Transport of UDP-galactose in plants. Identification and functional characterization of *At*UTr1, an *Arabidopsis thaliana* UDP-galactos/UDP-glucose transporter. *J. Biol. Chem.,* **277**(36), 32923-32929.

Northcote, D. (1963) Changes in the cell walls of plants during differentiation. *Symp. Soc. Exp. Biol.,* **17**(1), 157-174.

Nützmann, H. and Osbourn, A. (2014) Gene clustering in plant specialized metabolism. *Curr. Opin. Biotechnol.,* **26**, 91-99.

O'Byrne, K. J. and Dalgleish, A. G. (2001) Chronic immune activation and inflammation as the cause of malignancy. *Br. J. Cancer,* **85**(4), 473-483.

Obembe, O. O., Popoola, J. O., Leelavathi, S. and Reddy, S. V. (2011) Advances in plant molecular farming. *Biotechnol. Adv.,* **29**(2), 210-222.

Offen, W., Martinez-Fleites, C., Yang, M., Kiat-Lim, E., Davis, B. G., Tarling, C. A., Ford, C. M., Bowles, D. J. and Davies, G. J. (2006) Structure of a flavonoid glucosyltransferase reveals the basis for plant natural product modification. *EMBO J.*, **25**(6), 1396-1405.

Ogura, K. and Koyama, T. (1998) Enzymatic aspects of isoprenoid chain elongation. *Chem. Rev.*, **98**(4), 1263-1276.

Oka, T. and Jigami, Y. (2006) Reconstruction of de novo pathway for synthesis of UDP-glucuronic acid and UDP-xylose from intrinsic UDP-glucose in *Saccharomyces cerevisiae*. *FEBS J.* **273**(12), 2645-2657.

Oka, T. and Jigami, Y. (2007) Method for producing UDP-rhamnose and enzyme used for the method. *US Patent 20080064069*.

Oka, T., Nemoto, T. and Jigami, Y. (2007) Functional analysis of *Arabidopsis thaliana* RHM2/MUM4, a multidomain protein involved in UDP-D-glucose to UDP-L-rhamnose conversion. *J. Biol. Chem.*, **282**(8), 5389-5403.

O'Maille, P. E., Chappell, J. and Noel, J. P. (2004) A single-vial analytical and quantitative gas chromatography-mass spectrometry assay for terpene synthases. *Anal. Biochem.*, **335**(2), 210-217.

Paddon, C. J. and Keasling, J. D. (2014) Semi-synthetic artemisinin: a model for the use of synthetic biology in pharmaceutical development. *Nature Rev. Microbiol.*, **12**(5), 355-367.

Paddon, C. J., Westfall, P. J., Pitera, D., Benjamin, K., Fisher, K., McPhee, D., Leavell, M., Tai, A., Main, A., Eng, D., Polichuk, D. R., Teoh, K. H., Reed, D. W., Treynor, T., Lenihan, J., Jiang, H., Fleck, M., Bajad, S., Dang, G., Dengrove, D., Diola, D., Dorin, G., Ellens, K. W., Fickes, S., Galazzo, J., Gaucher, S. P., Geistlinger, T., Henry, R., Hepp, M., Horning, T., Iqbal, T., Kizer, L., Lieu, B., Melis, D., Moss, N., Regentin, R., Secrest, S., Tsuruta, H., Vazquez, R., Westblade, L. F., Xu, L., Yu, M., Zhang, Y., Zhao, L., Lievense, J., Covello, P. S., Keasling, J. D., Reiling, K. K., Renninger, N. S. and Newman, J. D. (2013) High-level semi-synthetic production of the potent antimalarial artemisinin. *Nature*, **496**(7446), 528-532.

Paine, T. and Hanlon, C. (1994) Influence of oleoresin constituents from *Pinus ponderosa* and *Pinus jeffreyi* on growth of mycangial fungi from *Dendroctonus ponderosae* and *Dendroctonus jeffreyi*. *J. Chem. Ecol.*, **20**(10), 2551-2563.

Paine, T., Raffa, K. and Harrington, T. (1997) Interactions among scolytid bark beetles, their associated fungi, and live host conifers. *Annu. Rev. Entomol.*, **42**(1), 179-206.

Pan, Y., Wang, X., Liu, H., Zhang, G. and Ma, Z. (2010) Molecular cloning of three UDP-glucuronate decarboxylase genes that are preferentially expressed in *Gossypium* fibers from elongation to secondary cell wall synthesis. *J. Plant Biol.*, **53**(5), 367-373.

Pandey, R. P., Malla, S., Simkhada, D., Kim, B. and Sohng, J. K. (2013) Production of 3-O-xylosyl quercetin in *Escherichia coli*. *Appl. Microbiol. Biotechnol.*, **97**(5), 1889-1901.

Parra, G., Bradnam, K. and Korf, I. (2007) CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics*, **23**(9), 1061-1067.

Paterson, A. H., Bowers, J. E., Bruggmann, R., Dubchak, I., Grimwood, J., Gundlach, H., Haberer, G., Hellsten, U., Mitros, T., Poliakov, A., Schmutz, J., Spannagl, M., Tang, H., Wang, X., Wicker, T., Bharti, A. K., Chapman, J., Feltus, F. A., Gowik, U., Grigoriev, I. V., Lyons, E., Maher, C. A., Martis, M., Narechania, A., Otillar, R. P., Penning, B. W., Salamov, A. A., Wang, Y., Zhang, L., Carpita, N. C., Freeling, M., Gingle, A. R., Hash, C. T., Keller, B., Klein, P., Kresovich, S., McCann, M. C., Ming, R., Peterson, D. G., Rahman, M., Ware, D., Westhoff, P., Mayer, K. F. X., Messing, J. and Rokhsar, D. S. (2009) The *Sorghum bicolor* genome and the diversification of grasses. *Nature*, **457**(7229), 551-556.

Patro, R., Mount, S. M. and Kingsford, C. (2014) Sailfish enables alignment-free isoform quantification from RNA-seq reads using lightweight algorithms. *Nat. Biotechnol.*, **32**(5), 462-464.

Pattathil, S., Harper, A. D. and Bar-Peled, M. (2005) Biosynthesis of UDP-xylose: characterization of membrane-bound *At*Uxs2. *Planta*, **221**(4), 538-548.

Pemberton, R. P. and Tantillo, D. J. (2014) Lifetimes of carbocations encountered along reaction coordinates for terpene formation. *Chem. Sci.*, **5**(8), 3301-3308.

Pettersson, E. M. (2001) Volatiles from potential hosts of *Rhopalicus tutela* a bark beetle parasitoid. *J. Chem. Ecol.*, **27**(11), 2219-2231.

Phillips, M. A. and Croteau, R. B. (1999) Resin-based defenses in conifers. *Trends Plant Sci.*, **4**(5), 184-190.

Pichersky, E. and Lewinsohn, E. (2011) Convergent evolution in plant specialized metabolism. *Annu. Rev. Plant Biol.*, **62**, 549-566.

Poppenberger, B., Fujioka, S., Soeno, K., George, G. L., Vaistij, F. E., Hiranuma, S., Seto, H., Takatsuto, S., Adam, G., Yoshida, S. and Bowles, D. (2005) The UGT73C5 of *Arabidopsis thaliana* glucosylates brassinosteroids. *Proc. Natl. Acad. Sci. U. S. A.*, **102**(42), 15253-15258.

Raffa, K. F. and Klepzig, K. D. (1989) Chiral escape of bark beetles from predators responding to a bark beetle pheromone. *Oecologia*, **80**(4), 566-569.

Ramos-Valdivia, A. C., van der Heijden, R. and Verpoorte, R. (1997) Isopentenyl diphosphate isomerase: A core enzyme in isoprenoid biosynthesis. A review of its biochemistry and function. *Nat. Prod. Rep.*, **14**(6), 591-603.

Reeves, P. R., Hobbs, M., Valvano, M. A., Skurnik, M., Whitfield, C., Coplin, D., Kido, N., Klena, J., Maskell, D. and Raetz, C. R. (1996) Bacterial polysaccharide synthesis and gene nomenclature. *Trends Microbiol.*, **4**(12), 495-503.

Richter, M., Ebermann, R. and Marian, B. (1999) Quercetin-induced apoptosis in colorectal tumor cells: possible role of EGF receptor signaling. *Nutr. Cancer*, **34**(1), 88-99.

Ridley, B. L., O'Neill, M. A. and Mohnen, D. (2001) Pectins: structure, biosynthesis, and oligogalacturonide-related signaling. *Phytochem.*, **57**(6), 929-967.

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W. and Smyth, G. K. (2015) Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.*, **43**(7), e47.

Ro, D., Paradise, E. M., Ouellet, M., Fisher, K. J., Newman, K. L., Ndungu, J. M., Ho, K. A., Eachus, R. A., Ham, T. S., Kirby, J., Chang, M. C. Y., Withers, S. T., Shiba, Y., Sarpong, R. and Keasling, J. D. (2006) Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature*, **440**(7086), 940-943.

Roach, C. R., Hall, D. E., Zerbe, P. and Bohlmann, J. (2014) Plasticity and evolution of (+)-3-carene synthase and (–)-sabinene synthase functions of a Sitka spruce monoterpene synthase gene family associated with weevil resistance. *J. Biol. Chem.*, **289**(34), 23859-23869.

Robert, J. A., Madilao, L. L., White, R., Yanchuk, A., King, J. and Bohlmann, J. (2010) Terpenoid metabolite profiling in Sitka spruce identifies association of dehydroabietic acid, (+)-3-carene, and terpinolene with resistance against white pine weevil. *Botany*, **88**(9), 810-820.

Robinson, M. D., McCarthy, D. J. and Smyth, G. K. (2010) edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, **26**(1), 139-140.

Rollwitz, I., Santaella, M., Hille, D., Flügge, U. and Fischer, K. (2006) Characterization of *At*NST-KT1, a novel UDP-galactose transporter from *Arabidopsis thaliana*. *FEBS Lett.*, **580**(17), 4246-4251.

Rossmann, M. G. and Argos, P. (1978) The taxonomy of binding sites in proteins. *Mol. Cell. Biochem.*, **21**(3), 161-182.

Russell, D. W. and Conn, E. E. (1967) The cinnamic acid 4-hydraxylase of pea seedlings. *Arch. Biochem. Biophys.*, **122**(1), 256-258.

Saito, K., Hirai, M. Y. and Yonekura-Sakakibara, K. (2008) Decoding genes with coexpression networks and metabolomics - "majority report by precogs". *Trends Plant Sci.*, **13**(1), 36-43.

Saito, K., Kobayashi, M., Gong, Z., Tanaka, Y. and Yamazaki, M. (1999) Direct evidence for anthocyanidin synthase as a 2-oxoglutarate-dependent oxygenase: molecular cloning and functional expression of cDNA from a red forma of *Perilla frutescens*. *Plant J.*, **17**(2), 181-189.

Sakakibara, H. (2006) Cytokinins: activity, biosynthesis, and translocation. *Annu. Rev. Plant Biol.*, **57**, 431-449.

Sarkar, F. H. and Li, Y. (2004) The role of isoflavones in cancer chemoprevention. *Front. Biosci.*, **9**(1), 2714-2724.

Schaefer, H., Hardy, O. J., Silva, L., Barraclough, T. G. and Savolainen, V. (2011) Testing darwin's naturalization hypothesis in the azores. *Ecol. Lett.*, **14**(4), 389-396.

Schmeller, T., El-Shazly, A. and Wink, M. (1997) Allelochemical activities of pyrrolizidine alkaloids: interactions with neuroreceptors and acetylcholine related enzymes. *J. Chem. Ecol.*, **23**(2), 399-416.

Schmutz, J., Cannon, S. B., Schlueter, J., Ma, J., Mitros, T., Nelson, W., Hyten, W. L., Song, Q., Thelen, J. J., Cheng, J., Xu, D., Hellsten, U., May, G. D., Yu, Y., Sakurai, T., Umezawa, T., Bhattacharyya, M. K., Sandhu, D., Valliyodan, B., Lindquist, E., Peto, M., Grant, D., Shu, S., Goodstein, D., Barry, K., Futrell-Griggs, M., Abernathy, B., Du, J., Tian, Z., Zhu, L., Gill, N., Joshi, T., Libault, M., Sethuraman, A., Zhang, Z., Shinozaki, K., Nguyen, H. T., Wing, R. A., Cregan, P., Specht, J., Grimwood, J., Rokhsar, D., Stacey, G., Shoemaker, R. C. and Jackson, S. A. (2010) Genome sequence of the palaeopolyploid soybean. *Nature*, **463**(7278), 178-183.

Schnable, P. S., Ware, D., Fulton, R. S., Stein, J. C., Wei, F., Pasternak, S., Liang, C., Zhang, J., Fulton, L., Graves, T. A., Minx, P., Reily, A. D., Courtney, L., Kruchowski, S. S., Tomlinson, C., Strong, C., Delehaunty, K., Fronick, C., Courtney, B., Rock, S. M., Belter, E., Du, F., Kim, K., Abbott, R. M., Cotton, M., Levy, A., Marchetto, P., Ochoa, K., Jackson, S. M., Gillam, B., Chen, W., Yan, L., Higginbotham, J., Cardenas, M., Waligorski, J., Applebaum, E., Phelps, L., Falcone, J., Kanchi, K., Thane, T., Scimone, A., Thane, N., Henke, J., Wang, T., Ruppert, J., Shah, N., Rotter, K., Hodges, J., Ingenthron, E., Cordes, M., Kohlberg, S., Sgro, J., Delgado, B., Mead, K., Chinwalla, A., Leonard, S., Crouse, K., Collura, K., Kudrna, D., Currie, J., He, R., Angelova, A., Rajasekar, S., Mueller, T., Lomeli, R., Scara, G., Ko, A., Delaney, K., Wissotski, M., Lopez, G., Campos, D., Braidotti, M., Ashley, E., Golser, W., Kim, H., Lee, S., Lin, J., Dujmic, Z., Kim, W., Talag, J., Zuccolo, A., Fan, C., Sebastian,

A., Kramer, M., Spiegel, L., Nascimento, L., Zutavern, T., Miller, B., Ambroise, C., Muller, S., Spooner, W., Narechania, A., Ren, L., Wei, S., Kumari, S., Faga, B., Levy, M. J., McMahan, L., Van Buren, P., Vaughn, M. W., Ying, K., Yeh, C. T., Emrich, S. J., Jia, Y., Kalyanaraman, A., Hsia, A. P., Barbazuk, W. B., Baucom, R. S., Brutnell, T. P., Carpita, N. C., Chaparro, C., Chia, J. M., Deragon, J. M., Estill, J. C., Fu, Y., Jeddeloh, J. A., Han, Y., Lee, H., Li, P., Lisch, D. R., Liu, S., Liu, Z., Nagel, D. H., McCann, M. C., SanMiguel, P., Myers, A. M., Nettleton, D., Nguyen, J., Penning, B. W., Ponnala, L., Schneider, K. L., Schwartz, D. C., Sharma, A., Soderlund, C., Springer, N. M., Sun, Q., Wang, H., Waterman, M., Westerman, R., Wolfgruber, T. K., Yang, L., Yu, Y., Zhang, L., Zhou, S., Zhu, Q., Bennetzen, J. L., Dawe, R. K., Jiang, J., Jiang, N., Presting, G. G., Wessler, S. R., Aluru, S., Martienssen, R. A., Clifton, S. W., McCombie, W. R., Wing, R. A. and Wilson, R. K. (2009) The B73 maize genome: complexity, diversity, and dynamics. *Science*, **326**(5956), 1112-1115.

Schoendorf, A., Rithner, C. D., Williams, R. M. and Croteau, R. B. (2001) Molecular cloning of a cytochrome P450 taxane 10 beta-hydroxylase cDNA from *Taxus* and functional expression in yeast. *Proc. Natl. Acad. Sci. U. S. A.*, **98**(4), 1501-1506.

Schreck, R., Rieber, P. and Baeuerle, P. A. (1991) Reactive oxygen intermediates as apparently widely used messengers in the activation of the NF-kappa B transcription factor and HIV-1. *EMBO J.*, **10**(8), 2247-2258.

Schuttelkopf, A. W. and Van Aalten, D. M. (2004) PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta Cryst. D*, **60**(8), 1355-1363.

Seifert, G. J. (2004) Nucleotide sugar interconversions and cell wall biosynthesis: how to bring the inside to the outside. *Curr. Opin. Plant Biol.*, **7**(3), 277-284.

Selway, J. W. (1986) Antiviral activity of flavones and flavans. *Prog. Clin. Biol. Res.*, **213**, 521-536.

Shao, H., He, X., Achnine, L., Blount, J. W., Dixon, R. A. and Wang, X. (2005) Crystal structures of a multifunctional triterpene/flavonoid glycosyltransferase from *Medicago truncatula*. *Plant Cell*, **17**(11), 3141-3154.

Sigala, P. A., Kraut, D. A., Caaveiro, J. M., Pybus, B., Ruben, E. A., Ringe, D., Petsko, G. A. and Herschlag, D. (2008) Testing geometrical discrimination within an enzyme active site: constrained hydrogen bonding in the ketosteroid isomerase oxyanion hole. *J. Am. Chem. Soc.*, **130**(41), 13696-13708.

Simao, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. and Zdobnov, E. M. (2015) BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, **31**(19), 3210-3212.

Smith, D. A. and Banks, S. W. (1986) Biosynthesis, elicitation and biological activity of isoflavonoid phytoalexins. *Phytochem.*, **25**(5), 979-995.

Song, C., Gu, L., Liu, J., Zhao, S., Hong, X., Schulenburg, K. and Schwab, W. (2015) Functional characterization and substrate promiscuity of UGT71 glycosyltransferases from strawberry (*Fragaria x ananassa*). *Plant Cell Physiol.*, **56**(12), 2478-2493.

Sonnhammer, E. L., Von Heijne, G. and Krogh, A. (1998) A hidden Markov model for predicting transmembrane helices in protein sequences. *Proc. Int. Conf. Intell. Syst. Mol. Biol.*, **6**(1), 175-182.

Souza-Chies, T. T., Bittar, G., Nadot, S., Carter, L., Besin, E. and Lejeune, B. (1997) Phylogenetic analysis of *Iridaceae* with parsimony and distance methods using the plastid gene *rps4*. *Plant Syst. Evol.*, **204**(1), 109-123.

Starks, C. M., Back, K., Chappell, J. and Noel, J. P. (1997) Structural basis for cyclic terpene biosynthesis by tobacco 5-epi-aristolochene synthase. *Science*, **277**(5333), 1815-1820.

Steele, C. L., Gijzen, M., Qutob, D. and Dixon, R. A. (1999) Molecular characterization of the enzyme catalyzing the aryl migration reaction of isoflavonoid biosynthesis in soybean. *Arch. Biochem. Biophys.*, **367**(1), 146-150.

Stevenson, G., Neal, B., Liu, D., Hobbs, M., Packer, N. H., Batley, M., Redmond, J. W., Lindquist, L. and Reeves, P. (1994) Structure of the O antigen of *Escherichia coli* K-12 and the sequence of its rfb gene cluster. *J. Bacteriol.*, **176**(13), 4144-4156.

Stierle, A., Strobel, G. and Stierle, D. (1993) Taxol and taxane production by *Taxomyces andreanae*, an endophytic fungus of pacific yew. *Science*, **260**(5105), 214-216.

Strasser, R., Stadlmann, J., Schähs, M., Stiegler, G., Quendler, H., Mach, L., Glössl, J., Weterings, K., Pabst, M. and Steinkellner, H. (2008) Generation of glyco-engineered *Nicotiana benthamiana* for the production of monoclonal antibodies with a homogeneous human-like N-glycan structure. *Plant Biotechnology J.*, **6**(4), 392-402.

Strickler, S. R., Bombarely, A. and Mueller, L. A. (2012) Designing a transcriptome next-generation sequencing project for a nonmodel plant species. *Am. J. Bot.*, **99**(2), 257-266.

Sumner, L. W., Lei, Z., Nikolau, B. J. and Saito, K. (2015) Modern plant metabolomics: advanced natural product gene discoveries, improved technologies, and future prospects. *Nat. Prod. Rep.*, **32**(2), 212-229.

Tanaka, T., Antonio, B. A., Kikuchi, S., Matsumoto, T., Nagamura, Y., Numa, H., Sakai, H., Wu, J., Itoh, T., Sasaki, T., Aono, R., Fujii, Y., Habara, T., Harada, E., Kanno, M., Kawahara, Y., Kawashima, H., Kubooka, H., Matsuya, A., Nakaoka, H., Saichi, N., Sanbonmatsu, R., Sato, Y., Shinso, Y., Suzuki, M., Takeda, J., Tanino,

M., Todokoro, F., Yamaguchi, K., Yamamoto, N., Yamasaki, C., Imanishi, T., Okido, T., Tada, M., Ikeo, K., Tateno, Y., Gojobori, T., Lin, Y., Wei, F., Hsing, Y., Zhao, Q., Bin, H., Kramer, M. R., McCombie, R. W., Lonsdale, D., O'Donovan, C. C., Whitfield, E. J., Apweiler, R., Koyanagi, K. O., Khurana, J. P., Raghuvanshi, S., Singh, N. K., Tyagi, A. K., Haberer, G., Fujisawa, M., Hosokawa, S., Ito, Y., Ikawa, H., Shibata, M., Yamamoto, M., Bruskiewich, R. M., Hoen, D. R., Bureau, T. E., Namiki, N., Ohyanagi, H., Sakai, Y., Nobushima, S., Sakata, K., Barrero, R. A., Sato, Y., Souvorov, A., Smith-White, B., Tatusova, T., An, S., An, G., Ota, S., Fuks, G., Messing, J., Christie, K. R., Lieberherr, D., Kim, H., Zuccolo, A., Wing, R. A., Nobuta, K., Green, P. J., Lu, C., Meyers, B. C., Chaparro, C., Piegu, B., Panaud, O. and Echeverria, M. (2008) The rice annotation project database (RAP-DB): 2008 update. *Nucleic Acids Res.*, **36**(Supp 1), D1028-D1033.

Tanner, G. and Kristiansen, K. (1993) Synthesis of 3,4-cis-[^{3}H] leucocyanidin and enzymatic reduction to catechin. *Anal. Biochem.*, **209**(2), 274-277.

Tantillo, D. J. (2010) The carbocation continuum in terpene biosynthesis - where are the secondary cations? *Chem. Soc. Rev.*, **39**(8), 2847-2854.

Tarling, C. A., Woods, K., Zhang, R., Brastianos, H. C., Brayer, G. D., Andersen, R. J. and Withers, S. G. (2008) The search for novel human pancreatic α-amylase inhibitors: high-throughput screening of terrestrial and marine natural product extracts. *ChemBioChem*, **9**(3), 433-438.

Tarshis, L., Yan, M., Poulter, C. D. and Sacchettini, J. C. (1994) Crystal structure of recombinant farnesyl diphosphate synthase at 2.6Å resolution. *Biochem.*, **33**(36), 10871-10877.

Tatusov, R., Natale, D., Garkavtsev, I., Tatusova, T., Shankavaram, U., Rao, B., Kiryutin, B., Galperin, M., Fedorova, N. and Koonin, E. (2001) Genetic diversity of aegilops variabilis for wheat improvement. *Nucleic Acids Res.*, **29**(1), 7.

Teoh, K. H., Polichuk, D. R., Reed, D. W. and Covello, P. S. (2009) Molecular cloning of an aldehyde dehydrogenase implicated in artemisinin biosynthesis in *Artemisia annua*. *Botany*, **87**(6), 635-642.

Thoma, R., Schulz-Gasch, T., D'Arcy, B., Benz, J., Aebi, J., Dehmlow, H., Hennig, M., Stihle, M. and Ruf, A. (2004) Insight into steroid scaffold formation from the structure of human oxidosqualene cyclase. *Nature*, **432**(7013), 118-122.

Thompson, J. D., Higgins, D. G. and Gibson, T. J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.*, **22**(22), 4673-4680.

Thompson, A. M. G, Iancu, C. V., Neet, K. E., Dean, J. V. and Choe, J. (2017) Differences in salicylic acid glucose conjugations by UGT74F1 and UGT74F2 from *Arabidopsis thaliana*. *Sci. Rep.,* **7**, 46629.

Tikkanen, M. J., Wahala, K., Ojala, S., Vihma, V. and Adlercreutz, H. (1998) Effect of soybean phytoestrogen intake on low density lipoprotein oxidation resistance. *Proc. Natl. Acad. Sci. U. S. A.,* **95**(6), 3106-3110.

Tohge, T., Watanabe, M., Hoefgen, R. and Fernie, A. R. (2013) The evolution of phenylpropanoid metabolism in the green lineage. *Crit. Rev. Biochem. Mol. Biol.,* **48**(2), 123-152.

Tomlin, E. S., Borden, J. H. and Pierce Jr., H. D. (1996) Relationship between cortical resin acids and resistance of Sitka spruce to the white pine weevil. *Can. J. Bot.,* **74**(4), 599-606.

Trapero, A., Ahrazem, O., Rubio-Moraga, A., Jimeno, M. L., Gomez, M. D. and Gomez-Gomez, L. (2012) Characterization of a glucosyltransferase enzyme involved in the formation of kaempferol and quercetin sophorosides in crocus sativus. *Plant Physiol.,* **159**(4), 1335-1354.

Trapp, S. and Croteau, R. (2001) Defensive resin biosynthesis in conifers. *Annu. Rev. Plant Biol.,* **52**(1), 689-724.

Tsuruta, H., Paddon, C. J., Eng, D., Lenihan, J. R., Horning, T., Anthony, L. C., Regentin, R., Keasling, J. D., Renninger, N. S. and Newman, J. D. (2009) High-level production of amorpha-4,11-diene, a precursor of the antimalarial agent artemisinin, in *Escherichia coli*. *PLoS One,* **4**(2), e4489.

Tuskan, G. A., Difazio, S., Jansson, S., Bohlmann, J., Grigoriev, I., Hellsten, U., Putnam, N., Ralph, S., Rombauts, S., Salamov, A., Schein, J., Sterck, L., Aerts, A., Bhalerao, R. R., Bhalerao, R. P., Blaudez, D., Boerjan, W., Brun, A., Brunner, A., Busov, V., Campbell, M., Carlson, J., Chalot, M., Chapman, J., Chen, G. L., Cooper, D., Coutinho, P. M., Couturier, J., Covert, S., Cronk, Q., Cunningham, R., Davis, J., Degroeve, S., Dejardin, A., Depamphilis, C., Detter, J., Dirks, B., Dubchak, I., Duplessis, S., Ehlting, J., Ellis, B., Gendler, K., Goodstein, D., Gribskov, M., Grimwood, J., Groover, A., Gunter, L., Hamberger, B., Heinze, B., Helariutta, Y., Henrissat, B., Holligan, D., Holt, R., Huang, W., Islam-Faridi, N., Jones, S., Jones-Rhoades, M., Jorgensen, R., Joshi, C., Kangasjarvi, J., Karlsson, J., Kelleher, C., Kirkpatrick, R., Kirst, M., Kohler, A., Kalluri, U., Larimer, F., Leebens-Mack, J., Leple, J. C., Locascio, P., Lou, Y., Lucas, S., Martin, F., Montanini, B., Napoli, C., Nelson, D. R., Nelson, C., Nieminen, K., Nilsson, O., Pereda, V., Peter, G., Philippe, R., Pilate, G., Poliakov, A., Razumovskaya, J., Richardson, P., Rinaldi, C., Ritland, K., Rouze, P., Ryaboy, D., Schmutz, J., Schrader, J., Segerman, B., Shin, H., Siddiqui, A., Sterky, F., Terry, A., Tsai, C. J., Uberbacher, E., Unneberg, P., Vahala, J., Wall, K., Wessler, S., Yang, G., Yin, T., Douglas, C., Marra, M., Sandberg, G.,

Van de Peer, Y. and Rokhsar, D. (2006) The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science,* **313**(5793), 1596-1604.

Urbanczyk-Wochniak, E., Luedemann, A., Kopka, J., Selbig, J., Roessner-Tunali, U., Willmitzer, L. and Fernie, A. R. (2003) Parallel analysis of transcript and metabolic profiles: a new approach in systems biology. *EMBO Rep.,* **4**(10), 989-993.

van de Laar, F. A. (2008) Alpha-glucosidase inhibitors in the early treatment of type 2 diabetes. *Vasc. Health. Risk Manag.,* **4**(6), 1189-1195.

Vedula, L. S., Rynkiewicz, M. J., Pyun, H., Coates, R. M., Cane, D. E. and Christianson, D. W. (2005) Molecular recognition of the substrate diphosphate group governs product diversity in trichodiene synthase mutants. *Biochemistry,* **44**(16), 6153-6163.

Velasco, R., Zharkikh, A., Affourtit, J., Dhingra, A., Cestaro, A., Kalyanaraman, A., Fontana, P., Bhatnagar, S. K., Troggio, M., Pruss, D., Salvi, S., Pindo, M., Baldi, P., Castelletti, S., Cavaiuolo, M., Coppola, G., Costa, F., Cova, V., Dal Ri, A., Goremykin, V., Komjanc, M., Longhi, S., Magnago, P., Malacarne, G., Malnoy, M., Micheletti, D., Moretto, M., Perazzolli, M., Si-Ammour, A., Vezzulli, S., Zini, E., Eldredge, G., Fitzgerald, L. M., Gutin, N., Lanchbury, J., Macalma, T., Mitchell, J. T., Reid, J., Wardell, B., Kodira, C., Chen, Z., Desany, B., Niazi, F., Palmer, M., Koepke, T., Jiwan, D., Schaeffer, S., Krishnan, V., Wu, C., Chu, V. T., King, S. T., Vick, J., Tao, Q., Mraz, A., Stormo, A., Stormo, K., Bogden, R., Ederle, D., Stella, D., Vecchietti, A., Kater, M. M., Masiero, S., Lasserre, P., Lespinasse, Y., Allan, A. C., Bus, V., Chagné, D., Crowhurst, R. N., Gleave, A. P., Lavezzo, E., Fawcett, J. A., Proost, S., Rouzé, P., Sterck, L., Toppo, S., Lazzari, B., Hellens, R. P., Durel, C. E., Gutin, A., Bumgarner, R. E., Gardiner, S. E., Skolnick, M., Egholm, M., Van de Peer, Y., Salamini, F. and Viola., R. (2010) The genome of the domesticated apple (*Malus x domestica* Borkh.). *Nat. Genet.,* **42**(10), 833-839.

Verpoorte, R., van der Heijden, R. and Memelink, J. (2000) Engineering the plant cell factory for secondary metabolite production. *Transgenic Res.,* **9**(4), 323-343.

The International Brachypodium Initiative (2010) Genome sequencing and analysis of the model grass *Brachypodium distachyon*. *Nature,* **463**(7282), 763-768.

Vogt, T. (2010) Phenylpropanoid biosynthesis. *Molecular Plant,* **3**(1), 2-20.

Vogt, T. and Jones, P. (2000) Glycosyltransferases in plant natural product synthesis: characterization of a supergene family. *Trends Plant Sci.,* **5**(9), 380-386.

Walker, K., Schoendorf, A. and Croteau, R. (2000) Molecular cloning of a taxa-4(20), 11(12)-dien-5α-ol-O-acetyl transferase cDNA from *Taxus* and functional expression in *Escherichia coli*. *Arch. Biochem. Biophys.,* **374**(2), 371-380.

Walker, K. and Croteau, R. (2000) Taxol biosynthesis: molecular cloning of a benzoyl-CoA:taxane-2α-O-benzoyltransferase cDNA from *Taxus* and functional expression in *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.,* **97**(25), 13591-13596.

Wang, D., Du, F., Liu, H. and Liang, Z. (2010a) Drought stress increases iridoid glycosides biosynthesis in the roots of *Scrophularia ningpoensis* seedlings. *J. Med. Plants Res.,* **4**(24), 2691-2699.

Wang, W., Wang, Y., Zhang, Q., Qi, Y. and Guo, D. (2009a) Global characterization of artemisia annua glandular trichome transcriptome using 454 pyrosequencing. *BMC Genomics,* **10**, 465.

Wang, X. (2009) Structure, mechanism and engineering of plant natural product glycosyltransferases. *FEBS Lett.,* **583**(20), 3303-3309.

Wang, X., Han, J., Yang, J., Pan, J. and Borchers, C. H. (2015) Matrix coating assisted by an electric field (MCAEF) for enhanced tissue imaging by MALDI-MS. *Chem. Sci.,* **6**(1), 729-738.

Wang, Z., Fang, B., Chen, J., Zhang, X., Luo, Z., Huang, L., Chen, X. and Li, Y. (2010b) *De novo* assembly and characterization of root transcriptome using Illumina paired-end sequencing and development of cSSR markers in sweetpotato (*Ipomoea batatas*). *BMC Genomics,* **11**, 726.

Wang, Z., Hobson, N., Galindo, L., Zhu, S., Shi, D., McDill, J., Yang, L., Hawkins, S., Neutelings, G., Datla, R., Lambert, G., Galbraith, D. W., Grassa, C. J., Geraldes, A., Cronk, Q. C., Cullis, C., Dash, P. K., Kumar, P. A., Cloutier, S., Sharpe, A. G., Wong, G. K. S., Wang, J. and Deyholos. K. S. (2012) The genome of flax (*Linum usitatissimum*) assembled *de novo* from short shotgun sequence reads. *The Plant Journal,* **72**(3), 461-473.

Wang, Z., Gerstein, M. and Snyder, M. (2009b) RNA-seq: a revolutionary tool for transcriptomics. *Nature Rev. Gen.,* **10**(1), 57-63.

Wang, J., Ma, X. M., Kojima, M., Sakakibara, H. and Hou, B. K. (2011) N-glucosyltransferase UGT76C2 is involved in cytokinin homeostasis and cytokinin response in *Arabidopsis thaliana*. *Plant Cell Physiol.,* **52**(12), 2200-2213.

Wani, M. C., Taylor, H. L., Wall, M. E., Coggon, P. and McPhail, A. T. (1971) Plant antitumor agents. VI. isolation and structure of taxol, a novel antileukemic and antitumor agent from *Taxus brevifolia*. *J. Am. Chem. Soc.,* **93**(9), 2325-2327.

Watt, G., Leoff, C., Harper, A. D. and Bar-Peled, M. (2004) A bifunctional 3,5-epimerase/4-keto reductase for nucleotide-rhamnose synthesis in *Arabidopsis*. *Plant Physiol.,* **134**(4), 1337-1346.

Wei, Y., Liu, L., Zhou, X., Lin, J., Sun, X. and Tang, K. (2012) Engineering taxol biosynthetic pathway for improving taxol yield in taxol-producing endophytic fungus EFY-21 (*Ozonium sp.*). *Afr. J. Biotechnol.,* **11**(37), 9094-9101.

Weitman, M. and Major, D. T. (2010) Challenges posed to bornyl diphosphate synthase: diverging reaction mechanisms in monoterpenes. *J. Am. Chem. Soc.,* **132**(18), 6349-6360.

Wetterhorn K. M., Newmister, S. A., Caniza, R. K., Busman, M., McCormick, S. P., Berthiller, F., Adam, G. and Rayment, I. (2016) Crystal structure of Os79 (Os04g0206600) from *Oryza sativa*: a UDP-glucosyltransferase involved in the detoxification of deoxynivalenol. *Biochemistry,* **55**(44), 6175-6186.

Whittington, D. A., Wise, M. L., Urbansky, M., Coates, R. M., Croteau, R. B. and Christianson, D. W. (2002) Bornyl diphosphate synthase: Structure and strategy for carbocation manipulation by a terpenoid cyclase. *Proc. Natl. Acad. Sci. U.S.A.,* **99**(24), 15375-15380.

Wildung, M. R. and Croteau, R. (1996) A cDNA clone for taxadiene synthase, the diterpene cyclase that catalyzes the committed step of taxol biosynthesis. *J. Biol. Chem.,* **271**(16), 9201-9204.

Williams, L. K., Zhang, X., Caner, S., Tysoe, C., Nguyen, N. T., Wicki, J., Williams, D. E., Coleman, J., McNeill, J. H., Yuen, V., Anderson, R. J., Withers, S. G. and Brayer Gary D (2015) The amylase inhibitor montbretin A reveals a new glycosidase inhibition motif. *Nat. Chem. Biol.,* **11**(9), 691-696.

Winkel-Shirley, B. (2001) Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology. *Plant Physiol.,* **126**(2), 485-493.

Wise, M. L., Pyun, H., Helms, G., Assink, B., Coates, R. M. and Croteau, R. B. (2001) Stereochemical disposition of the geminal dimethyl groups in the enzymatic cyclization of geranyl diphosphate to (+)-bornyl diphosphate by recombinant (+)-bornyl diphosphate synthase from *Salvia officinalis*. *Tetrahedron,* **57**(25), 5327-5334.

Wu, T., Luo, S., Wang, R., Zhong, Y., Xu, X., Lin, Y., He, X., Sun, B. and Huang, H. (2014) The first Illumina-based *de novo* transcriptome sequencing and analysis of pumpkin (*Cucurbita moschata* Duch.) and SSR marker development. *Mol. Breed.,* **34**(3), 1437-1447.

Xiao, M., Zhang, Y., Chen, X., Lee, E., Barber, C. J., Chakrabarty, R., Desgagné-Penix, I., Haslam, T. M., Kim, Y., Liu, E. and MacNevin, G. (2013) Transcriptome analysis based on next-generation sequencing of non-model plants producing specialized metabolites of biotechnological interest. *J. Biotechnol.,* **166**(3), 122-134.

Xu, P., Koffas, M. A. G. (2010) Metabolic engineering of *Escherichia coli* for biofuel production. *Biofuels,* **1**(3), 493-504

Xu, D., Long, H., Liang, J., Zhang, J., Chen, X., Li, J., Pan, Z., Deng, G. and Yu, M. (2012) *De novo* assembly and characterization of the root transcriptome of *Aegilops variabilis* during an interaction with the cereal cyst nematode. *BMC Genomics*, **13**, 133.

Xua, P., Ranganathan, S., Fowler, Z. L., Maranas, C. D., Koffas, M. A. G (2011). Genome-scale metabolic network modeling results in minimal interventions that cooperatively force carbon flux towards malonyl-CoA. *Metab. Eng.*, **13(**5), 578-587.

Ye, J., Fang, L., Zheng, H., Zhang, Y., Chen, J., Zhang, Z., Wang, J., Li, S., Li, R., Bolund, L. and Wang, J. (2006) WEGO: a web tool for plotting GO annotations. *Nucleic Acids Res.*, **34**(suppl_2), W293-W297.

Yin, S. and Kong, J. (2016) Transcriptome-guided gene isolation and functional characterization of UDP-xylose synthase and UDP-d-apiose/UDP-d-xylose synthase families from *Ornithogalum caudatum* Ait. *Plant Cell Rep.*, **35**(11), 2403-2421.

Yin, S., Liu, M. and Kong, J. (2016) Functional analyses of *Oc*RhS1 and *Oc*UER1 involved in UDP-L-rhamnose biosynthesis in *Ornithogalum caudatum*. *Plant Physiol. Biochem.*, **109**(1), 536-548.

Yin, Y., Huang, J., Gu, X., Bar-Peled, M. and Xu, Y. (2011) Evolution of plant nucleotide-sugar interconversion enzymes. *PLoS One*, **6**(11), e27995.

Yonekura-Sakakibara, K. and Hanada, K. (2011) An evolutionary view of functional diversity in family 1 glycosyltransferases. *Plant J.*, **66**(1), 182-193.

Yonekura-Sakakibara, K., Fukushima, A., Nakabayashi, R., Hanada, K., Matsuda, F., Sugawara, S., Inoue, E., Kuromori, T., Ito, T., Shinozaki, K., Wangwattana, B., Yamazaki, M. and Saito, K. (2012) Two glycosyltransferases involved in anthocyanin modification delineated by transcriptome independent component analysis in *Arabidopsis thaliana*. *Plant J.*, **69**(1), 154-167.

Yonekura-Sakakibara, K. and Saito, K. (2009) Functional genomics for plant natural product biosynthesis. *Nat. Prod. Rep.*, **26**(11), 1466-1487.

Yoshikuni, Y., Ferrin, T. E. and Keasling, J. D. (2006) Designed divergent evolution of enzyme function. *Nature*, **440**(7087), 1078-1082.

Young, M. and Neish, A. (1966) Properties of the ammonia-lyases deaminating phenylalanine and related compounds in *Triticum aestivum* and *Pteridium aquilinum*. *Phytochem.*, **5**(6), 1121-1132.

Yuen, V. G., Coleman, J., Withers, S. G., Andersen, R. J., Brayer, G. D., Mustafa, S. and McNeill, J. H. (2016) Glucose lowering effect of montbretin A in zucker diabetic fatty rats. *Mol. Cell. Biochem.,* **411**(1), 373-381.

Zerbe, P., Chiang, A., Dullat, H., O'Neil-Johnson, M., Starks, C., Hamberger, B. and Bohlmann, J. (2014) Diterpene synthases of the biosynthetic system of medicinally active diterpenoids in marrubium vulgare. *Plant J.,* **79**(6), 914-927.

Zerbe, P., Chiang, A., Yuen, M., Hamberger, B., Hamberger, B., Draper, J. A., Britton, R. and Bohlmann, J. (2012) Bifunctional cis-abienol synthase from *Abies balsamea* discovered by transcriptome sequencing and its implications for diterpenoid fragrance production. *J. Biol. Chem.,* **287**(15), 12121-12131.

Zerbe, P., Hamberger, B., Yuen, M. M., Chiang, A., Sandhu, H. K., Madilao, L. L., Nguyen, A., Hamberger, B., Bach, S. S. and Bohlmann, J. (2013) Gene discovery of modular diterpene metabolism in nonmodel systems. *Plant Physiol.,* **162**(2), 1073-1091.

Zhang, J., Liang, S., Duan, J., Wang, J., Chen, S., Cheng, Z., Zhang, Q., Liang, X. and Li, Y. (2012) *De novo* assembly and characterisation of the transcriptome during seed development, and generation of genic-SSR markers in peanut (*Arachis hypogaea L.*). *BMC Genomics,* **13**, 90.

Zhang, W., Seki, M. and Furusaki, S. (1997) Effect of temperature and its shift on growth and anthocyanin production in suspension cultures of strawberry cells. *Plant Science,* **127**(2), 207-214.

Zhang, Q., Shirley, N., Lahnstein, J. and Fincher, G. B. (2005) Characterization and expression patterns of UDP-D-glucuronate decarboxylase genes in barley. *Plant Physiol.,* **138**(1), 131-141.

Zhang, X., Li, M., Agrawal, A., San, A-Y. (2011). Efficient free fatty acid production in Escherichia coli using plant acyl-ACP thioesterases. *Metab. Eng.*, **13**(6), 713-722.

Zhao, K., Ping, W., Zhang, L., Liu, J., Lin, Y., Jin, T. and Zhou, D. (2008) Screening and breeding of high taxol producing fungi by genome shuffling. *Sci. China, C, Life Sci.,* **51**(3), 222-231.

Zhou, S., Chen, L., Liu, S., Wang, X. and Sun, X. (2015) *De novo* assembly and annotation of the Chinese chive (*Allium tuberosum* rottler ex spr.) transcriptome using the Illumina platform. *PLoS One,* **10**(7), e0133312.

Zulak, K. G., Lippert, D. N., Kuzyk, M. A., Domanski, D., Chou, T., Borchers, C. H. and Bohlmann, J. (2009) Targeted proteomics using selected reaction monitoring reveals the induction of specific terpene synthases in a multi-level study of methyl jasmonate-treated Norway spruce (*Picea abies*). *Plant J.,* **60**(6), 1015-1030.

APPENDIX

SUPPLEMENTARY TABLES

Table S2.1: Primers used for cloning *Crocosmia. x crocosmiiflora* early biosynthetic pathway genes

Target Gene	Primer (F = forward primer; R = reverse primer)
CxcPAL1	F: 5′-ATGGAGTTCCAGCACGACAGC-3′; R: 3′-CTATGATATGGGTAGAGGAGCTCCATTCCAC-5′
CxcPAL2	F: 5′-ATGGAGAAGGGCAACGGCAATG-3′; R: 3′-GCTAACATATGGGTAGGTGGGCACCATTCCAATCC-5′
CxcPAL3	F: 5′-ATGGAGTTCGAGAAAGGTAACGG-3′; R: 3′-CTAGGATATGGGAAGAGGAGCACCATTCCAT-5′
CxcPAL4	F: 5′-ATGGAGTTCCAGCACGACAG-3′; R: 3′-CTATGATATGGGTAGAGGAGCTCCATTCCAC-5′
CxcPAL5	F: 5′-ATGGAGAACGGTAACGGAAATGG-3′; R: 3′-CTAACATATGGGTAGGGTTGCACCATTCCAACCC-5′
CxcC4H1	F: 5′-ATGGATATTGACTTCCAGTTCC-3′; R: 3′-TCAAAACACTCTAGGTTTGGCC-5′
CxcC4H2	F: 5′-ATGGATCTTCGTTTTCTTGAG-3′; R: 3′-TTAAAACACTCTAGGTTTGGCC-5′
Cxc4CL1	F: 5′-ATGGGTTCCATTCCTTCGGAG-3′; R: 3′-TCACAGCTGCTGCCCCTTAGC-5′
Cxc4CL2	F: 5′-ATGGCAGTAGATCCAAGAAGC-3′; R: 3′-TCACATCTTGGAGGTAGAGGTG-5′
Cxc4CL3	F: 5′-ATGGAGGGAACCCTAACATCG-3′; R: 3′-CTACATCCTGGAACTTCCATGG-5′
Cxc4CL4	F: 5′-ATGGCAAACTCTGGCAATGGCC-3′; R: 3′-CTATAATTTTGATTTCACTTTC-5′
Cxc4CL5	F: 5′-ATGATCACTATAGCAACTCATG-3′; R: 3′-TTAAGTAGAAACAACCACTCTTG-5′
Cxc4CL6	F: 5′-ATGGCAACATTTTCTGGTTTCG-3′; R: 3′-CTACATTTTTGATCTTACTTTC-5′
CxcC3H1	F: 5′-ATGACCAAAGTGACAACAATGG-3′; R: 3′-TTACATATCGGCAGCGACACGGTTG-5′
CxcF3H1	F: 5′-ATGGGTCAAGAAGTAGATGCAGC-3′; R: 3′-TCACTTGGTCTTCCTAAAATG-5′
CxcF3H2	F: 5′-ATGGGTTTAGAAGTGGATTCAGC-3′; R: 3′-TCACTGGGTCTTCCTAAAATG-5′
CxcF3H3	F: 5′-ATGGCGCCGGGTGCAACTGCG-3′; R: 3′-TTAAGCCAAAATTTCACCGAGG-5′
CxcF3H4	F: 5′-ATGGACATAGAAGTAGACCCAG-3′; R: 3′-TTACTGGTTCCTCCTAAAATG-5′
CxcF3H5	F: 5′-ATGGCGCCGGTCGCGACTACA-3′; R: 3′-TCAAGCAAGGATTTCACTCAGG-5′
CxcF3′H1	F: 5′-ATGCTTACCTTCTTCTTCCTCTG-3′; R: 3′-TTAATAAGCCTTATGCGATAGC-5′
CxcF3′H2	F: 5′-ATGCTATATATAACGTTGCACAC-3′; R: 3′-TCAGAGACGGTAAACCGGGGTTG-5′
CxcF3′H3	F: 5′-ATGACAATGACTTCCCTTGATAT-3′; R: 3′-TCATTTCATCTTGGGCGAATAAG-5′
CxcF3′5′H1	F: 5′-ATGGAACAAACCACCCAGTACTC-3′; R: 3′-CTAAACAAACTGGTCAGAGTTG-5′
CxcF3′5′H2	F: 5′-ATGATCATGGACGTCATTACATC-3′; R: 3′-TTAAGCATAATACTCCAACCTTC-5′
CxcF3′5′H3	F: 5′-ATGGATATTGGCAGCAGCAATTA-3′; R: 3′-CTAAGAATAAAGATCAAGGCTAG-5′
CxcF3′5′H4	F: 5′-ATGGAATACACAGGCTTCAACTGC-3′; R: 3′-CTAAGAATACTGTTGAGGG-5′
CxcF3′5′H5	F: 5′-ATGGAGTACACCAAAATGCACTAT-3′; R: 3′-CTAAGAATATTGTTCAGGA-5′
CxcFLS1	F: 5′-ATGGAGGTGGAGAGAGTTCAAG-3′; R: 3′-TCACTGGGGAAGCTTGTTAAG-5′
CxcFLS2	F: 5′-ATGGATTGCCTTCAGGATTGGC-3′; R: 3′-TCAATTACTCTTCAGTGACTCC-5′

Target Gene	Primer (F = forward primer; R = reverse primer)
CxcFLS3	F: 5′-ATGGATTTGCTACAAGACTGG-3′; R: 3′-TTACATGGCCGTCAATGACTCG-5′
CxcFLS4	F: 5′-ATGGAGGCGGTGAGCAATCCC-3′; R: 3′-TCAAATCTTGAGGGCATCGATA-5′
CxcFLS5	F: 5′-ATGGAGGCGGTGAGCAATCCC-3′; R: 3′-TCAAATCTTGAGGGCATCGATA-5′
CxcRHM1	F: 5′-ATGGCGACATATAAGCCGAAGAAC-3′; R: 3′-TTAGCAGGAACCTTCCTGTTTGG-5′
CxcRHM2	F: 5′-ATGGCGACATATAAGCCGAAGAAC-3′; R: 3′-TTAGCAGGAACCTTCCTGTTTGG-5′
CxcRHM3	F: 5′-ATGACTAATCATACACCGAAGAACA-3′; R: 3′-TCAGACCTTCTTGTTGGGTTCAAAGAC-5′
CxcRHM4	F: 5′-ATGACAACCCACACACCGAAGAA-3′; R: 3′-TCAGACCTTCTTGTTGGGTTCGAAAAC-5′
CxcRHM5	F: 5′-ATGACTAATCATACACCGAAGAACA-3′; R: 3′-TCAGACCTTCTTGTTGGGTTCAAAGAC-5′
CxcUGDH1	F: 5′-ATGGTGAAGATCTGCTGCATTGG-3′; R: 3′-TTAGGCCACAGCAGGCATGTCC-5′
CxcUGDH2	F: 5′-ATGGTGAAGATCTGCTGCATTGG-3′; R: 3′-TTAGGCCACAGCAGGCATGTCC-5′
CxcUGDH3	F: 5′-ATGGTGAAGATCTGTTGTATTGG-3′; R: 3′-TTAGACCACAGCAGGCATA-5′
CxcUGDH4	F: 5′-ATGGTGAAGATCTGTTGTATTGG-3′; R: 3′-TTAGGCCACAGCAGGCATG-5′
CxcUXS1	F: 5′-ATGATGAAACACCTCTCAAAGCAG-3′; R: 3′-TTAATTTCCCTCACTCAGAATTC-5′
CxcUXS2	F: 5′-ATGGTGTCGGAGCTGATCTTCCG-3′; R: 3′-TTAGACAGCAGATGTCGAGTCGG-5′
CxcUXS3	F: 5′-ATGGCCGGGAATGATTCGACC-3′; R: 3′-TTATGCATTCTTGGGAACACC-5′
CxcUXS4	F: 5′-ATGGCGGGGAAGGATTCATCCA-3′; R: 3′-TTATGGCTTCTTGGGGACGCCAAGC-5′
CxcUXS5	F: 5′-ATGGTGTCGGAGCTGATCTTCCGCGG-3′; R: 3′-TTAGACAGCGGATGTCGAGTCGGTGG-5′

Table S2.2: Average montbretin A accumulation levels in six major *Crocosmia x crocosmiiflora* organs over a 12-month period with standard error. "N/A" = organ not available for analysis at time point. "n.d." = not detected in organ.

Extraction Date	Average MbA / g of FW Tissue (mg/g)					
	Flower	Leaf	Stem	Corm	Stolon	Root
29-May	N/A	n.d.	0.003 ± 0.002	2.01 ± 0.05	0.007 ± 0.001	n.d.
18-Jun	N/A	n.d.	0.006 ± 0.001	2.29 ± 0.23	0.006 ± 0.002	n.d.
04-Jul	N/A	n.d.	0.005 ± 0.002	2.41 ± 0.29	0.008 ± 0.002	n.d.
22-Jul	0.145 ± 0.010	n.d.	0.007 ± 0.001	3.24 ± 0.53	0.005 ± 0.001	n.d.
13-Aug	0.038 ± 0.007	n.d.	0.011 ± 0.001	2.90 ± 0.86	0.010 ± 0.001	n.d.
03-Sep	0.0267 ± 0.012	n.d.	0.008 ± 0.003	3.57 ± 1.18	0.006 ± 0.001	n.d.
24-Sep	N/A	n.d.	0.010 ± 0.003	3.55 ± 0.47	0.002 ± 0.000	n.d.
13-Oct	N/A	n.d.	0.014 ± 0.004	4.31 ± 1.46	0.003 ± 0.001	n.d.
05-Nov	N/A	n.d.	0.047 ± 0.005	3.26 ± 0.87	0.009 ± 0.001	n.d.
27-Nov	N/A	n.d.	0.019 ± 0.005	2.34 ± 0.39	0.010 ± 0.003	n.d.
17-Dec	N/A	n.d.	N/A	2.02 ± 0.49	0.009 ± 0.001	n.d.
14-Jan	N/A	n.d.	N/A	3.33 ± 0.40	0.010 ± 0.006	n.d.
06-Feb	N/A	n.d.	N/A	2.94 ± 0.3	0.013 ± 0.001	n.d.
26-Feb	N/A	n.d.	N/A	3.40 ± 1.09	0.013 ± 0.002	n.d.
20-Mar	N/A	n.d.	N/A	2.28 ± 0.25	0.008 ± 0.003	n.d.
08-Apr	N/A	n.d.	N/A	2.22 ± 0.20	0.005 ± 0.003	n.d.
01-May	N/A	n.d.	N/A	1.85 ± 0.31	0.008 ± 0.001	n.d.

Table S2.3: Average montbretin A accumulation levels within corm segments of *Crocosmia x crocosmiiflora*.

Corm Segment	Average MbA /g of FW Tissue (mg/g)
1	0.30 ± 0.06
2	2.37 ± 0.44
3	2.56 ± 0.40
4	2.53 ± 0.41
5	3.19 ± 0.42
6	3.62 ± 0.46
7	3.86 ± 0.42

Table S3.1: Primers used for cloning *Crocosmia. x crocosmiiflora* Nucleotide Sugar Interconversion Enzyme Genes into pET28b(+) Vector

Target Enzyme	Primer (F = forward primer; R = reverse primer)
*Cxc*UXS1	F: 5′-CAACATGCTAGCATGAAACACCTCTCAAAGCAGTCC-3′; R: 3′-CAACATGCTAGCTTAATTTCCCTCACTCAGAATTC-5′
*Cxc*UXS1(Δ1-69)	F: 5′-CAACATGCTAGCATGTCCAGCCACTTCCCCCACATCCCCCGCC-3′; R: 3′-CAACATGCTAGCTTAATTTCCCTCACTCAGAATTC-5′
*Cxc*UXS1(Δ1-81)	F: 5′-CAACATGCTAGCATGCTGACCCAGACCAATACCCCTATCCTC-3′; R: 3′-CAACATGCTAGCTTAATTTCCCTCACTCAGAATTC-5′
*Cxc*UXS2	F: 5′-CAACATGCTAGCATGGCTCGAGTTTTTCAGCAAGATATGG-3′; R: 3′-CAACATGCTAGCTTAGACAGCAGATGTCGAGTCGG-5′
*Cxc*UXS2(Δ1-89)	F: 5′-CAACATGCTAGCATGTGGTACGGGGAGCAGCGGCGGCGGTCG-3′; R: 3′- CAACATGCTAGCTTAGACAGCAGATGTCGAGTCGG-5′
*Cxc*UXS2(Δ1-98)	F: 5′-CAACATGCTAGCATGGTCGCCGGGAAGATCCCTCTCGGCC-3′; R: 3′-CAACATGCTAGCTTAGACAGCAGATGTCGAGTCGG-5′
*Cxc*UXS3	F: 5′-CAACATGCTAGCATGGCTCGAGTTTTTCAGCAAGATATGG-3′; R: 3′-CAACATGCTAGCTTATGGCTTCTTGGGGACGCC-5′
*Cxc*UXS4	F: 5′-CAACATGCTAGCATGGCCGGGAATGATTCGACCAACGG-3′; R: 3′-CAACATGCTAGCTTATGGCTTCTTGGGGACGCC-5′
*Cxc*UXS5	F: 5′-CAACATGCTAGCATGGCGAACGAATCTACCAACGGCGAT-3′; R: 3′- GGGGGTGCCCACAAAACAGGCTACTTAGGCTAGCATGTTG-5′
*Cxc*RHM1	F: 5′-CAACATGCTAGCATGGCGACATATAAGCCGAAGAACATCC-3′; R: 3′-CAACATGCTAGCTTAATTAGCAGGAACCTTCCTG-5′
*Cxc*RHM2	F: 5′-CAACATGCTAGCATGGCGACATATAAGCCGAAGAAC-3′; R: 3′-CAACATGCTAGCTTAATTAGCAGGAACCTTCCTG-5′
*Cxc*RHM3	F: 5′-CAACATGCTAGCATGACTAATCATACACCGAAGAAC-3′; R: 3′-CAACATGCTAGCTCAGACCTTCTTGTTGGGTTCAAAG-5′
*Cxc*RHM4	F: 5′-CAACATGCTAGCATGACAACCCACACACCGAAGAAC-3′; R: 3′-CAACATGCTAGCTCAGACCTTCTTGTTGGGTTCG-5′
*Cxc*RHM5	F: 5′-CAACATGCTAGCATGACTAATCATACACCGAAGAAC-3′; R: 3′-CAACATGCTAGCTCAGACCTTCTTGTTGGGTTCAAAG-5′

Table S3.2: NCBI-nr identified putative and characterized plant UXS.

Abbreviation on Tree	Species	Gene Accession Number
*Al*UXS1	*Arabidopsis lyrata*	XP_020886747
*At*UXS1	*Arabidopsis thaliana*	NP_190920
*At*UXS2	*Arabidopsis thaliana*	AAK32785
*At*UXS3	*Arabidopsis thaliana*	NP_001078768
*At*UXS4	*Arabidopsis thaliana*	NP_182287
*At*UXS5	*Arabidopsis thaliana*	NP_190228
*At*UXS6	*Arabidopsis thaliana*	NP_001325413
*Cc*UXS1	*Cajanus cajan*	KYP58107
*Cxc*UXS1	*Crocosmia x crocosmiiflora*	N/A
*Cxc*UXS2	*Crocosmia x crocosmiiflora*	N/A
*Cxc*UXS3	*Crocosmia x crocosmiiflora*	N/A
*Cxc*UXS4	*Crocosmia x crocosmiiflora*	N/A
*Cxc*UXS5	*Crocosmia x crocosmiiflora*	N/A
*Eg*UXS1	*Eucalyptus grandis*	NP_001289643
*Ga*UXS1	*Gossypium arboreum*	KHG01436.1
*Gh*UXS1	*Gossypium hirsutum*	NP_001314595
*Gh*UXS2	*Gossypium hirsutum*	NP_001314162
*Gh*UXS3	*Gossypium hirsutum*	NP_001314618
*Gh*UXS6	*Gossypium hirsutum*	XP_016678896
*Md*UXS2	*Malus domestica*	XP_008369016
*Mt*UXS1	*Medicago truncatula*	XP_003606663
*Mt*UXS2	*Medicago truncatula*	XP_013468298
*Nt*UXS3	*Nicotiana tabacum*	NP_001312321
*Nt*UXS6	*Nicotiana tabacum*	NP_001313009
*Ol*UXS1	*Ornithogalum longebracteatum*	AMM04374
*Ol*UXS2	*Ornithogalum longebracteatum*	AMM04375
*Ol*UXS3	*Ornithogalum longebracteatum*	AMM04376
*Os*UXS1	*Oryza sativa*	XP_015621250
*Os*UXS2	*Oryza sativa*	XP_015639298
*Os*UXS4	*Oryza sativa*	XP_015647595
*Os*UXS6	*Oryza sativa*	XP_015632111
*Pt*UXS1	*Populus tomentosa*	AAX37334
*Pt*UXS2	*Populus tomentosa*	AAX37335
*Pt*UXS3	*Populus tomentosa*	AAX37336
*Pv*UXS1	*Phaseolus vulgaris*	AFW90530
*Zm*UXS1	*Zea mays*	NP_001151221

Table S3.3: Average *E. coli* pellet weight post-overnight protein expression culture.

Gene Expressed	Average Bacterial Pellet Weight (g)	Standard Error of Pellet Weight (g)
UXS-1	0.51	0.014
UXS-2	0.47	0.017
UXS-3	1.60	0.111
UXS-4	1.59	0.125
UXS-5	1.69	0.117

Table S3.4: Effects of temperature and pH on relative activity of *Cxc*UXS. The activities of recombinant *Cxc*UXSs were analyzed over different temperatures and buffer pHs. Values presented are the average relative activity +/- standard error based on quintuplicate reaction. 100% relative activity corresponds to the highest level of activity observed for that condition.

Enzyme	Temperature		pH	
	°C	Relative Activity (% ± SE)	pH	Relative Activity (% ± SE)
***Cxc*UXS1(Δ1–68)**	4	15.30 ± 1.15	3.5	20.54 ± 1.98
	12	21.11 ± 1.05	4.5	31.02 ± 4.67
	22	42.55 ± 2.59	5.5	60.29 ± 4.59
	30	67.04 ± 3.53	6.5	100.00
	37	100.00	7.0	66.18 ± 2.94
	45	20.75 ± 3.86	7.5	46.57 ± 4.94
	55	1.83 ± 0.48	8.0	37.28 ± 3.88
			8.5	19.35 ± 3.28
***Cxc*UXS2(Δ1–89)**	4	13.98 ± 2.14	3.5	7.48 ± 3.35
	12	34.94 ± 3.92	4.5	9.56 ± 2.76
	22	88.09 ± 3.67	5.5	26.92 ± 4.49
	30	100.00	6.5	48.15 ± 4.60
	37	82.70 ± 4.97	7.0	100.00
	45	53.90 ± 4.69	7.5	42.38 ± 4.65
	55	0.97 ± 0.12	8.0	34.80 ± 3.65
			8.5	29.91 ± 4.26
***Cxc*UXS3**	4	60.62 ± 4.75	3.5	1.26 ± 0.52
	12	63.13 ± 3.61	4.5	19.33 ± 2.63
	22	77.54 ± 2.63	5.5	59.77 ± 3.55
	30	100.00	6.5	84.90 ± 2.63
	37	80.95 ± 4.27	7.0	100.00
	45	47.72 ± 4.92	7.5	83.00 ± 3.43
	55	18.78 ± 4.19	8.0	75.79 ± 3.42
			8.5	51.48 ± 3.90
***Cxc*UXS4**	4	45.76 ± 3.98	3.5	4.33 ± 0.21
	12	58.63 ± 3.66	4.5	34.27 ± 3.47
	22	79.71 ± 3.97	5.5	76.59 ± 1.66
	30	100.00	6.5	92.55 ± 1.84
	37	95.93 ± 2.25	7.0	100.00
	45	78.06 ± 2.06	7.5	89.41 ± 3.19
	55	31.81 ± 2.80	8.0	70.66 ± 3.13
			8.5	60.65 ± 2.72
***Cxc*UXS5**	4	47.86 ± 3.57	5.5	54.68 ± 3.54
	12	62.41 ± 1.02	6.5	70.73 ± 2.91
	22	74.95 ± 2.28	7.0	100.00
	30	82.22 ± 2.37	7.5	89.43 ± 2.76
	37	100.00	8.0	78.41 ± 1.83
	45	82.94 ± 4.60	8.5	71.82 ± 2.37
	55	23.47 ± 2.61	9.0	59.33 ± 1.73
			9.5	46.68 ± 3.15

Table S3.5: NCBI-nr identified putative and characterized plant RHM.

Abbreviation on Tree	Species	Gene Accession Number
*At*RHM1	*Arabidopsis thaliana*	NP_177978
*At*RHM2	*Arabidopsis thaliana*	XP_020867140
*At*RHM3	*Arabidopsis thaliana*	NP_188097
*Bd*RHM1	*Brachypodium distachyon*	XP_003571828
*Cxc*RHM1	*Crocosmia x crocosmiiflora*	N/A
*Cxc*RHM2	*Crocosmia x crocosmiiflora*	N/A
*Cxc*RHM3	*Crocosmia x crocosmiiflora*	N/A
*Cxc*RHM4	*Crocosmia x crocosmiiflora*	N/A
*Cxc*RHM5	*Crocosmia x crocosmiiflora*	N/A
*Gh*RHM1	*Gossypium hirsutum*	NP_001314210
*Gh*RHM2	*Gossypium hirsutum*	NP_001313649
*Gh*RHM3	*Gossypium hirsutum*	NP_001313859
*Gh*RHM4	*Gossypium hirsutum*	NP_001314630
*Gm*RHM1	*Glycine max*	XP_003543185
*Gm*RHM2	*Glycine max*	XP_003546628
*Gm*RHM3	*Glycine max*	XP_003551914
*Gm*RHM4	*Glycine max*	XP_006585326
*Hv*RHM1	*Hordeum vulgare*	BAJ97692
*Os*RHM1	*Oryza sativa*	XP_015630773
*Pe*RHM1	*Populus euphratica*	XP_011045548
*Pe*RHM2	*Populus euphratica*	XP_011031381
*Rc*RHM1	*Ricinus communis*	XP_015582142
*Sb*RHM1	*Sorghum bicolor*	XP_002440852
*Sb*RHM3	*Sorghum bicolor*	XP_002468088
*Sm*RHM1	*Selaginella moellendorffii*	XP_002964674
*Vv*RHM1	*Vitis vinifera*	XP_002285634
*Vv*RHM2	*Vitis vinifera*	XP_002271970
*Zm*RHM1	*Zea mays*	NP_001151455
*Zm*RHM2	*Zea mays*	XP_020401614

Table S3.6: Effects of temperature and pH on relative activity of *Cxc*RHM. The activities of recombinant *Cxc*RHMs were analyzed over different temperatures and buffer pHs. Values presented are the average relative activity +/- standard error based on quintuplicate reaction. 100% relative activity corresponds to the highest level of activity observed for that condition.

Enzyme	Temperature		pH	
	°C	Relative Activity (% ± SE)	pH	Relative Activity (% ± SE)
***Cxc*RHM1**	4	38.65 ± 4.96	5.5	43.14 ± 2.56
	12	55.37 ± 4.77	6.5	59.14 ± 3.18
	22	71.91 ± 3.74	7.0	78.94 ± 4.11
	30	84.18 ± 3.47	7.5	93.65 ± 0.83
	37	100.00	8.0	100.00
	45	87.74 ± 3.14	8.5	85.98 ± 2.65
	55	16.02 ± 4.01	9.0	73.93 ± 2.44
			9.5	69.98 ± 3.04
***Cxc*RHM2**	4	16.24 ± 3.50	5.5	10.15 ± 0.93
	12	28.12 ± 4.89	6.5	30.62 ± 1.81
	22	55.93 ± 3.59	7.0	45.18 ± 4.62
	30	100.00	7.5	71.93 ± 2.90
	37	90.62 ± 3.34	8.0	89.01 ± 1.69
	45	49.13 ± 4.81	8.5	96.49 ± 1.04
	55	2.20 ± 1.33	9.0	100.00
			9.5	63.00 ± 4.13
***Cxc*RHM3**	4	9.15 ± 2.52	3.5	8.93 ± 1.57
	12	12.13 ± 1.16	4.5	20.95 ± 2.79
	22	52.09 ± 4.97	5.5	27.39 ± 2.96
	30	100.00	6.5	50.39 ± 4.36
	37	67.66 ± 4.85	7.0	70.81 ± 3.69
	45	18.91 ± 2.66	7.5	100.00
	55	0.00 ± 0.00	8.0	84.25 ± 4.69
			8.5	72.08 ± 3.39
***Cxc*RHM4**	4	13.69 ± 4.06	3.5	3.19 ± 0.84
	12	28.64 ± 4.34	4.5	19.91 ± 1.34
	22	76.09 ± 4.86	5.5	31.48 ± 2.70
	30	100.00	6.5	54.38 ± 4.08
	37	78.78 ± 4.49	7.0	100.00
	45	24.81 ± 3.57	7.5	53.03 ± 1.93
	55	3.68 ± 2.19	8.0	15.33 ± 3.21
			8.5	5.67 ± 0.87
***Cxc*RHM5**	4	10.81 ± 2.21	3.5	26.46 ± 1.92
	12	17.79 ± 3.38	4.5	34.91 ± 1.23
	22	50.27 ± 4.16	5.5	46.84 ± 2.28
	30	80.24 ± 4.12	6.5	65.80 ± 0.80
	37	100.00	7.0	83.13 ± 0.85
	45	66.75 ± 3.96	7.5	100.00
	55	0.11 ± 0.09	8.0	90.92 ± 4.77
			8.5	76.89 ± 4.20

Table S3.7: NCBI-nr identified putative and characterized plant RHM and UER.

Abbreviation on Tree	Species	Gene Accession Number	Amino Acid Sequence Used
*At*UER	*Arabidopsis thaliana*	AEE34035	FL
*Cxc*UER1	*Crocosmia x crocosmiiflora*	N/A	FL
*Gh*UER	*Gossypium hirsutum*	ACJ11713	FL
*Na*UER	*Nicotiana attenuata*	OIT01948	FL
*Ol*UER1	*Ornithogalum longebracteatum*	ANK57461	FL
*Os*UER	*Oryza sativa*	BAD29243	FL
*Pt*UER	*Populus trichocarpa*	XP_002299567	FL
*Rc*UER	*Ricinus communis*	XP_002509687	FL
*Sb*UER	*Sorghum bicolor*	XP_002454480	FL
*Tc*UER	*Theobroma cacao*	EOY24593	FL
*Vv*UER	*Vitis vinifera*	XP_002282339	FL
*Zm*UER	*Zea mays*	NP_001152718	FL
*At*RHM1	*Arabidopsis thaliana*	NP_177978	386-669
*At*RHM3	*Arabidopsis thaliana*	NP_188097	384-667
*Bd*RHM1	*Brachypodium distachyon*	XP_003571828	381-667
*Cxc*RHM1	*Crocosmia x crocosmiiflora*	N/A	386-672
*Cxc*RHM2	*Crocosmia x crocosmiiflora*	N/A	386-672
*Cxc*RHM3	*Crocosmia x crocosmiiflora*	N/A	389-672
*Cxc*RHM4	*Crocosmia x crocosmiiflora*	N/A	389-672
*Cxc*RHM5	*Crocosmia x crocosmiiflora*	N/A	389-672
*Gh*RHM1	*Gossypium hirsutum*	NP_001314210	384-667
*Gh*RHM2	*Gossypium hirsutum*	NP_001313649	384-667
*Gh*RHM4	*Gossypium hirsutum*	NP_001314630	391-681
*Gm*RHM1	*Glycine max*	XP_003543185	385-669
*Gm*RHM2	*Glycine max*	XP_003546628	385-668
*Hv*RHM1	*Hordeum vulgare*	BAJ97692	381-667
*Os*RHM1	*Oryza sativa*	XP_015630773	393-679
*Pe*RHM1	*Populus euphratica*	XP_011045548	387-670
*Rc*RHM1	*Ricinus communis*	XP_015582142	364-647
*Sb*RHM1	*Sorghum bicolor*	XP_002440852	380-666
*Sb*RHM3	*Sorghum bicolor*	XP_002468088	386-672
*Sb*RHM6	*Sorghum bicolor*	XP_002468080	386-672
*Vv*RHM1	*Vitis vinifera*	XP_002285634	389-675
*Zm*RHM1	*Zea mays*	NP_001151455	386-672
*Zm*RHM2	*Zea mays*	XP_020401614	386-672
*Zm*RHM3	*Zea mays*	NP_001130297	390-676

Table S4.1: Primers used for cloning *Crocosmia. x crocosmiiflora* UDP-glycosyltransferase sequences into pASKIBA37+ Vector.

Target Enzyme	Primer (F = forward primer; R = reverse primer)
*Cxc*UGT1	F: 5′-CAACATGGTCTCAGAGTATGCAAGTTGCACAAAATCAGC-3′; R: 3′-GGTTTGGCATACAATGAGGCTTGATAGCTGAGACCATGTTG-5′
*Cxc*UGT2	F: 5′-CAACATGGTCTCAGAGTATGGGTTCCGAAGGAAATACATTAAACATGC-3′; R: 3′-GATGTTGTGGTTGGAAATTGATAGCTGAGACCATGTTG-5′
*Cxc*UGT3	F: 5′-CAACATGGTCTCAGAGTATGGGTTCCGAAGGAAATACATTAAACATGC-3′; R: 3′-GATGTTGTGGTTGGAAATTGATAGCTGAGACCATGTTG-5′
*Cxc*UGT4	F: 5′-CAACATGGTCTCAGAGTATGGGTTCCGAAGGAAATACATTAAACATGC-3′; R: 3′-GATGTTGTGGTTGGAAATTGATAGCTGAGACCATGTTG-5′
*Cxc*UGT5	F: 5′-CAACATGGTCTCAGAGTATGGAAGCTCAACCTCCACTT-3′; R: 3′-AGTGGAGAGGGTGCAAAATAGTAGCTGAGACCATGTTG-5′
*Cxc*UGT6	F: 5′-CAACATGGTCTCAGAGTATGAAGAAAGACCACATAGTG-3′; R: 3′-GACGAGATTCAGATTGCATGATTAGCTGAGACCATGTTG-5′
*Cxc*UGT7	F: 5′- CAACATGGTCTCAGAGTATGGAAGCTCAACCTCC -3′; R: 3′- GGAGAGGGTGCAAAATAGTAGCTGAGACCATGTTG -5′
*Cxc*UGT8	F: 5′-CAACATGGTCTCAGAGTATGGAGGGTTTGGAAGAGTCTAACG-3′; R: 3′-GTTCTGGAGGTCCTGAGGAAAGGAAGTGATAGCTGAGACCATGTTG-5′
*Cxc*UGT9	F: 5′-CAACATGGTCTCAGAGTATGGGCTCTACGGATGA-3′; R: 3′-GCTCACAGAAGTGAAGAGTAAGGCCAAGCAGTGATAGCTGAGACCATGTTG-5′
*Cxc*UGT10	F: 5′-CAACATGGTCTCAGAGTATGTCTCAATTCAATGCCATGGAAG-3′; R: 3′-GAATGTTTTGAAAAACTTGAGCTCAGATTAGTAGCTGAGACCATGTTG-5′
*Cxc*UGT11	F: 5′-CAACATGGTCTCAGAGTATGGATCACCAAGAGCAAGAGCAG-3′; R: 3′-CAACATGGTCTCAGCTATCAACATCGATCGTTTTGGATTGGATTGG-5′
*Cxc*UGT12	F: 5′-CAACATGGTCTCAGAGTATGGAAAACCAGAAGCAGGAGCTCC-3′; R: 3′-CAAGCCAAGAGTCCATCCAATCCCAATTGATAGCTGAGACCATGTTG-5′
*Cxc*UGT13	F: 5′-CAACATGGTCTCAGAGTATGAGCTCCGACGACGAAGTGCAC-3′; R: 3′-CAAGTCCAAGGAGAGCTCTGTTCTTGTCTGATAGCTGAGACCATGTTG-5′
*Cxc*UGT14	F: 5′-CAACATGGTCTCAGAGTATGGTGAAAGATCAGGAAAAGGAC-3′; R: 3′-GAGCTTCGAGAGGAAGAAAAGAAGTGACTAATAGCTGAGACCATGTTG-5′
*Cxc*UGT1(C432ΔA)	5′-GGGCCCAGCGTGCTTTGGTCTACACTGCGCACACAGCATCATG-3′
*Cxc*UGT6(A1255ΔC)	5′-GAAGACGATTCTCGGGCGACCGACGGCGTTACACACG-3′
*Cxc*UGT7(C858ΔT)	5′-GATATTGCTCTTGGTCTTGAAGCGTCGGGCCACCC-3′
*Cxc*UGT11(C981ΔA)	5′-GAGTGGGCACCGCAGGTATCGATCCTGAACCATCCGTC-3′
*Cxc*UGT13(C468ΔA)	5′-GATCTATCAGCTGGGTCTATTCGAGGGGCGTGGGGACC-3′
*Cxc*UGT14(T294ΔC)	5′-GCTTTCTACGACGTAGGGACCTTCCGGGCGGCTTTC-3′
*Cxc*UGT14(A582ΔG)	5′-GTCGTGGATTCGTTAGGGTCCCGATCAAGAGAGCAC-3′

Table S4.2: Distribution of GT1 UGTs in the different phylogenetic groups across multiple plant species. Distribution of GT1 UGT into phylogenetic groups were obtained from (Barvkar *et al.*, 2012; Caputi *et al.*, 2012; Huang *et al.*, 2015; Li *et al.*, 2001; Li *et al.*, 2014b).

Species	GT1 UGT Group																	Total	Genome Size
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q		
Arabidopsis thaliana	14	3	3	13	22	3	6	19	1	2	2	17	1	1	-	-	-	107	135 Mb
Malus domestica	33	4	7	13	55	6	40	14	11	12	6	16	13	1	5	5	-	241	750 Mb
Vitis vinifera	23	3	4	8	46	5	15	7	14	4	2	31	5	1	2	11	-	181	487 Mb
Populus trichocarpa	12	2	6	14	49	0	42	5	5	6	2	23	6	1	3	2	-	178	485 Mb
Glycine Max	25	3	1	43	36	1	15	3	18	3	2	19	4	1	5	3	-	182	1115 Mb
Cucumis sativus	10	1	2	12	13	0	11	5	0	2	1	17	2	1	3	5	-	85	243 Mb
Mimulus guttatus	10	2	3	11	14	0	12	1	4	2	9	17	2	1	9	3	-	100	322 Mb
Linum usitatissimum	15	5	6	21	22	1	19	6	9	4	5	19	3	1	-	-	-	136	373 Mb
Gossypium spp.	17	12	0	36	38	8	20	16	10	4	2	18	2	2	1	0	0	186	1,746 – 885 Mb
Oryza sativa	14	9	8	26	38	0	20	7	9	3	1	23	5	2	6	9	-	180	389 Mb
Sorghum bicolor	10	4	6	24	50	0	17	12	8	3	1	26	6	3	8	2	-	180	730 Mb
Zea mays	8	3	4	15	33	2	12	9	9	3	1	22	3	4	4	1	7	140	2058 Mb
Crocosmia x crocosmiiflora	9	5	4	59	13	8	15	0	2	5	1	10	5	2	7	15	-	160	Unknown

Table S4.3: *Crocosmia x crocosmiiflora* GT1 UGT phylogenetic tree legend.

Sequence ID	Listed on Phylogeny	Accession Number
103404_c0_g1_i1	1	TBD
103404_c0_g1_i2	2	TBD
110300_c0_g23_i4	3	TBD
110300_c0_g6_i4	4	TBD
107423_c0_g1_i1	5	TBD
103232_c2_g1_i1	6	TBD
109405_c3_g4_i1	7	TBD
96283_c0_g1_i1	8	TBD
96283_c0_g1_i2	9	TBD
106290_c0_g1_i1	10	TBD
108425_c2_g30_i1	11	TBD
AtUGT73B1_Group_D	12	OAP00589
106061_c0_g4_i1	13	TBD
106061_c0_g4_i2	14	TBD
108425_c2_g35_i3	15	TBD
108425_c2_g35_i5	16	TBD
96852_c0_g2_i1	17	TBD
99708_c0_g2_i1	18	TBD
101583_c0_g1_i1	19	TBD
101583_c0_g1_i2	20	TBD
111470_c0_g1_i1	21	TBD
111470_c0_g1_i2	22	TBD
111470_c0_g1_i4	23	TBD
109389_c0_g1_i1	24	TBD
109191_c0_g2_i4	25	TBD
80442_c0_g1_i1	26	TBD
111473_c5_g2_i1	27	TBD
111473_c5_g2_i7	28	TBD
111410_c2_g5_i5	29	TBD
11410_c2_g5_i2	30	TBD
111410_c2_g5_i4	31	TBD
109746_c3_g8_i1	32	TBD
106826_c0_g1_i1	33	TBD
109746_c3_g1_i2	34	TBD
109746_c3_g6_i1	35	TBD
233282_c0_g1_i1	36	TBD
109746_c3_g1_i1	37	TBD

Sequence ID	Listed on Phylogeny	Accession Number
109746_c3_g1_i4	38	TBD
96884_c0_g1_i1	39	TBD
AtUGT73C1_Group_D	40	OAP11697
109405_c3_g7_i1	41	TBD
95060_c0_g1_i1	42	TBD
109405_c3_g9_i2	43	TBD
GRMZM2G035755_P01_Group_D	44	NP_001140972
107423_c1_g1_i1	45	TBD
GRMZM2G479038_P01_Group_D	46	NP_001148090
110539_c0_g2_i2	47	TBD
110539_c0_g4_i1	48	TBD
83138_c0_g1_i1	49	TBD
110539_c0_g1_i1	50	TBD
19879_c0_g1_i1	51	TBD
108925_c0_g1_i1	52	TBD
GRMZM2G175910_P01_Group_C	53	ONM54727
AtUGT90A1_Group_C	54	Q9ZVX4
108156_c0_g1_i3	55	TBD
108156_c0_g2_i1	56	TBD
GRMZM2G301148_P01_Group_B	57	AQK58136
104519_c0_g1_i1	58	TBD
104519_c0_g1_i4	59	TBD
95748_c1_g1_i1	60	TBD
AtUGT89B1_Group_B	61	OAP14423
106481_c2_g2_i1	62	TBD
106481_c2_g2_i2	63	TBD
110860_c1_g1_i3	64	TBD
110860_c1_g1_i9	65	TBD
110860_c1_g1_i10	66	TBD
110860_c1_g1_i2	67	TBD
101680_c0_g1_i1	68	TBD
AtUGT92A1_Group_M	69	Q9LXV0
GRMZM2G159404_P01_Group_M	70	NP_001148083
109013_c5_g2_i2	71	TBD
109013_c5_g2_i3	72	TBD
111562_c4_g15_i1	73	TBD
111562_c4_g7_i5	74	TBD
108400_c0_g1_i1	75	TBD

Sequence ID	Listed on Phylogeny	Accession Number
108400_c0_g1_i2	76	TBD
AtUGT91A1_Group_A	77	Q940V3
64576_c0_g1_i1	78	TBD
GRMZM2G061289_P01_Group_A	79	ONM36204
111438_c6_g4_i1	80	TBD
111438_c6_g9_i8	81	TBD
106539_c5_g1_i1	82	TBD
GRMZM2G120016_P01_Group_O	83	NP_001141165
Os07g46610_Group_O	84	XP_015646369
100212_c0_g1_i1	85	TBD
106688_c0_g3_i1	86	TBD
110632_c3_g7_i3	87	TBD
110632_c3_g6_i1	88	TBD
88334_c0_g1_i1	89	TBD
96927_c0_g1_i1	90	TBD
106678_c7_g12_i7	91	TBD
106678_c6_g1_i1	92	TBD
110940_c0_g2_i1	93	TBD
110940_c0_g2_i3	94	TBD
GRMZM2G058314_P01_Group_E	95	NP_001149283
85901_c1_g1_i1	96	TBD
109457_c2_g8_i11	97	TBD
109457_c2_g8_i9	98	TBD
AtUGT72C1_Group_E	99	O23205
89445_c0_g1_i1	100	TBD
92712_c1_g1_i1	101	TBD
102407_c2_g3_i2	102	TBD
102407_c2_g3_i3	103	TBD
109669_c0_g3_i1	104	TBD
98439_c0_g1_i1	105	TBD
103947_c0_g1_i1	106	TBD
105388_c0_g5_i1	107	TBD
106481_c1_g2_i1	108	TBD
106409_c0_g2_i1	109	TBD
106409_c0_g3_i2	110	TBD
AtUGT75D1_Group_L	111	O23406
97222_c1_g6_i1	112	TBD
102411_c0_g3_i1	113	TBD

Sequence ID	Listed on Phylogeny	Accession Number
102411_c0_g3_i2	114	TBD
GRMZM2G050748_P01_Group_L	115	AQK83517
104689_c0_g1_i1	116	TBD
111335_c2_g1_i4	117	TBD
104749_c0_g2_i1	118	TBD
GRMZM2G173315_P01_Group_K	119	NP_001169578
AtUGT86A1_Group_K	120	Q9SJL0
AtUGT87A1_Group_J	121	O64732
100329_c0_g1_i1	122	TBD
105134_c0_g1_i1	123	TBD
86379_c0_g1_i1	124	TBD
GRMZM2G073376_P01_Group_J	125	NP_001148167
108899_c0_g1_i1	126	TBD
108899_c0_g1_i2	127	TBD
110344_c2_g2_i3	128	TBD
110344_c2_g2_i9	129	TBD
110344_c2_g2_i6	130	TBD
110344_c1_g2_i1	131	TBD
106902_c2_g1_i2	132	TBD
106902_c2_g1_i1	133	TBD
106902_c2_g1_i5	134	TBD
101271_c0_g2_i1	135	TBD
AtUGT78D3_Group_F	136	OAO94865
GRMZM2G022242_P01_Group_F	137	NP_001137065
102764_c0_g1_i1	138	TBD
102764_c0_g1_i3	139	TBD
GRMZM2G021786_P01_Group_N	140	NP_001131410
AtUGT82A1_Group_N	141	Q9LHJ2
AtUGT83A1_Group_I	142	Q9SGA8
GRMZM2G046994_P01_Group_I	143	AQK51585
105481_c1_g1_i1	144	TBD
107566_c1_g1_i1	145	TBD
AtUGT76B1_Group_H	146	OAP05179
GRMZM5G892627_P02_Group_H	147	ONM52585
103372_c0_g1_i1	148	TBD
95258_c0_g1_i1	149	TBD
108074_c1_g1_i1	150	TBD
108074_c1_g1_i2	151	TBD

Sequence ID	Listed on Phylogeny	Accession Number
98996_c0_g1_i1	152	TBD
106992_c1_g2_i1	153	TBD
99562_c2_g1_i1	154	TBD
86843_c0_g1_i1	155	TBD
95258_c0_g1_i2	156	TBD
102420_c0_g1_i1	157	TBD
44324_c0_g1_i1	158	TBD
101539_c3_g3_i1	159	TBD
110738_c4_g1_i1	160	TBD
138206_c0_g1_i1	161	TBD
GRMZM5G834303_P01_Group_P	162	NP_001148991
Os07g30690_Group_P	163	XP_015647802
AtUGT85A4_Group_G	164	OAP13096
109680_c1_g1_i1	165	TBD
109680_c1_g1_i3	166	TBD
82487_c0_g1_i1	167	TBD
GRMZM2G041699_P01_Group_G	168	NP_001149205
109680_c2_g4_i1	169	TBD
94406_c0_g1_i1	170	TBD
103030_c0_g1_i1	171	TBD
88206_c0_g1_i2	172	TBD
94275_c0_g1_i1	173	TBD
109028_c2_g4_i5	174	TBD
109028_c2_g4_i8	175	TBD
99057_c3_g1_i1	176	TBD
99057_c3_g1_i2	177	TBD
105847_c0_g1_i1	178	TBD
105847_c0_g1_i2	179	TBD
105847_c0_g1_i5	180	TBD
104375_c1_g24_i1	*Cxc*UGT1	TBD
194224_c0_g1_i1	*Cxc*UGT10	TBD
106438_c1_g2_i2	*Cxc*UGT11	TBD
89930_c0_g2_i1	*Cxc*UGT12	TBD
91935_c0_g2_i1	*Cxc*UGT13	TBD
100141_c0_g1_i3	*Cxc*UGT14	TBD
110300_c0_g5_i4	*Cxc*UGT2	TBD
110300_c0_g5_i1	*Cxc*UGT3	TBD
110300_c0_g5_i2	*Cxc*UGT4	TBD

Sequence ID	Listed on Phylogeny	Accession Number
93401_c1_g1_i1	*Cxc*UGT5	TBD
108925_c0_g1_i2	*Cxc*UGT6	TBD
93401_c0_g1_i1	*Cxc*UGT7	TBD
98575_c0_g5_i1	*Cxc*UGT8	TBD
109405_c3_g4_i2	*Cxc*UGT9	TBD

Table S4.4: Figure S4.12 phylogenetic tree legend.

Sequence ID	Listed on Phylogeny	Accession Number
109746_c3_g1_i1	1	TBD
109746_c3_g1_i4	2	TBD
233282_c0_g1_i1	3	TBD
109746_c3_g6_i1	4	TBD
109746_c3_g1_i2	5	TBD
106826_c0_g1_i1	6	TBD
109746_c3_g8_i1	7	TBD
111410_c2_g5_i4	8	TBD
111410_c2_g5_i5	9	TBD
11410_c2_g5_i2	10	TBD
106438_c1_g2_i2	11	TBD
89930_c0_g2_i1	12	TBD
111473_c5_g2_i1	13	TBD
111473_c5_g2_i7	14	TBD
109389_c0_g1_i1	15	TBD
93401_c1_g1_i1	16	TBD
93401_c0_g1_i1	17	TBD
106290_c0_g1_i1	18	TBD
108425_c2_g30_i1	19	TBD
96283_c0_g1_i1	20	TBD
96283_c0_g1_i2	21	TBD
91935_c0_g2_i1	22	TBD
109405_c3_g4_i1	23	TBD
109405_c3_g4_i2	24	TBD
103232_c2_g1_i1	25	TBD
107423_c0_g1_i1	26	TBD
110300_c0_g23_i4	27	TBD
110300_c0_g6_i4	28	TBD
103404_c0_g1_i1	29	TBD
103404_c0_g1_i2	30	TBD
110300_c0_g5_i4	31	TBD
110300_c0_g5_i2	32	TBD
110300_c0_g5_i1	33	TBD
*At*UGT73B1_Group_D	34	OAP00589
111470_c0_g1_i2	35	TBD
111470_c0_g1_i4	36	TBD
111470_c0_g1_i1	37	TBD

Sequence ID	Listed on Phylogeny	Accession Number
101583_c0_g1_i1	38	TBD
101583_c0_g1_i2	39	TBD
96852_c0_g2_i1	40	TBD
99708_c0_g2_i1	41	TBD
106061_c0_g4_i1	42	TBD
106061_c0_g4_i2	43	TBD
108425_c2_g35_i3	44	TBD
108425_c2_g35_i5	45	TBD
96884_c0_g1_i1	46	TBD
109405_c3_g7_i1	47	TBD
95060_c0_g1_i1	48	TBD
109405_c3_g9_i2	49	TBD
GRMZM2G035755_P01_Group_D	50	NP_001140972
*At*UGT73C1_Group_D	51	OAP11697
GRMZM2G479038_P01_Group_D	52	NP_001148090
107423_c1_g1_i1	53	TBD
104375_c1_g24_i1	54	TBD
110539_c0_g2_i2	55	TBD
110539_c0_g4_i1	56	TBD
83138_c0_g1_i1	57	TBD
110539_c0_g1_i1	58	TBD
19879_c0_g1_i1	59	TBD
109191_c0_g2_i4	60	TBD
80442_c0_g1_i1	61	TBD
108925_c0_g1_i1	62	TBD
108925_c0_g1_i2	63	TBD
GRMZM2G175910_P01_Group_C	64	ONM54727
*At*UGT90A1_Group_C	65	Q9ZVX4
108156_c0_g1_i3	66	TBD
108156_c0_g2_i1	67	TBD
GRMZM2G301148_P01_Group_B	68	AQK58136
104519_c0_g1_i1	69	TBD
104519_c0_g1_i4	70	TBD
95748_c1_g1_i1	71	TBD
*At*UGT89B1_Group_B	72	OAP14423
106481_c2_g2_i1	73	TBD
106481_c2_g2_i2	74	TBD
GRMZM2G159404_P01_Group_M	75	NP_001148083

Sequence ID	Listed on Phylogeny	Accession Number
*At*UGT92A1_Group_M	76	Q9LXV0
101680_c0_g1_i1	77	TBD
110860_c1_g1_i2	78	TBD
110860_c1_g1_i10	79	TBD
110860_c1_g1_i3	80	TBD
110860_c1_g1_i9	81	TBD
194224_c0_g1_i1	82	TBD
98575_c0_g5_i1	83	TBD
109013_c5_g2_i2	84	TBD
109013_c5_g2_i3	85	TBD
111562_c4_g15_i1	86	TBD
111562_c4_g7_i5	87	TBD
108400_c0_g1_i1	88	TBD
108400_c0_g1_i2	89	TBD
*At*UGT91A1_Group_A	90	Q940V3
64576_c0_g1_i1	91	TBD
GRMZM2G061289_P01_Group_A	92	ONM36204
111438_c6_g4_i1	93	TBD
111438_c6_g9_i8	94	TBD
102407_c2_g3_i2	95	TBD
102407_c2_g3_i3	96	TBD
92712_c1_g1_i1	97	TBD
89445_c0_g1_i1	98	TBD
*At*UGT72C1_Group_E	99	O23205
109669_c0_g3_i1	100	TBD
98439_c0_g1_i1	101	TBD
109457_c2_g8_i11	102	TBD
109457_c2_g8_i9	103	TBD
85901_c1_g1_i1	104	TBD
GRMZM2G058314_P01_Group_E	105	NP_001149283
106678_c7_g12_i7	106	TBD
106678_c6_g1_i1	107	TBD
110940_c0_g2_i1	108	TBD
110940_c0_g2_i3	109	TBD
110632_c3_g7_i3	110	TBD
110632_c3_g6_i1	111	TBD
88334_c0_g1_i1	112	TBD
96927_c0_g1_i1	113	TBD

Sequence ID	Listed on Phylogeny	Accession Number
100212_c0_g1_i1	114	TBD
106688_c0_g3_i1	115	TBD
GRMZM2G120016_P01_Group_O	116	NP_001141165
Os_07g46610_Group_O	117	XP_015646369
106539_c5_g1_i1	118	TBD
110344_c2_g2_i3	119	TBD
110344_c2_g2_i9	120	TBD
110344_c2_g2_i6	121	TBD
110344_c1_g2_i1	122	TBD
106902_c2_g1_i2	123	TBD
106902_c2_g1_i1	124	TBD
106902_c2_g1_i5	125	TBD
101271_c0_g2_i1	126	TBD
*At*UGT78D3_Group_F	127	OAO94865
GRMZM2G022242_P01_Group_F	128	NP_001137065
103947_c0_g1_i1	129	TBD
105388_c0_g5_i1	130	TBD
106481_c1_g2_i1	131	TBD
106409_c0_g2_i1	132	TBD
106409_c0_g3_i2	133	TBD
*At*UGT75D1_Group_L	134	O23406
97222_c1_g6_i1	135	TBD
102411_c0_g3_i1	136	TBD
102411_c0_g3_i2	137	TBD
GRMZM2G050748_P01_Group_L	138	AQK83517
104689_c0_g1_i1	139	TBD
111335_c2_g1_i4	140	TBD
104749_c0_g2_i1	141	TBD
GRMZM2G173315_P01_Group_K	142	NP_001169578
*At*UGT86A1_Group_K	143	Q9SJL0
*At*UGT87A1_Group_J	144	O64732
100329_c0_g1_i1	145	TBD
105134_c0_g1_i1	146	TBD
86379_c0_g1_i1	147	TBD
GRMZM2G073376_P01_Group_J	148	NP_001148167
108899_c0_g1_i1	149	TBD
108899_c0_g1_i2	150	TBD
102764_c0_g1_i1	151	TBD

Sequence ID	Listed on Phylogeny	Accession Number
102764_c0_g1_i3	152	TBD
GRMZM2G021786_P01_Group_N	153	NP_001131410
*AtU*GT82A1_Group_N	154	Q9LHJ2
*At*UGT83A1_Group_I	155	Q9SGA8
GRMZM2G046994_P01_Group_I	156	AQK51585
105481_c1_g1_i1	157	TBD
107566_c1_g1_i1	158	TBD
*At*UGT76B1_Group_H	159	OAP05179
GRMZM5G892627_P02_Group_H	160	ONM52585
103372_c0_g1_i1	161	TBD
95258_c0_g1_i1	162	TBD
108074_c1_g1_i1	163	TBD
108074_c1_g1_i2	164	TBD
98996_c0_g1_i1	165	TBD
106992_c1_g2_i1	166	TBD
99562_c2_g1_i1	167	TBD
86843_c0_g1_i1	168	TBD
95258_c0_g1_i2	169	TBD
100141_c0_g1_i3	170	TBD
102420_c0_g1_i1	171	TBD
44324_c0_g1_i1	172	TBD
101539_c3_g3_i1	173	TBD
110738_c4_g1_i1	174	TBD
138206_c0_g1_i1	175	TBD
GRMZM5G834303_P01_Group_P	176	NP_001148991
Os_07g30690_Group_P	177	XP_015647802
*At*UGT85A4_Group_G	178	OAP13096
109680_c1_g1_i1	179	TBD
109680_c1_g1_i3	180	TBD
82487_c0_g1_i1	181	TBD
GRMZM2G041699_P01_Group_G	182	NP_001149205
103030_c0_g1_i1	183	TBD
94406_c0_g1_i1	184	TBD
109680_c2_g4_i1	185	TBD
88206_c0_g1_i2	186	TBD
94275_c0_g1_i1	187	TBD
109028_c2_g4_i5	188	TBD
109028_c2_g4_i8	189	TBD

Sequence ID	Listed on Phylogeny	Accession Number
99057_c3_g1_i1	190	TBD
99057_c3_g1_i2	191	TBD
105847_c0_g1_i1	192	TBD
105847_c0_g1_i2	193	TBD
105847_c0_g1_i5	194	TBD
*Zm*Anthocyanidin-3-O-glucosyltransferase	*Zm*UGT6	AQK87021
*Zm*Cytokinin-O-glucosyltransferase1	*Zm*UGT7	AQK94606
*Zm*UDP-glycosyltransferase-73B3	*Zm*UGT2	AQL08831
*Zm*Uncharacterized_protein_LOC100217043	*Zm*UGT5	NP_001136887
*Zm*Cytokinin-O-glucosyltransferase1	*Zm*UGT4	NP_001149762
*Zm*Flavonol-3-O-glycoside-7-O-glucosyltransferase	*Zm*UGT3	NP_001151546
*Zm*UDP-glycosyltransferase_73B3	*Zm*UGT1	ONM39936

Table S5.1: Percent total of individual monoterpenes of the product profiles produced by *Ps*TPS-3car1, *Ps*TPS-3car2, *Ps*TPS-3car3, *Ps*TPS-sab, and their variants.

		Percent Total of Individual Monoterpene Produced											
Variant #	**Enzyme Variant**	**(–)-α-pinene**	**α-thujene**	**(–)-β-pinene**	**(-)-sabinene**	**(+)-3-carene**	**α-phellandrene**	**myrcene**	**α-terpinene**	**(–)-limonene**	**(–)-β-phellandrene**	**γ-terpinene**	**α-terpinolene**
	***Ps*TPS-3car1**	3.31	0.50	0.47	8.73	**49.24**	0.64	4.20	0.74	0.77	5.48	1.21	**24.71**
	***Ps*TPS-3car2**	1.25	0.48	0.44	6.85	**67.52**	0.41	2.74	0.61	0.30	2.84	1.14	**15.43**
	***Ps*TPS-3car3**	0.00	0.00	0.44	8.76	**46.23**	0.00	8.31	0.62	4.64	1.26	0.00	**29.74**
	***Ps*TPS-sab**	4.88	1.99	2.14	**44.67**	1.29	0.54	0.65	1.66	0.62	3.24	2.42	**35.91**
1	***Ps*TPS-sab (helix A - helix K)**	1.68	0.38	0.58	6.94	**66.11**	0.36	2.71	0.59	0.39	3.17	1.09	**16.00**
2	***Ps*TPS-sab (helix A - helix E)**	3.67	0.39	1.19	7.48	**56.21**	0.51	3.45	0.57	0.36	4.85	1.16	**20.16**
3	***Ps*TPS-sab (helix F - helix $G_{1/2}$)**	9.25	0.98	3.57	7.91	**44.13**	1.16	3.19	1.02	1.04	9.75	1.84	**16.17**
4	***Ps*TPS-sab (helix H1 - helix I)**	9.54	1.36	3.94	7.67	**41.29**	0.95	4.61	1.11	0.55	10.36	1.92	**16.69**
5	***Ps*TPS-sab (helix J - helix K)**	7.91	0.80	3.02	9.46	**44.86**	0.56	1.91	0.97	0.45	8.74	1.32	**20.00**
6	***Ps*TPS-sab (helix J)**	8.92	0.60	3.84	9.20	**39.67**	0.45	3.93	0.78	1.02	9.67	1.10	**20.80**
34	***Ps*TPS-sab (J/K loop)**	4.70	2.00	1.88	**46.23**	1.31	0.50	0.37	1.56	0.51	3.03	2.35	**35.56**
35	***Ps*TPS-sab (helix K)**	4.70	1.82	2.01	**45.90**	1.24	0.49	0.50	1.50	0.64	3.08	2.30	**35.82**
7	***Ps*TPS-sab (S589A)**	4.48	1.56	1.80	**39.02**	2.38	1.35	1.25	3.92	0.82	4.64	5.52	**33.27**
8	***Ps*TPS-sab (A595G)**	5.66	2.06	2.59	**42.91**	1.19	0.62	0.28	2.10	0.59	3.57	2.95	**35.48**
9	***Ps*TPS-sab (F596L)**	6.49	1.15	2.38	18.74	**28.63**	0.62	5.95	1.17	0.85	7.52	1.55	**24.94**
10	***Ps*TPS-sab (L599F)**	4.68	1.52	2.27	**42.81**	1.47	0.64	0.21	2.36	0.69	3.62	3.30	**36.43**
36	***Ps*TPS-sab (S589A/F596L)**	7.76	0.98	2.57	16.25	**31.16**	1.01	6.14	1.53	0.94	9.02	2.16	**20.47**
37	***Ps*TPS-sab (A595G/F596L)**	7.23	1.04	2.52	13.29	**34.98**	0.95	6.75	1.11	1.08	8.52	2.07	**20.44**
38	***Ps*TPS-sab (F596L/L599F)**	6.71	0.70	2.54	17.04	**29.16**	0.63	6.95	1.06	0.93	7.29	1.49	**25.50**

		Percent Total of Individual Monoterpene Produced											
Variant #	**Enzyme Variant**	**(–)-α-pinene**	**α-thujene**	**(–)-β-pinene**	**(-)-sabinene**	**(+)-3-carene**	**α-phellandrene**	**myrcene**	**α-terpinene**	**(–)-limonene**	**(–)-β-phellandrene**	**γ-terpinene**	**α-terpinolene**
39	***Ps*TPS-sab (S589A/ A595G/F596L)**	7.75	0.93	2.87	14.25	**34.13**	0.92	5.85	1.05	0.93	9.91	1.89	**19.53**
40	***Ps*TPS-sab (S589A/ F596L/L599F)**	11.98	2.04	5.27	9.60	**25.86**	2.50	7.76	2.30	4.33	8.65	3.80	**15.93**
11	***Ps*TPS-sab (A595G/ F596L/L599F)**	8.67	0.70	3.11	7.28	**42.29**	0.75	5.42	1.06	0.98	7.84	1.84	**20.07**
15	***Ps*TPS-sab (helix A + helix J)**	10.10	1.25	3.99	8.04	**39.44**	0.70	5.03	0.75	0.80	11.08	1.17	**17.65**
12	***Ps*TPS-sab (helix A - helix E + helix J)**	3.67	0.39	0	1.19	7.48	56.21	0.51	3.45	0.57	0.36	4.85	1.16
13	***Ps*TPS-sab (helix F - helix $G_{1/2}$ + helix J)**	9.25	0.98	0	3.57	7.91	44.13	1.16	3.19	1.02	1.04	9.75	1.84
14	***Ps*TPS-sab (helix H1 - helix I + helix J)**	9.54	1.36	0	3.94	7.67	41.29	0.95	4.61	1.11	0.55	10.36	1.92
15	***Ps*TPS-sab (helix A + helix J)**	10.10	1.25	3.99	8.04	**39.44**	0.70	5.03	0.75	0.80	11.08	1.17	**17.65**
16	***Ps*TPS-sab (A/C loop + helix J)**	8.71	0.97	3.19	5.40	**35.57**	1.59	5.55	2.51	1.32	12.09	2.62	**20.48**
17	***Ps*TPS-sab (helix C + helix J)**	6.96	0.86	2.27	5.16	**45.36**	1.29	3.17	2.41	1.25	8.37	2.70	**20.20**
18	***Ps*TPS-sab (helix D + helix J)**	5.77	0.74	2.14	4.09	**50.83**	1.12	4.09	1.98	1.07	6.23	2.39	**19.55**
19	***Ps*TPS-sab (helix D_1 - helix D_2 + helix J)**	7.68	0.40	3.05	8.18	**45.40**	0.53	3.87	0.67	0.75	9.00	1.09	**19.37**
20	***Ps*TPS-sab (helix E + helix J)**	8.25	0.26	2.82	8.20	**43.80**	0.61	3.56	0.70	0.83	10.68	1.06	**19.23**
21	***Ps*TPS-3car1 (helix J)**	4.54	2.39	0.33	**23.93**	1.36	0.98	1.80	0.75	2.29	3.90	1.32	**56.39**
22	***Ps*TPS-3car1 (L596F)**	6.28	1.65	0.42	**20.88**	5.03	0.45	2.31	1.38	2.74	2.85	2.51	**53.49**

		Percent Total of Individual Monoterpene Produced											
Variant #	**Enzyme Variant**	**(–)-α-pinene**	**α-thujene**	**(–)-β-pinene**	**(-)-sabinene**	**(+)-3-carene**	**α-phellandrene**	**myrcene**	**α-terpinene**	**(–)-limonene**	**(–)-β-phellandrene**	**γ-terpinene**	**α-terpinolene**
23	***Ps*TPS-3car1 (A595G/ L596F/ F599L)**	5.25	2.53	0.39	**23.28**	1.84	0.87	2.01	0.97	2.31	3.77	1.62	**55.16**
24	***Ps*TPS-3car2 (helix J)**	2.15	2.63	0.54	**47.88**	4.01	0.38	1.00	0.88	0.42	1.48	2.27	**36.37**
25	***Ps*TPS-3car2 (L596F)**	2.64	1.93	0.89	**37.36**	12.27	0.50	0.94	1.56	0.79	1.87	3.89	**35.36**
26	***Ps*TPS-3car2 (A595G/ L596F/ F599L)**	2.16	2.62	0.56	**47.37**	4.66	0.38	1.24	1.17	0.44	1.65	2.51	**35.22**
27	***Ps*TPS-3car3 (helix J)**	0.00	0.00	0.00	**37.70**	8.43	0.00	21.80	3.89	0.35	4.01	0.00	**23.82**
28	***Ps*TPS-3car3 (L596F)**	0.00	0.00	0.00	20.23	9.23	0.00	**25.44**	3.53	7.35	1.96	0.00	**32.26**
29	***Ps*TPS-3car3 (A595G/ L596F/ F599L)**	0.00	0.00	0.00	**29.01**	5.39	0.00	21.40	1.92	11.32	4.60	0.00	**26.36**
30	***Ps*TPS-sab (F596E)**	1.73	0.34	0.22	10.21	0.45	0.21	6.60	0.68	**70.98**	2.06	0.85	5.68
31	***Ps*TPS-sab (F596H)**	4.63	1.03	1.43	**46.46**	2.06	0.61	4.65	0.54	2.91	7.54	1.00	**27.13**
32	***Ps*TPS-sab (F596R)**	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
33	***Ps*TPS-sab (F596G)**	4.40	0.91	1.35	11.22	9.97	0.40	**26.95**	1.86	17.01	6.78	1.87	**17.27**

Table S5.2: Primers used for domain swaps and site directed mutagenesis.

Variant Number	Enzyme Variant Used For	Primer
1	*Ps*TPS-sab (Helix A – K)	Used mega primer created from 5′-CATAGGCACAGGTTCAAGAAC-3′ and 3′-TCCAAAAAACATGCTTTTGAC-5′
2	*Ps*TPS-sab (Helix A – E)	Used mega primer created from 5′-AATCAAATGCTAGATGTGAATACCGAGAAAGTTTTAGAACTTGCAAAA-3′ and 3′-ATAGTTGAGCGTGTCTCGGCCTTGTGACTT-5′
3	*Ps*TPS-sab (Helix F – $G_{1/2}$)	Used mega primer created from 5′-TCACAAGGCCGAGACACGCTCAACTAT-3′ and 3′-GATGTGATGAGGAAAGGGAATGTCCAA-5′
4	*Ps*TPS-sab (Helix H – I)	Used mega primer created from 5′-TTGGACATTCCCTTTCCTCATCACATC-3′ and 3′-GGAAGAGATTGGAACGTTGTTGTCAGGTTTAAGAAGCTC-5′
5	*Ps*TPS-sab (Helix J – K)	Used mega primer created from 5′-GATGTGAATACCGAGAAAGTT-3′ and 3′-TCCAAAAAACATGCTTTTGAC-5′
6	*Ps*TPS-sab (Helix J)	5′-ATCTCTTCCAAGAAGCATGCTTTTGACATAAGCAGAGGTTTGCATCACTTCTACAGGTACCGAGAT-3′
34	*Ps*TPS-sab (J/K Loop)	5′-AGAGCTTTTCATCACCTCTACAATTATCGAGATGGCTACACTGTTGCCAGCAATGAAACAAAGAAT-3′
35	*Ps*TPS-sab (Helix K)	5′-GATGGCTACACAGTTTCCAGCAATGAAACAAAGAATTTGGTGATAAAAACCGTTCTTGAACCTCTCCCTATGTAA-3′
7	*Ps*TPS-sab (S589A)	5′-TCTTCCAAGAAGCATGCTTTTGACATAAGTAGA-3′
8	*Ps*TPS-sab (A595G)	5′-TTTGACATAAGTAGAGGTTTTCATCACCTCTAC-3′
9	*Ps*TPS-sab (F596L)	5′-GACATAAGTAGAGCTTTGCATCACCTCTACAGG-3′
10	*Ps*TPS-sab (L599F)	5′-AGAGCTTTTCATCACTTCTACAGGTACCGAGAT-3′
36	*Ps*TPS-sab (S589A|F596L)	5′-ATCTCTTCCAAGAAGCATGCTTTTGACATAAGTAGAGCTTTGCATCACCTCTACAGGTAC-3′
37	*Ps*TPS-sab (A595G|F596L)	5′-TCTTTTGACATAAGTAGAGGTTTGCATCACCTCTACAGGTAC-3′
38	*Ps*TPS-sab (F596L|L599F)	5′-GACATAAGTAGAGCTTTGCATCACTTCTACAGGTACCGAGAT-3′
39	*Ps*TPS-sab (S589A|A595G|F596L)	5′-ATCTCTTCCAAGAAGCATGCTTTTGACATAAGTAGAGGTTTGCATCACCTCTACAGGTAC-3′
40	*Ps*TPS-sab (S589A|F596L|L599F)	5′-ATCTCTTCCAAGAAGCATGCTTTTGACATAAGTAGAGCTTTGCATCACTTCTACAGGTACCGAGATGGC-3′
11	*Ps*TPS-sab (A595G|F596L|L599F)	5′-TCTTTTGACATAAGTAGAGGTTTGCATCACTTCTACAGGTACCGAGATGGC-3′
N/A	*Ps*TPS-sab (Helix A)	Used mega primer created from 5′-AATCAAATGCTAGATGTGAAT-3′ and 3′-ACCTGAATCTTTCCACCATCT-5′

Variant Number	Enzyme Variant Used For	Primer
N/A	*Ps*TPS-sab (A/C Loop)	5′-TGGTGGAAAGAATCAGGTTTCCCTGATCTAAACTTCATTCGACATCGTCAC-3′
N/A	*Ps*TPS-sab (Helix C)	5′-TTCATTCGACATCGTCACGTGGAATTCTACACTTTGGTCTCTGGAATTGACATGGAG-3′
N/A	*Ps*TPS-sab (Helix D)	5′-GACATGGAGCCTAAACATTCTACATTCAGATTGAGCTTTGTCAAAATG-3′ and 5′-CTTATCACAGTTTTGGATGATATGTATGACACCTTTGGAACAATGG-3′
N/A	*Ps*TPS-sab (Helix D_1 – D_2)	5′-GACACCTTTGGAACAATAGACGAGCTCCGACTCTTCACAGCTGCAGTTAAGAGATGGGAT-3′ and 5′-GCAGTTAAGAGATGGGATCCGTCGACAACTCAGTGTCTTCCAGAATATATGAAAGGTG-3′
N/A	*Ps*TPS-sab (Helix E)	5′-CAGAATATATGAAAGGTGTTTATATAGTTCTTTATGAAACCGTAAATGAAATGGCGAAAGAGGCACAAAAGTCACAAGGCCGAGAC-3′
21	*Ps*TPS-3car1 (Helix J)	5′-AATCTCTTCCAAGAAACATTCTTTTAACATAAGCAGAGCTTTTCACCACCTCTACAAATACCGAGATGG-3′
22	*Ps*TPS-3car1 (L596F)	5′-CTTTTAACATAAGCAGAGGTTTTCACCACTTCTACAAATAC-3′
23	*Ps*TPS-3car1 (G595A|L596F|F599L)	5′-GCTTTTAACATAAGCAGAGCTTTTCACCACCTCTACAAATACCGAGATGG-3′
24	*Ps*TPS-3car2 (Helix J)	5′-ATCTCTTCCAAAAAACATTCTTTTGACATAAGCAGAGCTTTTCATCACCTCTACAATTATCGAGATGGC-3′
25	*Ps*TPS-3car2 (L596F)	5′-TTTGACATAAGCAGAGGTTTTCATCACTTCTACAATTAT-3′
26	*Ps*TPS-3car2 (G595A|L596F|F599L)	5′-GCTTTTGACATAAGCAGAGCTTTTCATCACCTCTACAATTATCGAGATGGC-3′
27	*Ps*TPS-3car3 (Helix J)	5′-ATCTCTTCCAAGAAACCTTCTTTTGACATAAGCAGAGCTTTTCATCACCTCTACAATTATCGTGATGGC-3′
28	*Ps*TPS-3car3 (L596F)	5′-CTTTTGACATAAGCAGAGGTTTTCATCACTTCTACAATTATC-3′
29	*Ps*TPS-3car3 (G595A|L596F|F599L)	5′-GCTTTTGACATAAGCAGAGCTTTTCATCACCTCTACAATTATCGTGATGGC-3′
30	*Ps*TPS-sab (F596E)	5′-CTTTTGACATAAGTAGAGCTGAACATCACCTCTACAGGTACCGAG-3′
31	*Ps*TPS-sab (F596H)	5′-CTTTTGACATAAGTAGAGCTCATCATCACCTCTACAGGTACCGAG-3′
32	*Ps*TPS-sab (F596R)	5′-CTTTTGACATAAGTAGAGCTCGTCATCACCTCTACAGGTACCGAG-3′
33	*Ps*TPS-sab (F596G)	5′-CTTTTGACATAAGTAGAGCTGGTCATCACCTCTACAGGTACCGAG-3′

SUPPLEMENTARY FIGURES

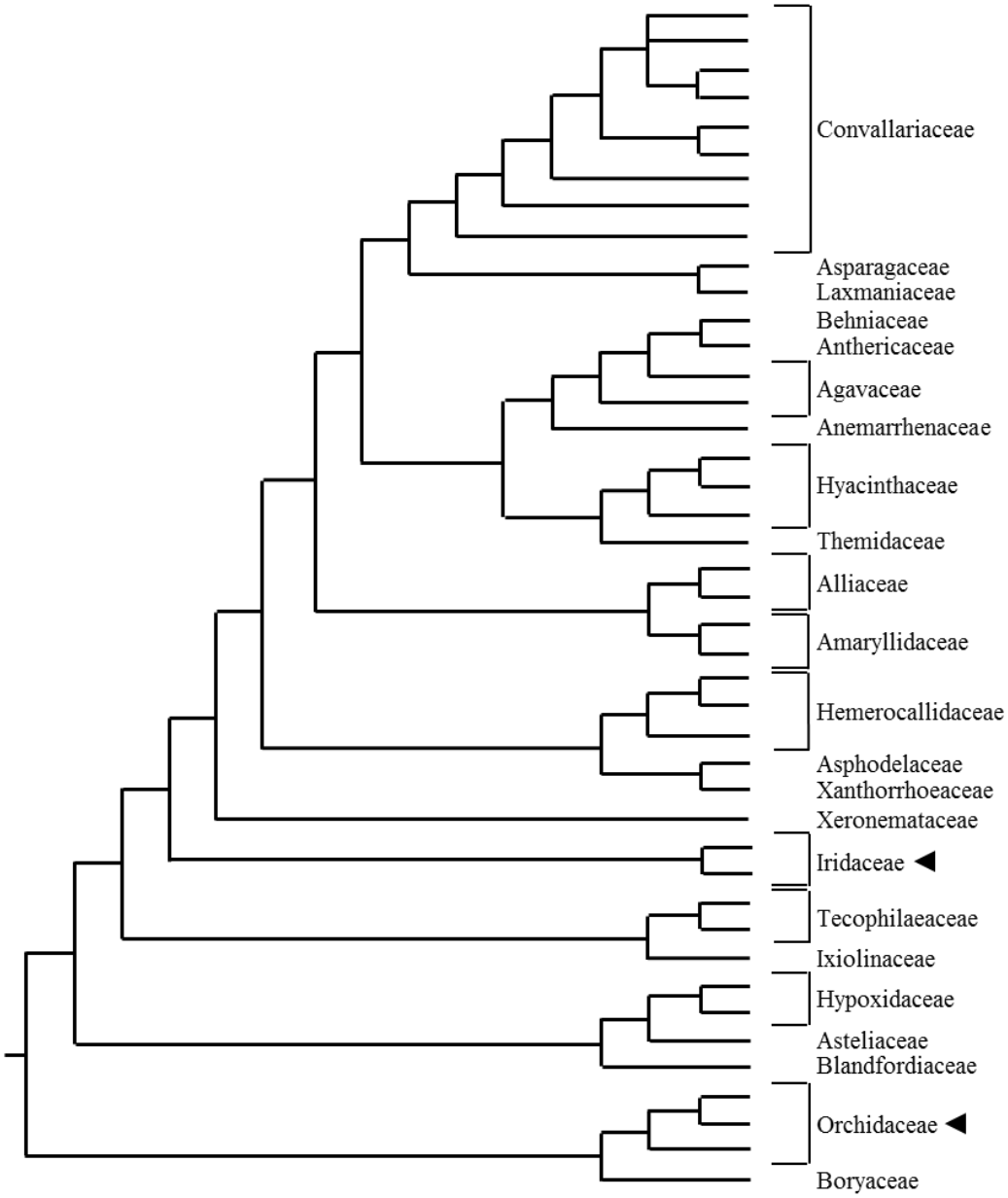


Figure S2.1. Phylogenetic tree of the Asparagales order adapted from Fay *et al.* (2000). Families within the order are listed to the right of the tree. Arrowheads indicate the Iridaceae and Orchidaceae families.

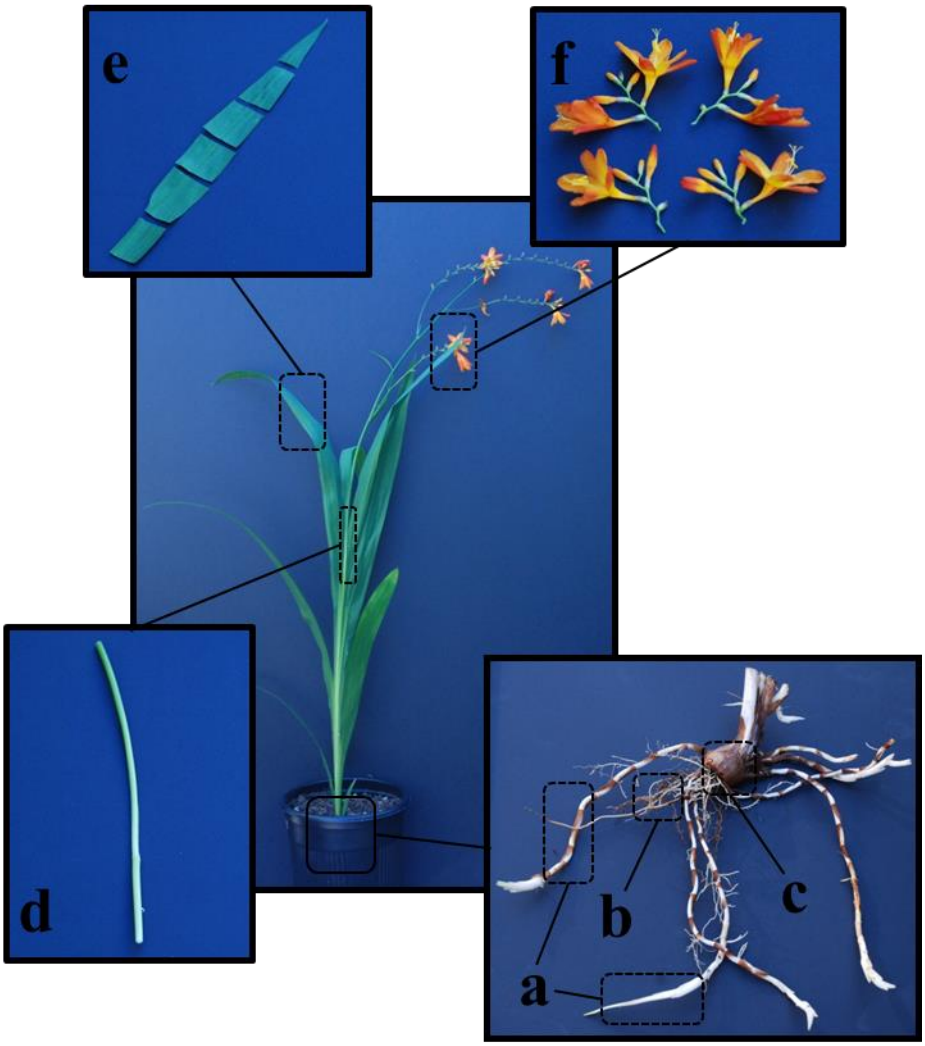


Figure S2.2: Organs of *Crocosmia x crocosmiiflora*. (a) stolon, (b) root, (c) corm, (d) stem, (e) leaf, and (f) flower.

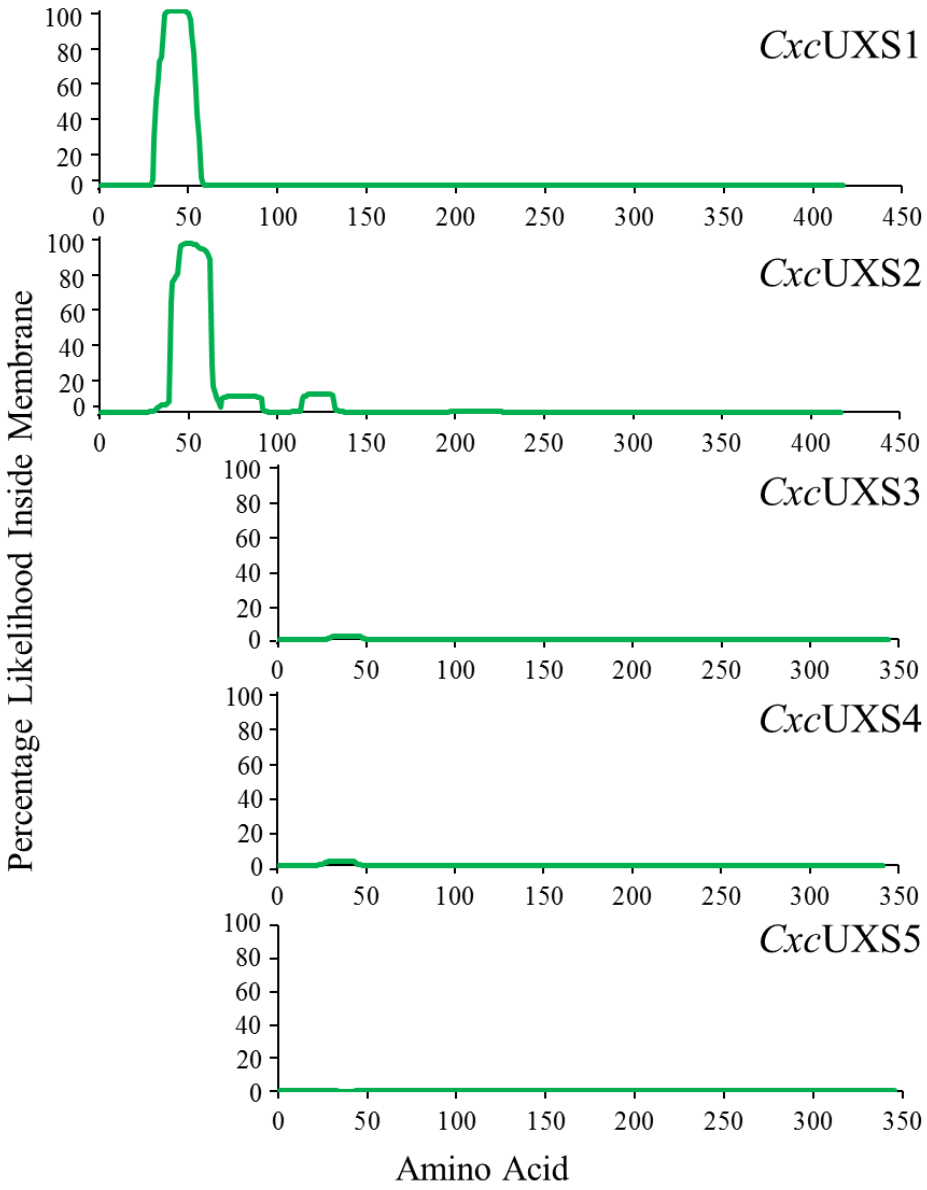
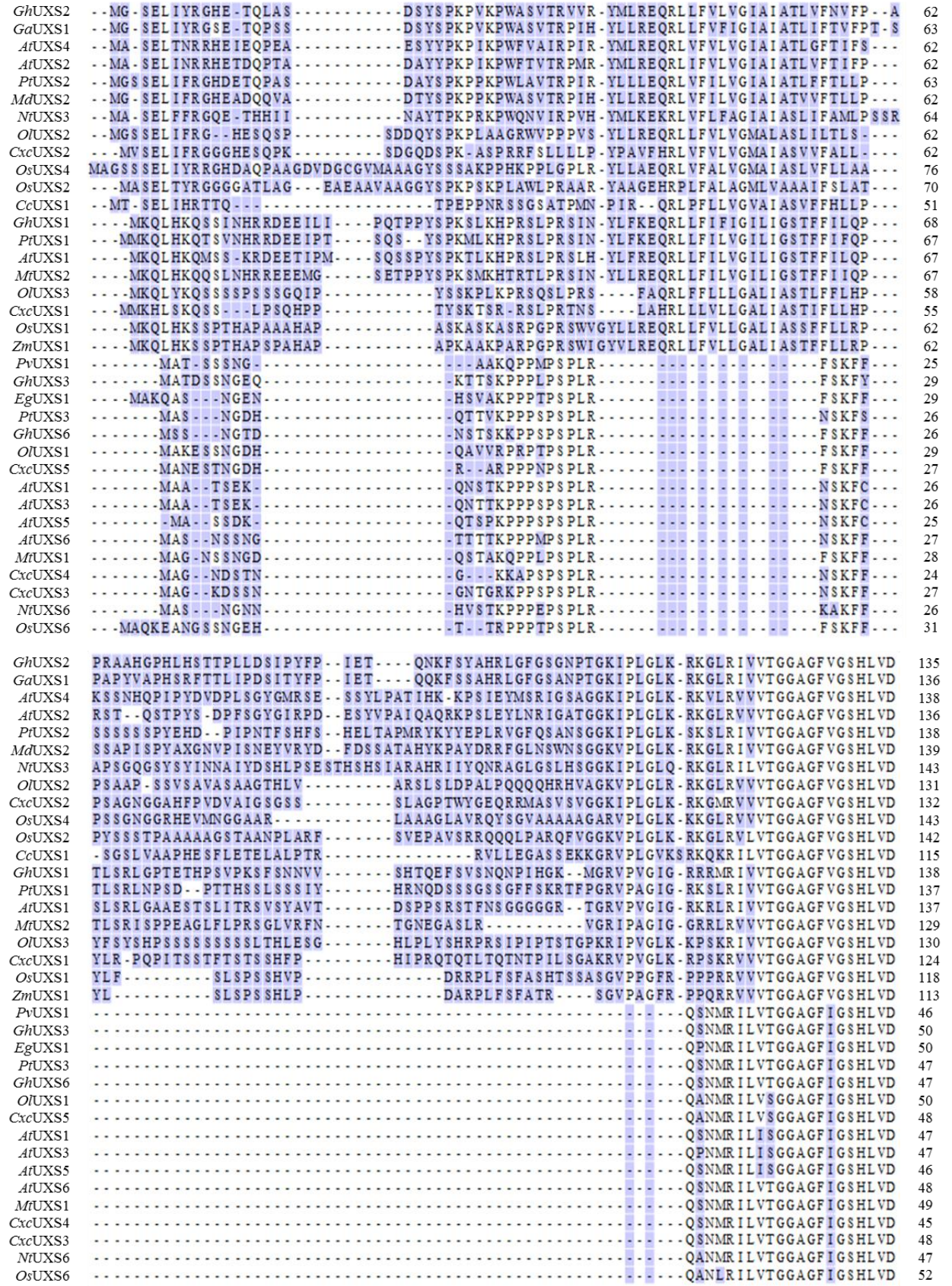
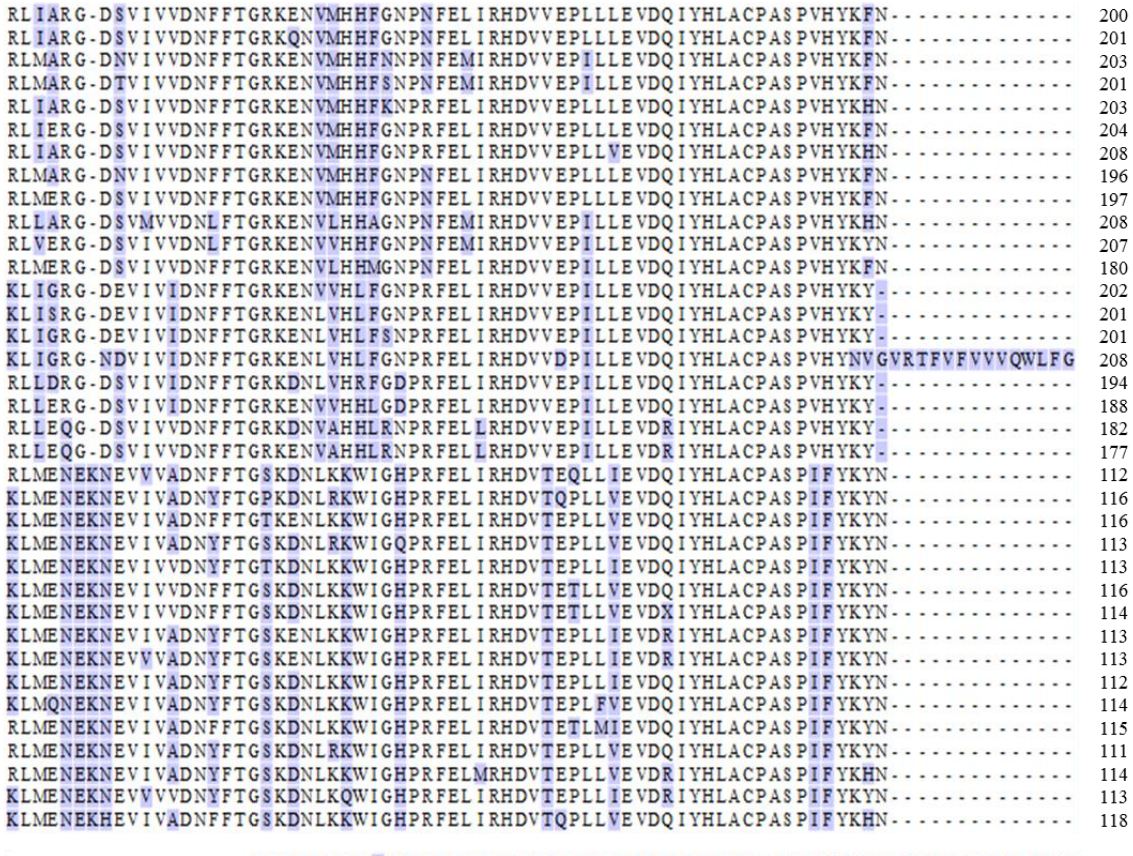
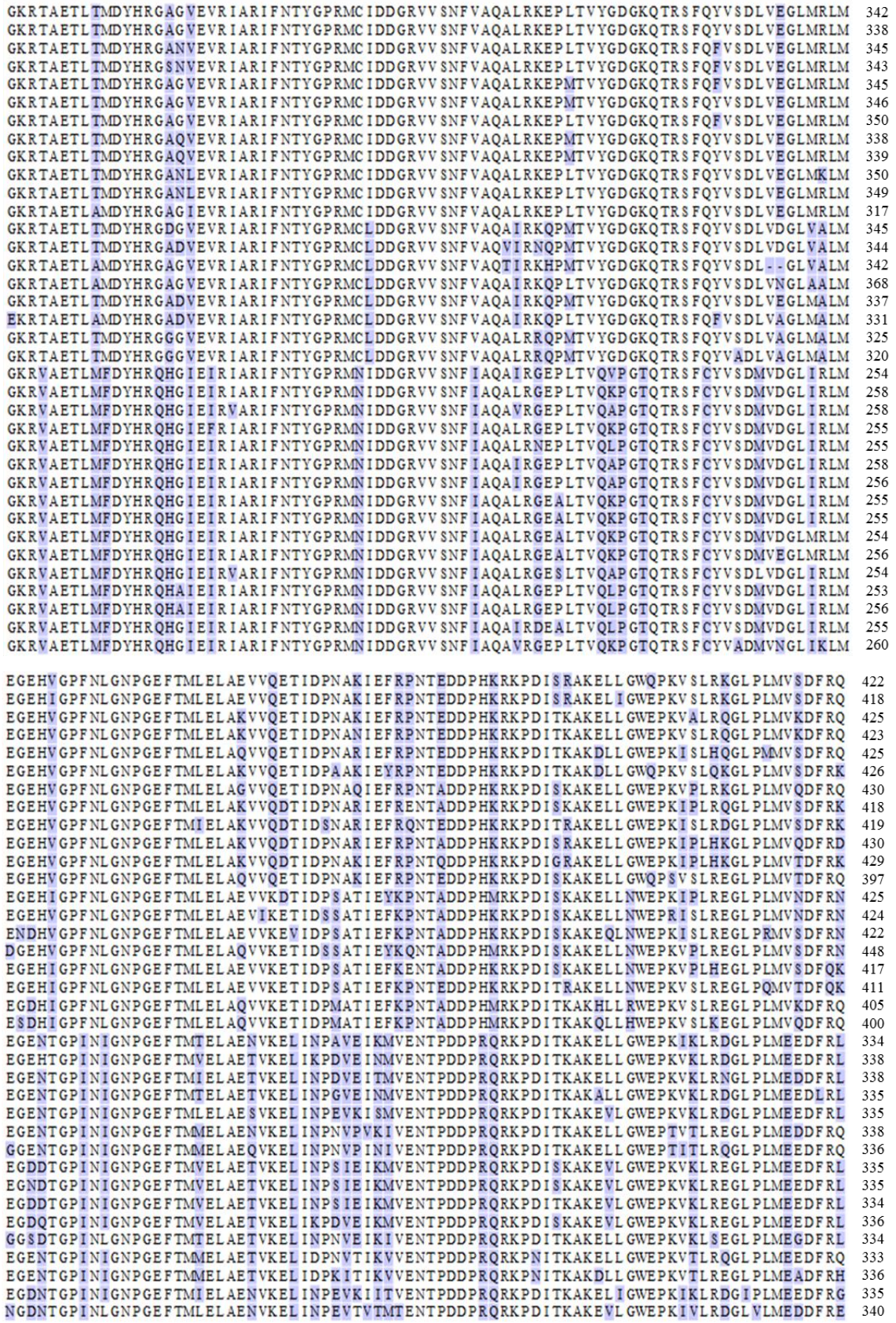


Figure S3.1: Hydropathy plots of *Cxc*UXS1 – 5. Plots show *Cxc*UXS1 and 2 appear to have a variable N-terminal region that spans amino acids 1 – 30 and 1 – 39 respectively and a 22 amino acid long hydrophobic transmembrane domain. *Cxc*UXS3, 4, and 5 show no hydrophobic transmembrane domain.




GhUXS2 RLIARG-DSVIVVDNFFTGRKENVMHHFGNPNFELIRHDVVEPLLLEVDQIYHLACPASPVHYKFN- 200
GaUXS1 RLIARG-DSVIVVDNFFTGRKQNVMHHFGNPNFELIRHDVVEPLLLEVDQIYHLACPASPVHYKFN- 201
AtUXS4 RLMARG-DNVIVVDNFFTGRKENVMHHFNNPNFEMIRHDVVEPILLEVDQIYHLACPASPVHYKFN- 203
AtUXS2 RLMARG-DTVIVVDNFFTGRKENVMHHFSNPNFEMIRHDVVEPILLEVDQIYHLACPASPVHYKFN- 201
PtUXS2 RLIARG-DSVIVVDNFFTGRKENVMHHFKNPRFELIRHDVVEPLLLEVDQIYHLACPASPVHYKHN- 203
MdUXS2 RLIERG-DSVIVVDNFFTGRKENVMHHFGNPRFELIRHDVVEPLLLEVDQIYHLACPASPVHYKFN- 204
NtUXS3 RLIARG-DSVIVVDNFFTGRKENVMHHFGNPRFELIRHDVVEPLLVEVDQIYHLACPASPVHYKHN- 208
OlUXS2 RLMARG-DNVIVVDNFFTGRKENVMHHFGNPNFELIRHDVVEPLLLEVDQIYHLACPASPVHYKFN- 196
CxcUXS2 RLMERG-DSVIVVDNFFTGRKENVMHHFGNPNFELIRHDVVEPLLLEVDQIYHLACPASPVHYKFN- 197
OsUXS4 RLLARG-DSVMVVDNLFTGRKENVLHHAGNPNFEMIRHDVVEPILLEVDQIYHLACPASPVHYKHN- 208
OsUXS2 RLVERG-DSVIVVDNLFTGRKENVVHHFGNPNFEMIRHDVVEPILLEVDQIYHLACPASPVHYKYN- 207
CcUXS1 RLMERG-DSVIVVDNFFTGRKENVLHHMGNPNFELIRHDVVEPILLEVDQIYHLACPASPVHYKFN- 180
GhUXS1 KLIGRG-DEVIVIDNFFTGRKENVVHLFGNPRFELIRHDVVEPILLEVDQIYHLACPASPVHYKY- 202
PtUXS1 KLISRG-DEVIVIDNFFTGRKENLVHLFGNPRFELIRHDVVEPILLEVDQIYHLACPASPVHYKY- 201
AtUXS1 KLIGRG-DEVIVIDNFFTGRKENLVHLFSNPRFELIRHDVVEPILLEVDQIYHLACPASPVHYKY- 201
MtUXS2 KLIGRG-NDVIVIDNFFTGRKENLVHLFGNPRFELIRHDVVDPILLEVDQIYHLACPASPVHYNVGVRTFVFVVVQWLFG 208
OlUXS3 RLLDRG-DSVIVIDNFFTGRKDNLVHRFGDPRFELIRHDVVEPILLEVDQIYHLACPASPVHYKY- 194
CxcUXS1 RLLERG-DSVIVIDNFFTGRKENVVHHLGDPRFELIRHDVVEPILLEVDQIYHLACPASPVHYKY- 188
OsUXS1 RLLEQG-DSVIVVDNFFTGRKDNVAHHLRNPRFELLRHDVVEPILLEVDRIYHLACPASPVHYKY- 182
ZmUXS1 RLLEQG-DSVIVVDNFFTGRKENVAHHLRNPRFELLRHDVVEPILLEVDRIYHLACPASPVHYKY- 177
PvUXS1 RLMENEKNEVVVADNFFTGSKDNLKKWIGHPRFELIRHDVTEQLLIEVDQIYHLACPASPIFYKYN- 112
GhUXS3 KLMENEKNEVIVADNYFTGPKDNLRKWIGHPRFELIRHDVTQPLLVEVDQIYHLACPASPIFYKYN- 116
EgUXS1 KLMENEKNEVIVADNFFTGTKENLKKWIGHPRFELIRHDVTEPLLVEVDQIYHLACPASPIFYKYN- 116
PtUXS3 KLMENEKNEVIVADNYFTGSKDNLRKWIGQPRFELIRHDVTEPLLVEVDQIYHLACPASPIFYKYN- 113
GhUXS6 KLMENEKNEVIVVDNYFTGTKDNLKKWIGHPRFELIRHDVTEPLLIEVDQIYHLACPASPIFYKYN- 113
OlUXS1 KLMENEKNEVIVVDNFFTGSKDNLKKWIGHPRFELIRHDVTETLLVEVDQIYHLACPASPIFYKYN- 116
CxcUXS5 KLMENEKNEVIVVDNFFTGSKDNLKKWIGHPRFELIRHDVTETLLVEVDXIYHLACPASPIFYKYN- 114
AtUXS1 KLMENEKNEVIVADNYFTGSKENLKKWIGHPRFELIRHDVTEPLLIEVDRIYHLACPASPIFYKYN- 113
AtUXS3 KLMENEKNEVVVADNYFTGSKENLKKWIGHPRFELIRHDVTEPLLIEVDRIYHLACPASPIFYKYN- 113
AtUXS5 KLMENEKNEVIVADNYFTGSKDNLKKWIGHPRFELIRHDVTEPLLIEVDQIYHLACPASPIFYKYN- 112
AtUXS6 KLMQNEKNEVIVADNYFTGSKDNLKKWIGHPRFELIRHDVTEPLFVEVDQIYHLACPASPIFYKYN- 114
MtUXS1 RLMENEKNEVIVADNFFTGSKDNLKKWIGHPRFELIRHDVTETLMIEVDQIYHLACPASPIFYKYN- 115
CxcUXS4 RLMENEKNEVIVADNYFTGSKDNLRKWIGHPRFELIRHDVTEPLLVEVDQIYHLACPASPIFYKYN- 111
CxcUXS3 RLMENEKNEVIVADNYFTGSKDNLKKWIGHPRFELMRHDVTEPLLVEVDRIYHLACPASPIFYKHN- 114
NtUXS6 KLMENEKNEVVVVDNYFTGSKDNLKQWIGHPRFELIRHDVTEPLLIEVDRIYHLACPASPIFYKYN- 113
OsUXS6 KLMENEKHEVIVADNFFTGSKDNLKKWIGHPRFELIRHDVTQPLLVEVDQIYHLACPASPIFYKHN- 118
GhUXS2 ----------PVKTIKTNVVGTLNMLGLAKRVGARFLLTSTSEVYGDPLQHPQKETYWGNVNPIGVRSCYDE 262
GaUXS1 ----------P-----TNVVGTLNMLGLAKRVGARFLLTSTSEVYGDPLQHPQKETYWGNVNPIGVRSCYDE 258
AtUXS4 ----------PVKTIKTNVVGTLNMLGLAKRVGARFLLTSTSEVYGDPLQHPQVETYWGNVNPIGVRSCYDE 265
AtUXS2 ----------PVKTIKTNVVGTLNMLGLAKRVGARFLLTSTSEVYGDPLQHPQVETYWGNVNPIGVRSCYDE 263
PtUXS2 ----------PVKTIKTNVVGTLNMLGLAKRVGARFLLTSTSEVYGDPLQHPQVETYWGNVNPIGVRSCYDE 265
MdUXS2 ----------PVKTIKTNVVGTLNMLGLAKRIGARFLLTSTSEVYGDPLQHPQVETYWGNVNPIGVRSCYDE 266
NtUXS3 ----------PVKTIKTNVVGTLNMLGLAKRVGARFLLTSTSEVYGDPLQHPQKETYWGNVNPIGVRSCYDE 270
OlUXS2 ----------PVKTIKTNVVGTLNMLGLAKRVGARFLLTSTSEVYGDPLQHPQVETYWGNVNPIGVRSCYDE 258
CxcUXS2 ----------PVKTIKTNVVGTLNMLGLAKRVGARFLLTSTSEVYGDPLQHPQVETYWGNVNPIGVRSCYDE 259
OsUXS4 ----------PVKTIKTNVVGTLNMLGLAKRVGARFLLTSTSEVYGDPLQHPQVETYWGNVNPIGVRSCYDE 270
OsUXS2 ----------PVKTIKTNVVGTLNMLGLAKRINARFLLTSTSEVYGDPLQHPQVETYWGNVNPIGVRSCYDE 269
CcUXS1 ----------P-----TNVVGTLNMLGLAKRIGARFLLTSTSEVYGDPLQHPQAETYWGNVNPIGVRSCYDE 237
GhUXS1 ---------NP--------VKTIKTNVMGTLNMLGLAKRVGARFLLTSTSEVYGDPLQHPQKETYWGNVNPIGERSCYDE 265
PtUXS1 ---------NP--------VKTIKTNVMGTLNMLGLAKRVGARFLLTSTSEVYGDPLEHPQKETYWGNVNPIGERSCYDE 264
AtUXS1 ---------NP--------VKTIKTNVMGTLNMLGLAKRVGARFLLTSTSEVYGDPLEHPQKETYWGNVNPIGERSCYDE 264
MtUXS2 ISNFVIGFLDPSLYYVIGDLQDKCTNVMGTLNMLGLAKRIGARFLLTSTSEVYGDPLEHPQKETYWGNVNPIGERSCYDE 288
OlUXS3 ---------NP--------IKTIKTNVMGTLNMLGLAKRIGARFLLTSTSEVYGDPLEHPQKETYWGHVNPIGVRSCYDE 257
CxcUXS1 ---------NP--------IKTIKTNVMGTLNMLGLAKRIGARFLLTSTSEVYGDPLEHPQKETYWGHVNPIGVRSCYDE 251
OsUXS1 ---------NP--------IKTIKTNVMGTLNMLGLAKRIGARFLLTSTSEVYGDPLEHPQKETYWGHVNPIGVRSCYDE 245
ZmUXS1 ---------NP--------IKTIKTNVMGTLNMLGLAKRIGARFLLTSTSEVYGDPLEHPQKESYWGHVNPIGVRSCYDE 240
PvUXS1 ----------PVKTIKTNVIGTLNMLGLAKRVGARILLTSTSEVYGDPLVHPQPESYWGNVNPIGVRSCYDE 174
GhUXS3 ----------PVKTIKTNVIGTLNMLGLAKRVGARILLTSTSEVYGDPLIHPQPETYWGNVNPIGVRSCYDE 178
EgUXS1 ----------PVKTIKTNVIGTLNMLGLAKRVGARILLTSTSEVYGDPLVHPQDESYWGNVNPIGVRSCYDE 178
PtUXS3 ----------PVKTIKTNVIGTLNMLGLAKRVGARILLTSTSEVYGDPLVHPQPESYWGNVNPIGVRSCYDE 175
GhUXS6 ----------PVKTIKTNVIGTLNMLGLAKRVGARILLTSTSEVYGDPLEHPQTESYWGNVNPIGVRSCYDE 175
OlUXS1 ----------PVKTIKTNVIGTLNMLGLAKRVGARILLTSTSEVYGDPLQHPQTEEYWGNVNPIGVRSCYDE 178
CxcUXS5 ----------PVKTIKTNVIGTLNMLGLAKRVGARILLTSTSEVYGDPLEHPQTEAYWGNVNPIGVRSCYDE 176
AtUXS1 ----------PVKTIKTNVIGTLNMLGLAKRVGARILLTSTSEVYGDPLIHPQPESYWGNVNPIGVRSCYDE 175
AtUXS3 ----------PVKTIKTNVIGTLNMLGLAKRVGARILLTSTSEVYGDPLIHPQPESYWGNVNPIGVRSCYDE 175
AtUXS5 ----------PVKTIKTNVIGTLNMLGLAKRVGARILLTSTSEVYGDPLIHPQPESYWGNVNPIGVRSCYDE 174
AtUXS6 ----------PVKTIKTNVIGTLNMLGLAKRVGARILLTSTSEVYGDPLVHPQTESYWGNVNPIGVRSCYDE 176
MtUXS1 ----------PVK---TNVIGTLNMLGLAKRVGARILLTSTSEVYGDPLEHPQPESYWGNVNPIGVRSCYDE 174
CxcUXS4 ----------PVKTIKTNVIGTLNMLGLAKRVGARILLTSTSEVYGDPLEHPQNEEYWGNVNPIGVRSCYDE 173
CxcUXS3 ----------PVKTIKTNVIGTLNMLGLAKRVGARILLTSTSEVYGDPLEHPQTEEYWGNVNPIGVRSCYDE 176
NtUXS6 ----------PVKTIKTNVLGTMNMLGLAKRTGARILLTSTSEVYGDPLVHPQDESYWGNVNPIGVRSCYDE 175
OsUXS6 ----------PVKTIKTNVIGTLNMLGLAKRVGARILLTSTSEVYGDPLEHPQTEAYWGNVNPIGVRSCYDE 180


GhUXS2 GKRTAETLTMDYHRGAGVEVRIARIFNTYGPRMCIDDGRVVSNFVAQALRKEPLTVYGDGKQTRSFQYVSDLVEGLMRLM 342
GaUXS1 GKRTAETLTMDYHRGAGVEVRIARIFNTYGPRMCIDDGRVVSNFVAQALRKEPLTVYGDGKQTRSFQYVSDLVEGLMRLM 338
AtUXS4 GKRTAETLTMDYHRGANVEVRIARIFNTYGPRMCIDDGRVVSNFVAQALRKEPLTVYGDGKQTRSFQFVSDLVEGLMRLM 345
AtUXS2 GKRTAETLTMDYHRGSNVEVRIARIFNTYGPRMCIDDGRVVSNFVAQALRKEPLTVYGDGKQTRSFQFVSDLVEGLMRLM 343
PtUXS2 GKRTAETLTMDYHRGAGVEVRIARIFNTYGPRMCIDDGRVVSNFVAQALRKEPMTVYGDGKQTRSFQFVSDLVEGLMRLM 345
MdUXS2 GKRTAETLTMDYHRGAGVEVRIARIFNTYGPRMCIDDGRVVSNFVAQALRKEPMTVYGDGKQTRSFQYVSDLVEGLMRLM 346
NtUXS3 GKRTAETLTMDYHRGAGVEVRIARIFNTYGPRMCIDDGRVVSNFVAQALRKEPLTVYGDGKQTRSFQFVSDLVEGLMRLM 350
OlUXS2 GKRTAETLTMDYHRGAQVEVRIARIFNTYGPRMCIDDGRVVSNFVAQALRKEPMTVYGDGKQTRSFQYVSDLVEGLMRLM 338
CxcUXS2 GKRTAETLTMDYHRGAQVEVRIARIFNTYGPRMCIDDGRVVSNFVAQALRKEPMTVYGDGKQTRSFQYVSDLVEGLMRLM 339
OsUXS4 GKRTAETLTMDYHRGANLEVRIARIFNTYGPRMCIDDGRVVSNFVAQALRKEPLTVYGDGKQTRSFQYVSDLVEGLMKLM 350
OsUXS2 GKRTAETLTMDYHRGANLEVRIARIFNTYGPRMCIDDGRVVSNFVAQALRKEPLTVYGDGKQTRSFQYVSDLVEGLMRLM 349
CcUXS1 GKRTAETLAMDYHRGAGIEVRIARIFNTYGPRMCIDDGRVVSNFVAQALRKEPLTVYGDGKQTRSFQYVSDLVEGLMRLM 317
GhUXS1 GKRTAETLTMDYHRGDGVEVRIARIFNTYGPRMCLDDGRVVSNFVAQAIRKQPMTVYGDGKQTRSFQYVSDLVDGLVALM 345
PtUXS1 GKRTAETLTMDYHRGADVEVRIARIFNTYGPRMCLDDGRVVSNFVAQVIRNQPMTVYGDGKQTRSFQYVSDLVDGLVALM 344
AtUXS1 GKRTAETLAMDYHRGAGVEVRIARIFNTYGPRMCLDDGRVVSNFVAQTIRKHPMTVYGDGKQTRSFQYVSDL--GLVALM 342
MtUXS2 GKRTAETLAMDYHRGAGVEVRIARIFNTYGPRMCLDDGRVVSNFVAQAIRKQPLTVYGDGKQTRSFQYVSDLVNGLAALM 368
OlUXS3 GKRTAETLTMDYHRGADVEVRIARIFNTYGPRMCLDDGRVVSNFVAQAIRKQPMTVYGDGKQTRSFQYVSDLVEGLMALM 337
CxcUXS1 EKRTAETLAMDYHRGADVEVRIARIFNTYGPRMCLDDGRVVSNFVAQAIRKQPLTVYGDGKQTRSFQFVSDLVAGLMALM 331
OsUXS1 GKRTAETLTMDYHRGGGVEVRIARIFNTYGPRMCLDDGRVVSNFVAQALRRQPMTVYGDGKQTRSFQYVSDLVAGLMALM 325
ZmUXS1 GKRTAETLTMDYHRGGGVEVRIARIFNTYGPRMCLDDGRVVSNFVAQALRRQPMTVYGDGKQTRSFQYVADLVAGLMALM 320
PvUXS1 GKRVAETLMFDYHRQHGIEIRIARIFNTYGPRMNIDDGRVVSNFIAQAIRGEPLTVQVPGTQTRSFCYVSDMVDGLIRLM 254
GhUXS3 GKRVAETLMFDYHRQHGIEIRIARIFNTYGPRMNIDDGRVVSNFIAQALRGEPLTVQKPGTQTRSFCYVSDMVDGLIRLM 258
EgUXS1 GKRVAETLMFDYHRQHGIEIRVARIFNTYGPRMNIDDGRVVSNFIAQAVRGEPLTVQAPGTQTRSFCYVSDMVDGLIRLM 258
PtUXS3 GKRVAETLMFDYHRQHGIEFRIARIFNTYGPRMNIDDGRVVSNFIAQALRGEPLTVQKPGTQTRSFCYVSDMVDGLIRLM 255
GhUXS6 GKRVAETLMFDYHRQHGIEIRIARIFNTYGPRMNIDDGRVVSNFIAQALRNEPLTVQLPGTQTRSFCYVSDMVDGLIRLM 255
OlUXS1 GKRVAETLMFDYHRQHGIEIRIARIFNTYGPRMNIDDGRVVSNFIAQAIRGEPLTVQAPGTQTRSFCYVSDMVDGLIRLM 258
CxcUXS5 GKRVAETLMFDYHRQHGIEIRIARIFNTYGPRMNIDDGRVVSNFIAQAIRGEPLTVQAPGTQTRSFCYVSDMVDGLIRLM 256
AtUXS1 GKRVAETLMFDYHRQHGIEIRIARIFNTYGPRMNIDDGRVVSNFIAQALRGEALTVQKPGTQTRSFCYVSDMVDGLIRLM 255
AtUXS3 GKRVAETLMFDYHRQHGIEIRIARIFNTYGPRMNIDDGRVVSNFIAQALRGEALTVQKPGTQTRSFCYVSDMVDGLIRLM 255
AtUXS5 GKRVAETLMFDYHRQHGIEIRIARIFNTYGPRMNIDDGRVVSNFIAQALRGEALTVQKPGTQTRSFCYVSDMVDGLMRLM 254
AtUXS6 GKRVAETLMFDYHRQHGIEIRIARIFNTYGPRMNIDDGRVVSNFIAQALRGEALTVQKPGTQTRSFCYVSDMVEGLMRLM 256
MtUXS1 GKRVAETLMFDYHRQHGIEIRVARIFNTYGPRMNIDDGRVVSNFIAQALRGE$LTVQAPGTQTRSFCYVSDLVDGLIRLM 254
CxcUXS4 GKRVAETLMFDYHRQHAIEIRIARIFNTYGPRMNIDDGRVVSNFIAQALRGEPLTVQLPGTQTRSFCYVSDMVDGLIRLM 253
CxcUXS3 GKRVAETLMFDYHRQHAIEIRIARIFNTYGPRMNIDDGRVVSNFIAQALRGEPLTVQLPGTQTRSFCYVSDMVDGLIRLM 256
NtUXS6 GKRVAETLMFDYHRQHGIEIRIARIFNTYGPRMNIDDGRVVSNFIAQAIRDEALTVQLPGTQTRSFCYVSDMVDGLIRLM 255
OsUXS6 GKRVAETLMFDYHRQHGIEIRIARIFNTYGPRMNIDDGRVVSNFIAQAVRGEPLTVQKPGTQTRSFCYVADMVNGLIKLM 260
GhUXS2 EGEHVGPFNLGNPGEFTMLELAEVVQETIDPNAKIEFRPNTEDDPHKRKPDI$RAKELLGWQPKV$LRKGLPLMV$DFRQ 422
GaUXS1 EGEHIGPFNLGNPGEFTMLELAEVVQETIDPNAKIEFRPNTEDDPHKRKPDIS RAKELIGWEPKVSLRKGLPLMVSDFRQ 418
AtUXS4 EGEHVGPFNLGNPGEFTMLELAKVVQETIDPNAKIEFRPNTEDDPHKRKPDITKAKELLGWEPKVALRQGLPLMVKDFRQ 425
AtUXS2 EGEHVGPFNLGNPGEFTMLELAKVVQETIDPNANIEFRPNTEDDPHKRKPDITKAKELLGWEPKVSLRQGLPLMVKDFRQ 423
PtUXS2 EGEHVGPFNLGNPGEFTMLELAQVVQETIDPNARIEFRPNTEDDPHKRKPDITKAKDLLGWEPKISLHQGLPMMVSDFRQ 425
MdUXS2 EGEHVGPFNLGNPGEFTMLELAKVVQETIDPAAKIEYRPNTEDDPHKRKPDITKAKDLLGWQPKVSLQKGLPLMV$DFRK 426
NtUXS3 EGEHVGPFNLGNPGEFTMLELAGVVQETIDPNAQIEFRPNTADDPHKRKPDI$KAKELLGWEPKVPLRKGLPLMVQDFRQ 430
OlUXS2 EGEHVGPFNLGNPGEFTMLELAKVVQDTIDPNARIEFRENTADDPHKRKPDI$KAKELLGWEPKIPLRQGLPLMV$DFRK 418
CxcUXS2 EGEHVGPFNLGNPGEFTMIELAKVVQDTID$NARIEFRQNTEDDPHKRKPDITRAKELLGWEPKISLRDGLPLMV$DFRK 419
OsUXS4 EGEHVGPFNLGNPGEFTMLELAKVVQDTIDPNARIEFRPNTADDPHKRKPDI$RAKELLGWEPKIPLHKGLPLMVQDFRD 430
OsUXS2 EGEHVGPFNLGNPGEFTMLELAKVVQDTIDPNAKIEFRPNTQDDPHKRKPDIGRAKELLGWEPKIPLHKGLPLMVTDFRK 429
CcUXS1 EGEHVGPFNLGNPGEFTMLELAQVVQETIDPNAKIEFRPNTEDDPHKRKPDI$KAKELLGWQP$VSLREGLPLMVTDFRQ 397
GhUXS1 EGEHIGPFNLGNPGEFTMLELAEVVKDTIDP$ATIEYKPNTADDPHMRKPDI$KAKELLNWEPKIPLREGLPLMVNDFRN 425
PtUXS1 EGEHVGPFNLGNPGEFTMLELAEVIKETID$$ATIEFKPNTADDPHKRKPDI$KAKELLNWEPRISLREGLPLMVNDFRN 424
AtUXS1 ENDHVGPFNLGNPGEFTMLELAEVVKEVIDP$ATIEFKPNTADDPHKRKPDI$KAKEQLNWEPKISLREGLPRMVSDFRN 422
MtUXS2 DGEHVGPFNLGNPGEFTMLELAQVVKETID$$ATIEYKQNTADDPHMRKPDI$KAKELLNWEPKVPLREGLPLMV$DFRN 448
OlUXS3 EGEHIGPFNLGNPGEFTMLELAEVVKETIDP$ATIEFKENTADDPHKRKPDI$KAKELLNWEPKVPLHEGLPLMV$DFQK 417
CxcUXS1 EGEHIGPFNLGNPGEFTMLELAEVVKETIDP$ATIEFKPNTEDDPHKRKPDITRAKELLNWEPKVSLREGLPQMVTDFQK 411
OsUXS1 EGDHIGPFNLGNPGEFTMLELAQVVKETIDPMATIEFKPNTADDPHMRKPDITKAKHLLRWEPKVSLREGLPLMVKDFRQ 405
ZmUXS1 E$DHIGPFNLGNPGEFTMLELAQVVKETIDPMATIEFKPNTADDPHMRKPDITKAKQLLHWEPKVSLKEGLPLMVQDFRQ 400
PvUXS1 EGENTGPINIGNPGEFTMTELAENVKELINPAVEIKMVENTPDDPRQRKPDITKAKELLGWEPKIKLRDGLPLMEEDFRL 334
GhUXS3 EGEHTGPINIGNPGEFTMVELAETVKELIKPDVEINMVENTPDDPRQRKPDITKAKELLGWEPKVKLRDGLPLMEEDFRL 338
EgUXS1 EGENTGPINIGNPGEFTMIELAETVKELINPDVEITMVENTPDDPRQRKPDITKAKELLGWEPKVKLRNGLPLMEDDFRL 338
PtUXS3 EGENTGPINIGNPGEFTMTELAETVKELINPGVEINMVENTPDDPRQRKPDITKAKALLGWEPKVKLRDGLPLMEEDLRL 335
GhUXS6 EGENTGPINIGNPGEFTMLELAE$VKELINPEVKI$MVENTPDDPRQRKPDITKAKEVLGWEPKVKLRDGLPLMEEDFRL 335
OlUXS1 EGENTGPINIGNPGEFTMMELAENVKELINPNVPVKIVENTPDDPRQRKPDITKAKELLGWEPTVTLREGLPLMEDDFRQ 338
CxcUXS5 GGENTGPINIGNPGEFTMMELAEQVKELINPNVPINIVENTPDDPRQRKPDITKAKELLGWEPTITLRQGLPLMEEDFRQ 336
AtUXS1 EGDDTGPINIGNPGEFTMVELAETVKELINP$IEIKMVENTPDDPRQRKPDI$KAKEVLGWEPKVKLREGLPLMEEDFRL 335
AtUXS3 EGNDTGPINIGNPGEFTMVELAETVKELINP$IEIKMVENTPDDPRQRKPDI$KAKEVLGWEPKVKLREGLPLMEEDFRL 335
AtUXS5 EGDDTGPINIGNPGEFTMVELAETVKELINP$IEIKMVENTPDDPRQRKPDITKAKEVLGWEPKVKLREGLPLMEEDFRL 334
AtUXS6 EGDQTGPINIGNPGEFTMVELAETVKELIKPDVEIKMVENTPDDPRQRKPDI$KAKEVLGWEPKVKLREGLPLMEEDFRL 336
MtUXS1 GG$DTGPINLGNPGEFTMTELAETVKELINPNVEIKIVENTPDDPRQRKPDITKAKELLGWEPKVKL$EGLPLMEGDFRL 334
CxcUXS4 EGENTGPINIGNPGEFTMMELAETVKELIDPNVTIKVVENTPDDPRQRKPNITKAKELLGWEPKVTLRQGLPLMEEDFRQ 333
CxcUXS3 EGENTGPINIGNPGEFTMMELAETVKELIDPKITIKVVENTPDDPRQRKPNITKAKDLLGWEPKVTLREGLPLMEADFRH 336
NtUXS6 EGDNTGPINIGNPGEFTMIELAENVKELINPEVKIITVENTPDDPRQRKPDITKAKELIGWEPKIKLRDGIPLMEEDFRG 335
OsUXS6 NGDNTGPINLGNPGEFTMLELAENVKELINPEVTVTMTENTPDDPRQRKPDITKAKEVLGWEPKIVLRDGLVLMEDDFRE 340

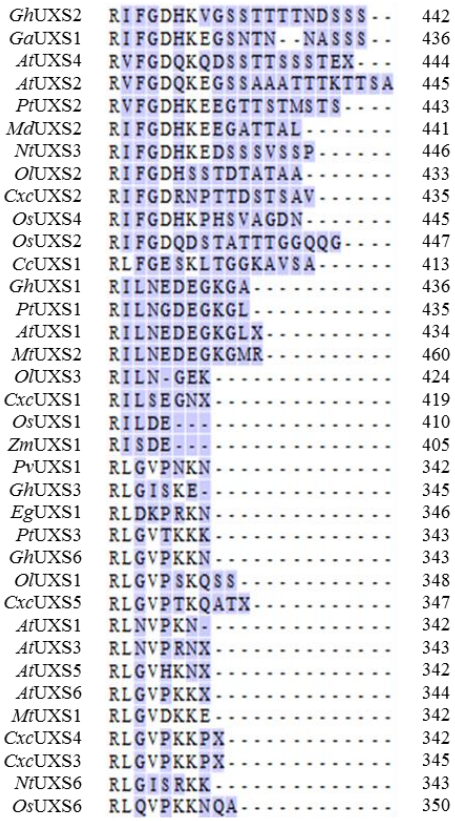


Figure S3.2: Amino acid sequence alignment of NCBI-nr identified putative and characterized plant UXS. The alignment includes protein sequences of putative and characterized plant UXSs found in the NCBI-nr database. Amino acids with highlighted with blue background colour are those different from the consensus.

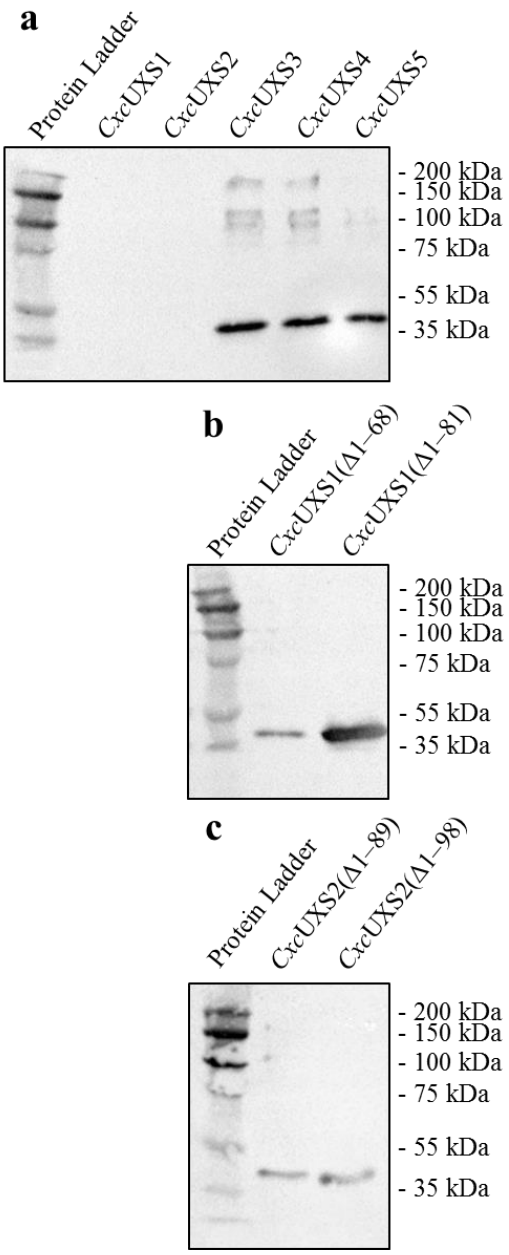


Figure S3.3: Western blot of purified *Cxc*UXS enzymes after purification. (a) Protein from *E. coli* expressing the full-length *Cxc*UXS1 – 5 was blotted using antibody specific for the his_6-tag. (b) and (c) show the same blotting for the truncated *Cxc*UXS1 and 2 enzymes, respectively.

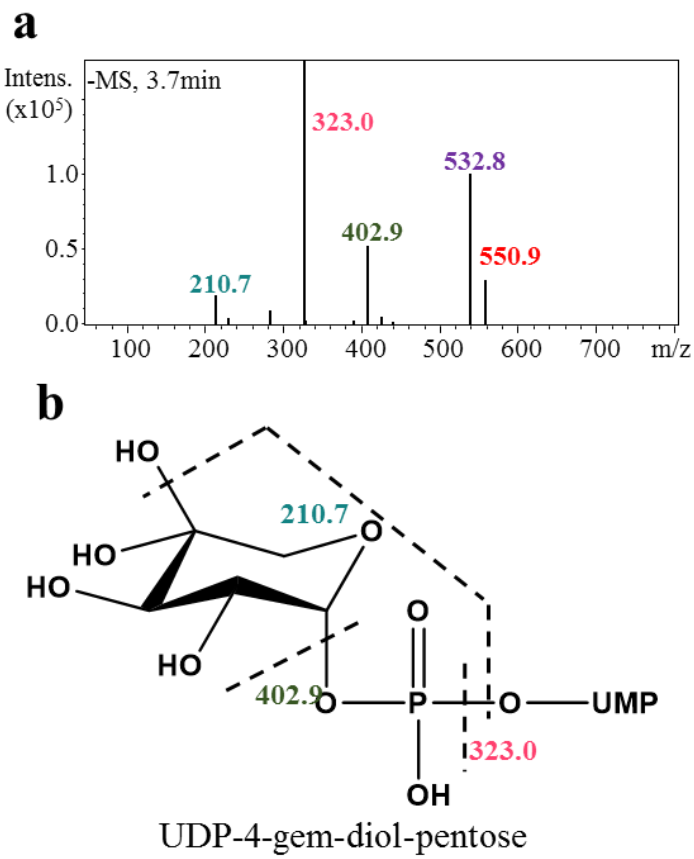


Figure S3.4. ESI LC-MS/MS analysis of the precursor ions at m/z 551.0 [M-H] generated by *Cxc*UXS1. (a) LC-MS/MS spectrum for the ion with m/z of 551.0 [M-H](superscript-) eluting at 3.7 minutes. (b) The proposed structure of the *Cxc*UXS1 enzymatic product. The ion with m/z of 533.0 [M-H](superscript-) is proposed to be UDP-4-keto pentose.

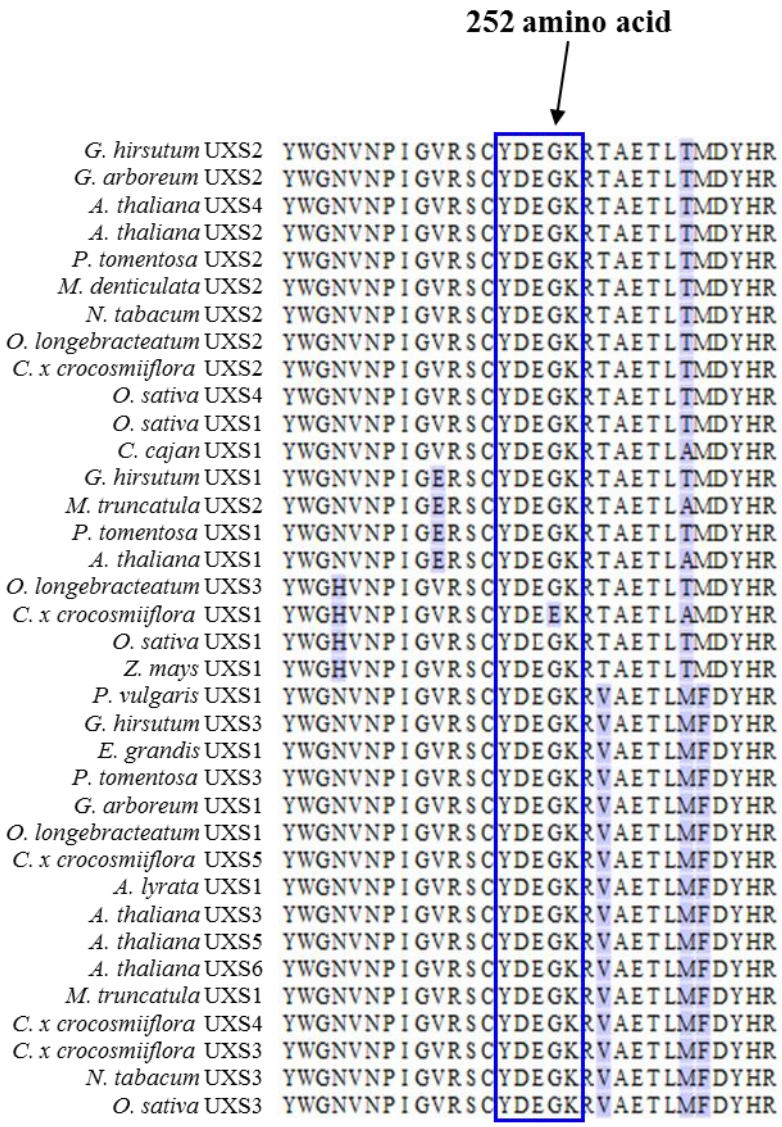
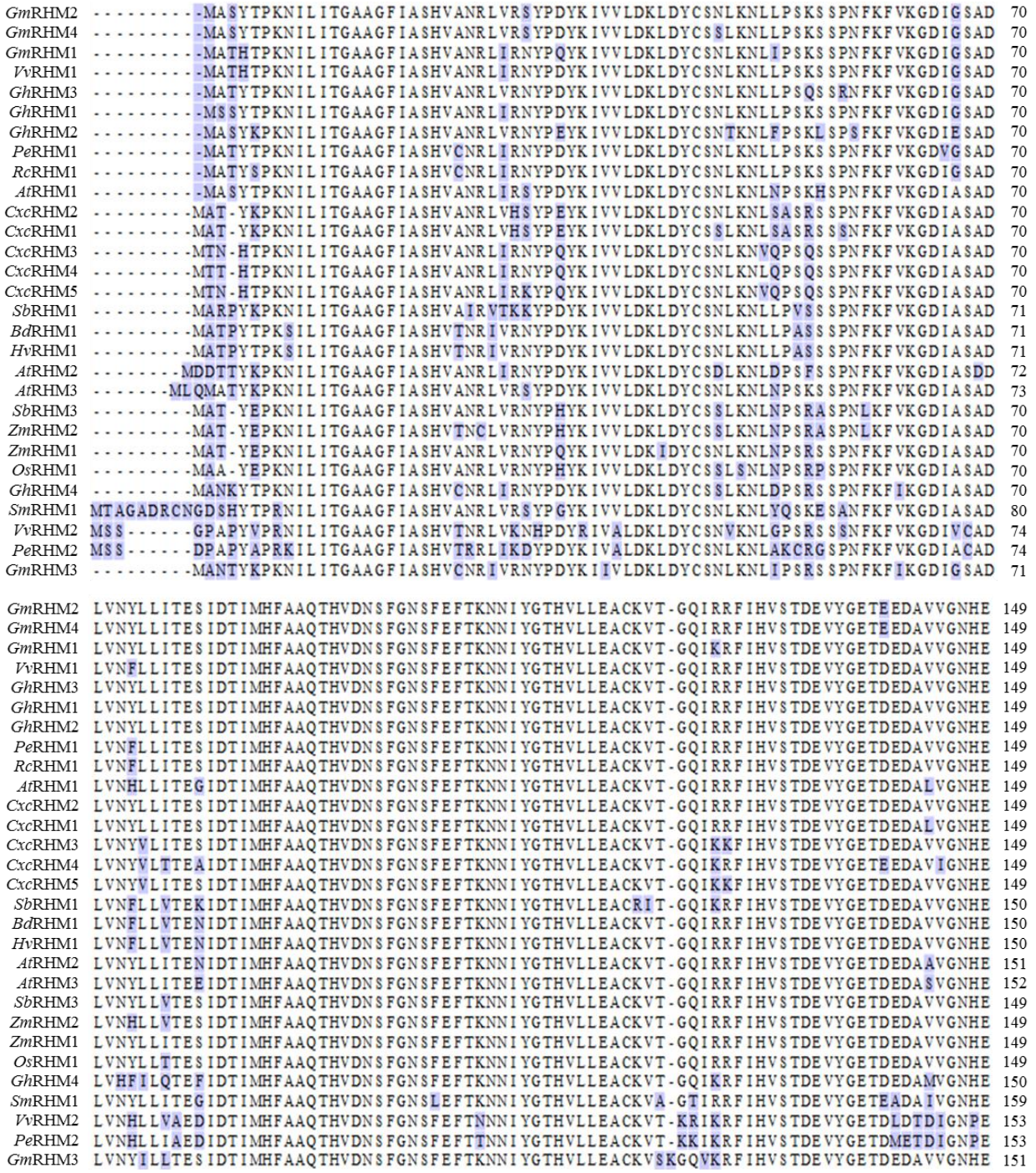
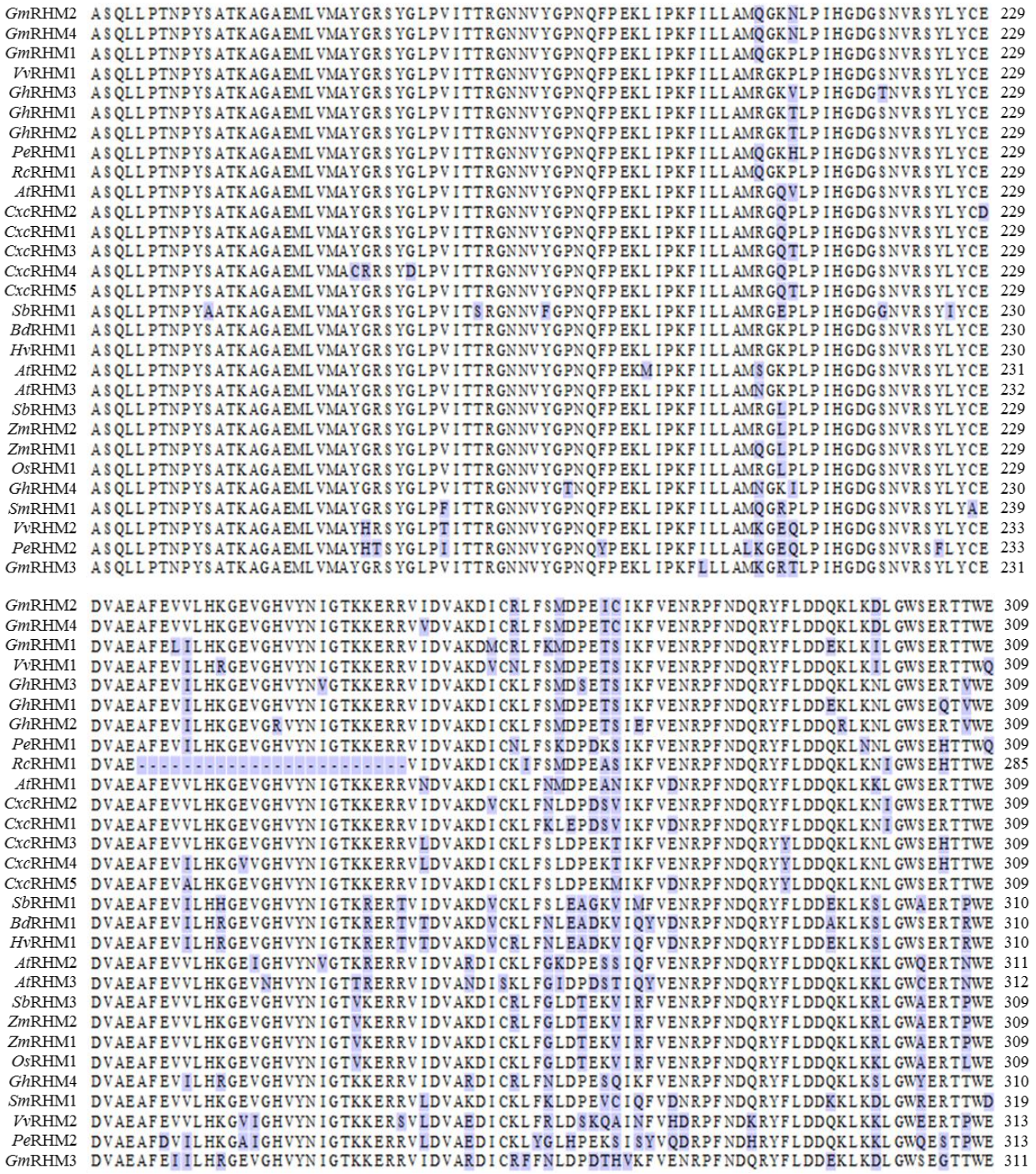
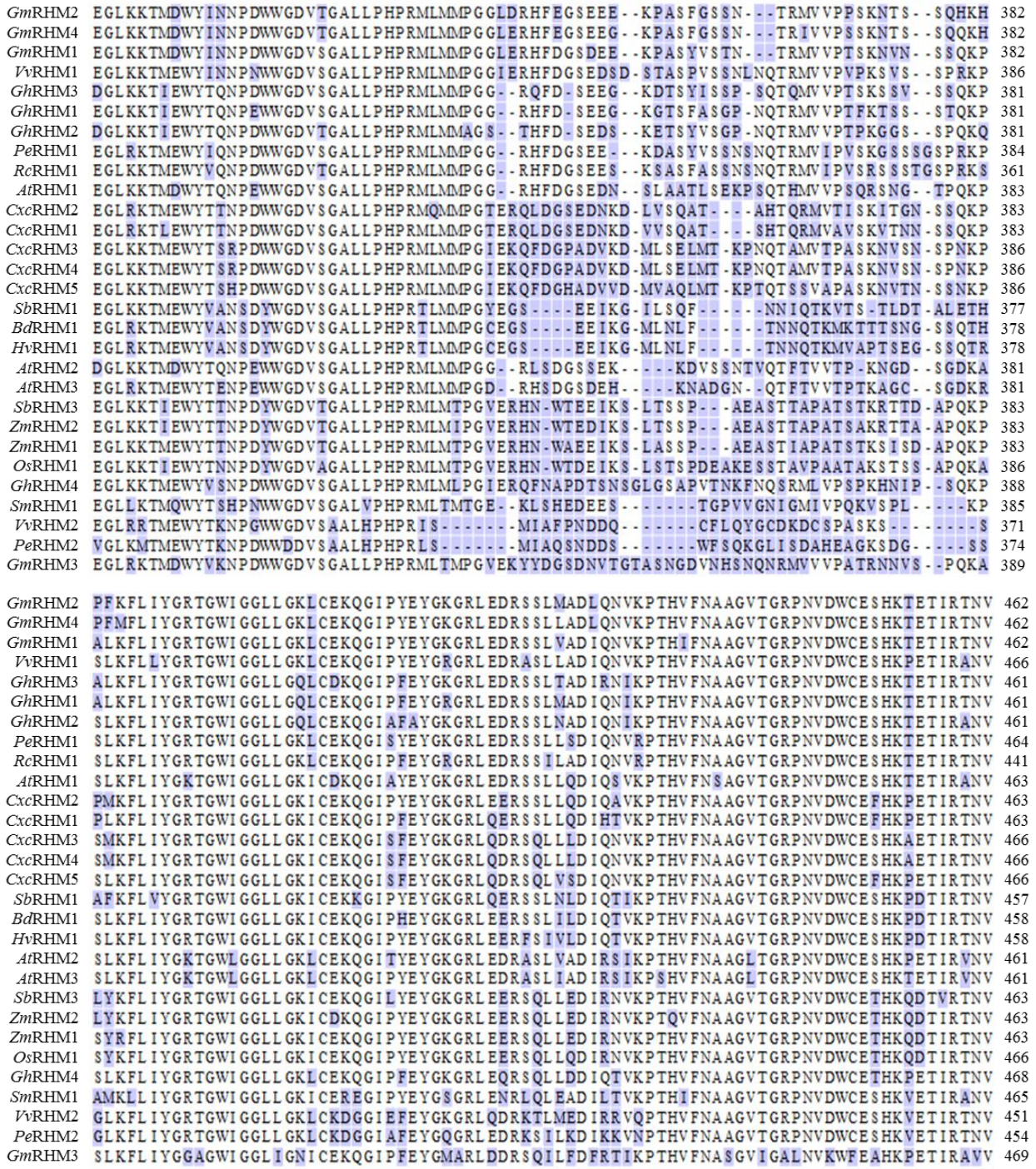
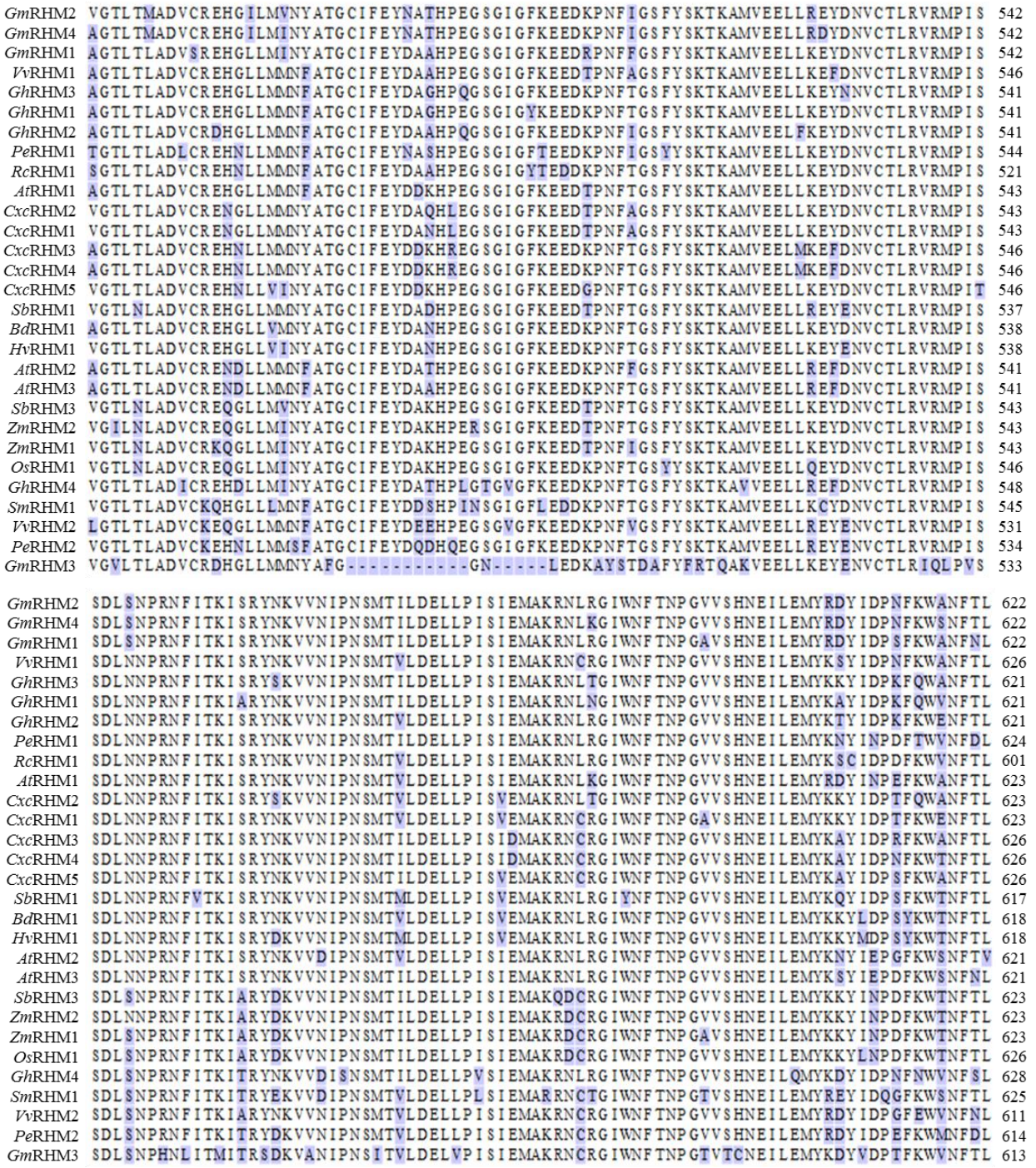


Figure S3.5: Amino acid sequence alignment of the YXXXK motifs of putative and characterized plant UXS. Amino acids with highlighted with blue background colour are those different from the consensus. The conserved YXXXK motif is identified by the blue box.


GmRHM2 ---------MASYTPKNILITGAAGFIASHVANRLVRSYPDYKIVVLDKLDYCSNLKNLLPSKSSPNFKFVKGDIGSAD 70
GmRHM4 ---------MASYTPKNILITGAAGFIASHVANRLVRSYPDYKIVVLDKLDYCSSLKNLLPSKSSPNFKFVKGDIGSAD 70
GmRHM1 ---------MATHTPKNILITGAAGFIASHVANRLIRNYPQYKIVVLDKLDYCSNLKNLIPSKSSPNFKFVKGDIGSAD 70
VvRHM1 ---------MATHTPKNILITGAAGFIASHVANRLIRNYPDYKIVVLDKLDYCSNLKNLLPSKSSPNFKFVKGDIGSAD 70
GhRHM3 ---------MATYTPKNILITGAAGFIASHVANRLVRNYPDYKIVVLDKLDYCSNLKNLLPSQSSRNFKFVKGDIGSAD 70
GhRHM1 ---------MSSYTPKNILITGAAGFIASHVANRLIRNYPDYKIVVLDKLDYCSNLKNLLPSKSSPNFKFVKGDIGSAD 70
GhRHM2 ---------MASYKPKNILITGAAGFIASHVANRLVRNYPEYKIVVLDKLDYCSNTKNLFPSKLSPSFKFVKGDIESAD 70
PeRHM1 ---------MATYTPKNILITGAAGFIASHVCNRLIRNYPDYKIVVLDKLDYCSNLKNLLPSKSSPNFKFVKGDVGSAD 70
RcRHM1 ---------MATYSPKNILITGAAGFIASHVCNRLIRNYPDYKIVVLDKLDYCSNLKNLLPSKSSPNFKFVKGDIGSAD 70
AtRHM1 ---------MASYTPKNILITGAAGFIASHVANRLIRSYPDYKIVVLDKLDYCSNLKNLNPSKHSPNFKFVKGDIASAD 70
CxcRHM2 ---------MAT-YKPKNILITGAAGFIASHVANRLVHSYPEYKIVVLDKLDYCSNLKNLSASRSSPNFKFVKGDIASAD 70
CxcRHM1 ---------MAT-YKPKNILITGAAGFIASHVANRLVHSYPEYKIVVLDKLDYCSSLKNLSASRSSSNFKFVKGDIASAD 70
CxcRHM3 ---------MTN-HTPKNILITGAAGFIASHVANRLIRNYPQYKIVVLDKLDYCSNLKNVQPSQSSPNFKFVKGDIASAD 70
CxcRHM4 ---------MTT-HTPKNILITGAAGFIASHVANRLIRNYPQYKIVVLDKLDYCSNLKNLQPSQSSPNFKFVKGDIASAD 70
CxcRHM5 ---------MTN-HTPKNILITGAAGFIASHVANRLIRKYPQYKIVVLDKLDYCSNLKNVQPSQSSPNFKFVKGDIASAD 70
SbRHM1 ---------MARPYKPKNILITGAAGFIASHVAIRVTKKYPDYKIVVLDKLDYCSNLKNLLPVSSSPNFKFVKGDIASAD 71
BdRHM1 ---------MATPYTPKSILITGAAGFIASHVTNRIVRNYPDYKIVVLDKLDYCSNLKNLLPASSSPNFKFVKGDIASAD 71
HvRHM1 ---------MATPYTPKSILITGAAGFIASHVTNRIVRNYPDYKIVVLDKLDYCSNLKNLLPASSSPNFKFVKGDIASAD 71
AtRHM2 --------MDDTTYKPKNILITGAAGFIASHVANRLIRNYPDYKIVVLDKLDYCSDLKNLDPSFSSPNFKFVKGDIASDD 72
AtRHM3 -------MLQMATYKPKNILITGAAGFIASHVANRLVRSYPDYKIVVLDKLDYCSNLKNLNPSKSSPNFKFVKGDIASAD 73
SbRHM3 ---------MAT-YEPKNILITGAAGFIASHVANRLVRNYPHYKIVVLDKLDYCSSLKNLNPSRASPNLKFVKGDIASAD 70
ZmRHM2 ---------MAT-YEPKNILITGAAGFIASHVTNCLVRNYPHYKIVVLDKLDYCSSLKNLNPSRASPNLKFVKGDIASAD 70
ZmRHM1 ---------MAT-YEPKNILITGAAGFIASHVANRLVRNYPQYKIVVLDKIDYCSNLKNLNPSRSSPNFKFVKGDIASAD 70
OsRHM1 ---------MAA-YEPKNILITGAAGFIASHVANRLVRNYPHYKIVVLDKLDYCSSLSNLNPSRPSPNFKFVKGDIASAD 70
GhRHM4 ---------MANKYTPKNILITGAAGFIASHVCNRLIRNYPDYKIVVLDKLDYCSSLKNLDPSRSSPNFKFIKGDIASAD 70
SmRHM1 MTAGADRCNGDSHYTPRNILITGAAGFIASHVANRLVRSYPGYKIVVLDKLDYCSNLKNLYQSKESANFKFVKGDIASAD 80
VvRHM2 MSS------GPAPYVPRNILITGAAGFIASHVTNRLVKNHPDYRIVALDKLDYCSNVKNLGPSRSSSNFKFVKGDIVCAD 74
PeRHM2 MSS------DPAPYAPRKILITGAAGFIASHVTRRLIKDYPDYKIVALDKLDYCSNLKNLAKCRGSPNFKFVKGDIACAD 74
GmRHM3 ---------MANTYKPKNILITGAAGFIASHVCNRIVRNYPDYKIIVLDKLDYCSNLKNLIPSRSSPNFKFIKGDIGSAD 71
GmRHM2 LVNYLLITESIDTIMHFAAQTHVDNSFGNSFEFTKNNIYGTHVLLEACKVT-GQIRRFIHVSTDEVYGETEEDAVVGNHE 149
GmRHM4 LVNYLLITESIDTIMHFAAQTHVDNSFGNSFEFTKNNIYGTHVLLEACKVT-GQIRRFIHVSTDEVYGETEEDAVVGNHE 149
GmRHM1 LVNYLLITESIDTIMHFAAQTHVDNSFGNSFEFTKNNIYGTHVLLEACKVT-GQIKRFIHVSTDEVYGETDEDAVVGNHE 149
VvRHM1 LVNFLLITESIDTIMHFAAQTHVDNSFGNSFEFTKNNIYGTHVLLEACKVT-GQIRRFIHVSTDEVYGETDEDAVVGNHE 149
GhRHM3 LVNYLLITESIDTIMHFAAQTHVDNSFGNSFEFTKNNIYGTHVLLEACKVT-GQIRRFIHVSTDEVYGETDEDAVVGNHE 149
GhRHM1 LVNYLLITESIDTIMHFAAQTHVDNSFGNSFEFTKNNIYGTHVLLEACKVT-GQIRRFIHVSTDEVYGETDEDAVVGNHE 149
GhRHM2 LVNYLLITESIDTIMHFAAQTHVDNSFGNSFEFTKNNIYGTHVLLEACKVT-GQIRRFIHVSTDEVYGETDEDAVVGNHE 149
PeRHM1 LVNFLLITESIDTIMHFAAQTHVDNSFGNSFEFTKNNIYGTHVLLEACKVT-GQIRRFIHVSTDEVYGETDEDAVVGNHE 149
RcRHM1 LVNFLLITESIDTIMHFAAQTHVDNSFGNSFEFTKNNIYGTHVLLEACKVT-GQIRRFIHVSTDEVYGETDEDAVVGNHE 149
AtRHM1 LVNHLLITEGIDTIMHFAAQTHVDNSFGNSFEFTKNNIYGTHVLLEACKVT-GQIRRFIHVSTDEVYGETDEDALVGNHE 149
CxcRHM2 LVNYLLITESIDTIMHFAAQTHVDNSFGNSFEFTKNNIYGTHVLLEACKVT-GQIRRFIHVSTDEVYGETDEDAVVGNHE 149
CxcRHM1 LVNYLLITESIDTIMHFAAQTHVDNSFGNSFEFTKNNIYGTHVLLEACKVT-GQIRRFIHVSTDEVYGETDEDALVGNHE 149
CxcRHM3 LVNYVLITESIDTIMHFAAQTHVDNSFGNSFEFTKNNIYGTHVLLEACKVT-GQIKKFIHVSTDEVYGETDEDAVVGNHE 149
CxcRHM4 LVNYVLTTEAIDTIMHFAAQTHVDNSFGNSFEFTKNNIYGTHVLLEACKVT-GQIKRFIHVSTDEVYGETEEDAVIGNHE 149
CxcRHM5 LVNYVLITESIDTIMHFAAQTHVDNSFGNSFEFTKNNIYGTHVLLEACKVT-GQIKKFIHVSTDEVYGETDEDAVVGNHE 149
SbRHM1 LVNFLLVTEKIDTIMHFAAQTHVDNSFGNSFEFTKNNIYGTHVLLEACRIT-GQIKRFIHVSTDEVYGETDEDAVVGNHE 150
BdRHM1 LVNFLLVTENIDTIMHFAAQTHVDNSFGNSFEFTKNNIYGTHVLLEACKVT-GQIRRFIHVSTDEVYGETDEDAVVGNHE 150
HvRHM1 LVNFLLVTENIDTIMHFAAQTHVDNSFGNSFEFTKNNIYGTHVLLEACKVT-GQIRRFIHVSTDEVYGETDEDAVVGNHE 150
AtRHM2 LVNYLLITENIDTIMHFAAQTHVDNSFGNSFEFTKNNIYGTHVLLEACKVT-GQIRRFIHVSTDEVYGETDEDAAVGNHE 151
AtRHM3 LVNYLLITEEIDTIMHFAAQTHVDNSFGNSFEFTKNNIYGTHVLLEACKVT-GQIRRFIHVSTDEVYGETDEDASVGNHE 152
SbRHM3 LVNYLLVTESIDTIMHFAAQTHVDNSFGNSFEFTKNNIYGTHVLLEACKVT-GQIRRFIHVSTDEVYGETDEDAVVGNHE 149
ZmRHM2 LVNHLLVTESIDTIMHFAAQTHVDNSFGNSFEFTKNNIYGTHVLLEACKVT-GQIRRFIHVSTDEVYGETDEDAVVGNHE 149
ZmRHM1 LVNYLLITESIDTIMHFAAQTHVDNSFGNSFEFTKNNIYGTHVLLEACKVT-GQIRRFIHVSTDEVYGETDEDAVVGNHE 149
OsRHM1 LVNYLLTTESIDTIMHFAAQTHVDNSFGNSFEFTKNNIYGTHVLLEACKVT-GQIRRFIHVSTDEVYGETDEDAVVGNHE 149
GhRHM4 LVHFILQTEFIDTIMHFAAQTHVDNSFGNSFEFTKNNIYGTHVLLEACKVT-GQIKRFIHVSTDEVYGETDEDAMVGNHE 150
SmRHM1 LVNYLLITEGIDTIMHFAAQTHVDNSFGNSLEFTKNNIYGTHVLLEACKVA-GTIRRFIHVSTDEVYGETEADAIVGNHE 159
VvRHM2 LVNHLLVAEDIDTIMHFAAQTHVDNSFGNSFEFTNNNIYGTHVLLEACKVT-KRIKRFIHVSTDEVYGETDLDTDIGNPE 153
PeRHM2 LVNHLLIAEDIDTIMHFAAQTHVDNSFGNSFEFTTNNIYGTHVLLEACKVT-KKIKRFIHVSTDEVYGETDMETDIGNPE 153
GmRHM3 LVNYILLTESIDTIMHFAAQTHVDNSFGNSFEFTKNNIYGTHVLLEACKVSKGQVKRFIHVSTDEVYGETDEDAVVGNHE 151


GmRHM2 ASQLLPTNPYSATKAGAEMLVMAYGRSYGLPVITTRGNNVYGPNQFPEKLIPKFILLAMQGKNLPIHGDGSNVRSYLYCE 229
GmRHM4 ASQLLPTNPYSATKAGAEMLVMAYGRSYGLPVITTRGNNVYGPNQFPEKLIPKFILLAMQGKNLPIHGDGSNVRSYLYCE 229
GmRHM1 ASQLLPTNPYSATKAGAEMLVMAYGRSYGLPVITTRGNNVYGPNQFPEKLIPKFILLAMQGKPLPIHGDGSNVRSYLYCE 229
VvRHM1 ASQLLPTNPYSATKAGAEMLVMAYGRSYGLPVITTRGNNVYGPNQFPEKLIPKFILLAMRGKPLPIHGDGSNVRSYLYCE 229
GhRHM3 ASQLLPTNPYSATKAGAEMLVMAYGRSYGLPVITTRGNNVYGPNQFPEKLIPKFILLAMRGKVLPIHGDGTNVRSYLYCE 229
GhRHM1 ASQLLPTNPYSATKAGAEMLVMAYGRSYGLPVITTRGNNVYGPNQFPEKLIPKFILLAMRGKTLPIHGDGSNVRSYLYCE 229
GhRHM2 ASQLLPTNPYSATKAGAEMLVMAYGRSYGLPVITTRGNNVYGPNQFPEKLIPKFILLAMRGKTLPIHGDGSNVRSYLYCE 229
PeRHM1 ASQLLPTNPYSATKAGAEMLVMAYGRSYGLPVITTRGNNVYGPNQFPEKLIPKFILLAMQGKHLPIHGDGSNVRSYLYCE 229
RcRHM1 ASQLLPTNPYSATKAGAEMLVMAYGRSYGLPVITTRGNNVYGPNQFPEKLIPKFILLAMQGKPLPIHGDGSNVRSYLYCE 229
AtRHM1 ASQLLPTNPYSATKAGAEMLVMAYGRSYGLPVITTRGNNVYGPNQFPEKLIPKFILLAMRGQVLPIHGDGSNVRSYLYCE 229
CxcRHM2 ASQLLPTNPYSATKAGAEMLVMAYGRSYGLPVITTRGNNVYGPNQFPEKLIPKFILLAMRGQPLPIHGDGSNVRSYLYCD 229
CxcRHM1 ASQLLPTNPYSATKAGAEMLVMAYGRSYGLPVITTRGNNVYGPNQFPEKLIPKFILLAMRGQPLPIHGDGSNVRSYLYCE 229
CxcRHM3 ASQLLPTNPYSATKAGAEMLVMAYGRSYGLPVITTRGNNVYGPNQFPEKLIPKFILLAMRGQTLPIHGDGSNVRSYLYCE 229
CxcRHM4 ASQLLPTNPYSATKAGAEMLVMACRRSYDLPVITTRGNNVYGPNQFPEKLIPKFILLAMRGQPLPIHGDGSNVRSYLYCE 229
CxcRHM5 ASQLLPTNPYSATKAGAEMLVMAYGRSYGLPVITTRGNNVYGPNQFPEKLIPKFILLAMRGQTLPIHGDGSNVRSYLYCE 229
SbRHM1 ASQLLPTNPYAATKAGAEMLVMAYGRSYGLPVITSRGNNVFGPNQFPEKLIPKFILLAMRGEPLPIHGDGGNVRSYIYCE 230
BdRHM1 ASQLLPTNPYSATKAGAEMLVMAYGRSYGLPVITTRGNNVYGPNQFPEKLIPKFILLAMRGKPLPIHGDGSNVRSYLYCE 230
HvRHM1 ASQLLPTNPYSATKAGAEMLVMAYGRSYGLPVITTRGNNVYGPNQFPEKLIPKFILLAMRGKPLPIHGDGSNVRSYLYCE 230
AtRHM2 ASQLLPTNPYSATKAGAEMLVMAYGRSYGLPVITTRGNNVYGPNQFPEKMIPKFILLAMSGKPLPIHGDGSNVRSYLYCE 231
AtRHM3 ASQLLPTNPYSATKAGAEMLVMAYGRSYGLPVITTRGNNVYGPNQFPEKLIPKFILLAMNGKPLPIHGDGSNVRSYLYCE 232
SbRHM3 ASQLLPTNPYSATKAGAEMLVMAYGRSYGLPVITTRGNNVYGPNQFPEKLIPKFILLAMRGLPLPIHGDGSNVRSYLYCE 229
ZmRHM2 ASQLLPTNPYSATKAGAEMLVMAYGRSYGLPVITTRGNNVYGPNQFPEKLIPKFILLAMRGLPLPIHGDGSNVRSYLYCE 229
ZmRHM1 ASQLLPTNPYSATKAGAEMLVMAYGRSYGLPVITTRGNNVYGPNQFPEKLIPKFILLAMQGLPLPIHGDGSNVRSYLYCE 229
OsRHM1 ASQLLPTNPYSATKAGAEMLVMAYGRSYGLPVITTRGNNVYGPNQFPEKLIPKFILLAMRGLPLPIHGDGSNVRSYLYCE 229
GhRHM4 ASQLLPTNPYSATKAGAEMLVMAYGRSYGLPVITTRGNNVYGTNQFPEKLIPKFILLAMNGKILPIHGDGSNVRSYLYCE 230
SmRHM1 ASQLLPTNPYSATKAGAEMLVMAYGRSYGLPFITTRGNNVYGPNQFPEKLIPKFILLAMQGRPLPIHGDGSNVRSYLYAE 239
VvRHM2 ASQLLPTNPYSATKAGAEMLVMAYHRSYGLPTITTRGNNVYGPNQFPEKLIPKFILLAMKGEQLPIHGDGSNVRSYLYCE 233
PeRHM2 ASQLLPTNPYSATKAGAEMLVMAYHTSYGLPIITTRGNNVYGPNQYPEKLIPKFILLALKGEQLPIHGDGSNVRSFLYCE 233
GmRHM3 ASQLLPTNPYSATKAGAEMLVMAYGRSYGLPVITTRGNNVYGPNQFPEKLIPKFLLLAMKGRTLPIHGDGSNVRSYLYCE 231
GmRHM2 DVAEAFEVVLHKGEVGHVYNIGTKKERRVIDVAKDICRLFSMDPEICIKFVENRPFNDQRYFLDDQKLKDLGWSERTTWE 309
GmRHM4 DVAEAFEVVLHKGEVGHVYNIGTKKERRVVDVAKDICRLFSMDPETCIKFVENRPFNDQRYFLDDQKLKDLGWSERTTWE 309
GmRHM1 DVAEAFELILHKGEVGHVYNIGTKKERRVIDVAKDMCRLFKMDPETSIKFVENRPFNDQRYFLDDEKLKILGWSERTTWE 309
VvRHM1 DVAEAFEVILHRGEVGHVYNIGTKKERRVIDVAKDVCNLFSMDPETSIKFVENRPFNDQRYFLDDQKLKILGWSERTTWQ 309
GhRHM3 DVAEAFEVILHKGEVGHVYNVGTKKERRVIDVAKDICKLFSMDSETSIKFVENRPFNDQRYFLDDQKLKNLGWSERTVWE 309
GhRHM1 DVAEAFEVILHKGEVGHVYNIGTKKERRVIDVAKDICKLFSMDPETSIKFVENRPFNDQRYFLDDEKLKNLGWSEQTVWE 309
GhRHM2 DVAEAFEVILHKGEVGRVYNIGTKKERRVIDVAKDICKLFSMDPETSIEFVENRPFNDQRYFLDDQRLKNLGWSERTVWE 309
PeRHM1 DVAEAFEVILHKGEVGHVYNIGTKKERRVIDVAKDICNLFSKDPDKSIKFVENRPFNDQRYFLDDQKLNNLGWSEHTTWQ 309
RcRHM1 DVAE----------VIDVAKDICKIFSMDPEASIKFVENRPFNDQRYFLDDQKLKNIGWSEHTTWE 285
AtRHM1 DVAEAFEVVLHKGEVGHVYNIGTKKERRVNDVAKDICKLFNMDPEANIKFVDNRPFNDQRYFLDDQKLKKLGWSERTTWE 309
CxcRHM2 DVAEAFEVVLHKGEVGHVYNIGTKKERRVIDVAKDVCKLFNLDPDSVIKFVENRPFNDQRYFLDDQKLKNIGWSERTTWE 309
CxcRHM1 DVAEAFEVVLHKGEVGHVYNIGTKKERRVIDVAKDICKLFKLEPDSVIKFVDNRPFNDQRYFLDDQKLKNIGWSERTTWE 309
CxcRHM3 DVAEAFEVVLHKGEVGHVYNIGTKKERRVLDVAKDICKLFSLDPEKTIKFVENRPFNDQRYYLDDQKLKNLGWSEHTTWE 309
CxcRHM4 DVAEAFEVILHKGVVGHVYNIGTKKERRVLDVAKDICKLFSLDPEKTIKFVENRPFNDQRYYLDDQKLKNLGWSEHTTWE 309
CxcRHM5 DVAEAFEVALHKGEVGHVYNIGTKKERRVIDVAKDICKLFSLDPEKMIKFVDNRPFNDQRYYLDDQKLKNLGWSERTTWE 309
SbRHM1 DVAEAFEVILHHGEVGHVYNIGTKRERTVIDVAKDVCKLFSLEAGKVIMFVENRPFNDQRYFLDDEKLKSLGWAERTPWE 310
BdRHM1 DVAEAFEVILHRGEVGHVYNIGTKRERTVTDVAKDVCKLFNLEADKVIQYVDNRPFNDQRYFLDDAKLKSLGWSERTRWE 310
HvRHM1 DVAEAFEVILHRGEVGHVYNIGTKRERTVTDVAKDVCRLFNLEADKVIQFVDNRPFNDQRYFLDDEKLKSLGWSERTRWE 310
AtRHM2 DVAEAFEVVLHKGEIGHVYNVGTKRERRVIDVARDICKLFGKDPESSIQFVENRPFNDQRYFLDDQKLKKLGWQERTNWE 311
AtRHM3 DVAEAFEVVLHKGEVNHVYNIGTTRERRVIDVANDISKLFGIDPDSTIQYVENRPFNDQRYFLDDQKLKKLGWCERTNWE 312
SbRHM3 DVAEAFEVVLHKGEVGHVYNIGTVKERRVIDVAKDICRLFGLDTEKVIRFVENRPFNDQRYFLDDQKLKRLGWAERTPWE 309
ZmRHM2 DVAEAFEVVLHKGEVGHVYNIGTVKERRVIDVAKDICRLFGLDTEKVIRFVENRPFNDQRYFLDDQKLKRLGWAERTPWE 309
ZmRHM1 DVAEAFEVVLHKGEVGHVYNIGTVKERRVIDVAKDICKLFGLDTEKVIRFVENRPFNDQRYFLDDQKLKRLGWAERTPWE 309
OsRHM1 DVAEAFEVVLHKGEVGHVYNIGTVKERRVIDVAKDICKLFGLDTEKVIRFVENRPFNDQRYFLDDQKLKKLGWAERTLWE 309
GhRHM4 DVAEAFEVILHRGEVGHVYNIGTKKERRVIDVARDICRLFNLDPESQIKFVENRPFNDQRYFLDDQKLKSLGWYERTTWE 310
SmRHM1 DVAEAFEVVLHKGEVGHVYNIGTKKERRVLDVAKDICKLFKLDPEVCIQFVDNRPFNDQRYFLDDKKLKDLGWRERTTWD 319
VvRHM2 DVAEAFEVVLHKGVIGHVYNIGTKKERSVLDVAEDICKLFRLDSKQAINFVHDRPFNDKRYFLDDQKLKKLGWEERTPWE 313
PeRHM2 DVAEAFDVILHKGAIGHVYNIGTKKERRVLDVAEDICKLYGLHPEKSISYVQDRPFNDHRYFLDDQKLKKLGWQESTPWE 313
GmRHM3 DVAEAFEIILHRGEVGHVYNIGTKKERRVIDVARDICRFFNLDPDTHVKFVENRPFNDQRYFLDDEKLKDLGWSEGTTWE 311


GmRHM2 EGLKKTMDWYINNPDWWGDVTGALLPHPRMLMMPGGLDRHFEGSEEE--KPASFGSSN---TRMVVPPSKNTS--SQHKH 382
GmRHM4 EGLKKTMDWYINNPDWWGDVTGALLPHPRMLMMPGGLERHFEGSEEG--KPASFGSSN---TRIVVPSSKNTS--SQQKH 382
GmRHM1 EGLKKTMDWYINNPDWWGDVSGALLPHPRMLMMPGGLERHFDGSDEE--KPASYVSTN---TRMVVPTSKNVN--SSQKP 382
VvRHM1 EGLKKTMEWYINNPNWWGDVSGALLPHPRMLMMPGGIERHFDGSEDSD-STASPVSSNLNQTRMVVPVPKSVS--SPRKP 386
GhRHM3 DGLKKTIEWYTQNPDWWGDVSGALLPHPRMLMMPGG--RQFD-SEEG--KDTSYISSP-SQTQMVVPTSKSSV--SSQKP 381
GhRHM1 EGLKKTIEWYTQNPEWWGDVSGALLPHPRMLMMPGG--RHFD-SEEG--KGTSFASGP-NQTRMVVPTFKTSS--STQKP 381
GhRHM2 DGLKKTIEWYTQNPDWWGDVTGALLPHPRMLMMAGS--THFD-SEDS--KETSYVSGP-NQTRMVVPTPKGGS--SPQKQ 381
PeRHM1 EGLRKTMEWYIQNPDWWGDVSGALLPHPRMLMMPGG--RHFDGSEE---KDASYVSSNSNQTRMVIPVSKGSSSGSPRKP 384
RcRHM1 EGLKKTMEWYVQNPDWWGDVTGALLPHPRMLMMPGG--RHFDGSEES--KSASFASSNSNQTRMVIPVSRSSSTGSPRKS 361
AtRHM1 EGLKKTMDWYTQNPEWWGDVSGALLPHPRMLMMPGG--RHFDGSEDN--SLAATLSEKPSQTHMVVPSQRSNG--TPQKP 383
CxcRHM2 EGLRKTMEWYTTNPDWWGDVSGALLPHPRMQMMPGTERQLDGSEDNKD-LVSQAT----AHTQRMVTISKITGN-SSQKP 383
CxcRHM1 EGLRKTLEWYTTNPDWWGDVSGALLPHPRMLMMPGTERQLDGSEDNKD-VVSQAT----SHTQRMVAVSKVTNN-SSQKP 383
CxcRHM3 EGLKKTMEWYTSRPDWWGDVSGALLPHPRMLMMPGIEKQFDGPADVKD-MLSELMT-KPNQTAMVTPASKNVSN-SPNKP 386
CxcRHM4 EGLKKTMEWYTSRPDWWGDVSGALLPHPRMLMMPGIEKQFDGPADVKD-MLSELMT-KPNQTAMVTPASKNVSN-SPNKP 386
CxcRHM5 EGLKKTMEWYTSHPDWWGDVSGALLPHPRMLMMPGIEKQFDGHADVVD-MVAQLMT-KPTQTSSVAPASKNVTN-SSNKP 386
SbRHM1 EGLKKTMEWYVANSDYWGDVSGALLPHPRTLMMPGYEGS----EEIKG-ILSQF------NNIQTKVTS-TLDT-ALETH 377
BdRHM1 EGLRKTMEWYVANSDYWGDVSGALLPHPRTLMMPGCEGS----EEIKG-MLNLF------TNNQTKMKTTTSNG-SSQTH 378
HvRHM1 EGLRKTMEWYVANSDYWGDVSGALLPHPRTLMMPGCEGS----EEIKG-MLNLF------TNNQTKMVAPTSEG-SSQTR 378
AtRHM2 DGLKKTMDWYTQNPEWWGDVSGALLPHPRMLMMPGG--RLSDGSSEK-----KDVSSNTVQTFTVVTP-KNGD--SGDKA 381
AtRHM3 EGLRKTMEWYTENPEWWGDVSGALLPHPRMLMMPGD--RHSDGSDEH-----KNADGN--QTFTVVTPTKAGC--SGDKR 381
SbRHM3 EGLKKTIEWYTTNPDYWGDVTGALLPHPRMLMTPGVERHN-WTEEIKS-LTSSP---AEASTTAPATSTKRTTD-APQKP 383
ZmRHM2 EGLKKTIEWYTTNPDYWGDVTGALLPHPRMLMIPGVERHN-WTEDIKS-LTSSP---AEASTTAPATSAKRTTA-APQKP 383
ZmRHM1 EGLKKTMEWYTTNPDYWGDVTGALLPHPRMLMTPGVERHN-WAEEIKS-LASSP---AEASTIAPATSTKSISD-APQKP 383
OsRHM1 EGLKKTIEWYTNNPDYWGDVAGALLPHPRMLMTPGVERHN-WTDEIKS-LSTSPDEAKESSTAVPAATAKSTSS-APQKA 386
GhRHM4 EGLKKTMEWYVSNPDWWGDVSGALLPHPRMLMLPGIERQFNAPDTSNSGLGSAPVTNKFNQSRMLVPSPKHNIP--SQKP 388
SmRHM1 EGLLKTMQWYTSHPNWWGDVSGALVPHPRMLTMTGE--KLSHEDEES-------TGPVVGNIGMIVPQKVSPL-----KP 385
VvRHM2 EGLRRTMEWYTKNPGWWGDVSAALHPHPRIS-------MIAFPNDDQ-------CFLQYGCDKDCSPASKS---------S 371
PeRHM2 VGLKMTMEWYTKNPDWWDDVSAALHPHPRLS-------MIAQSNDDS-------WFSQKGLISDAHEAGKSDG-----SS 374
GmRHM3 EGLRKTMDWYVKNPDWWGDVSGALLPHPRMLTMPGVEKYYDGSDNVTGTASNGDVNHSNQNRMVVVPATRNNVS--PQKA 389
GmRHM2 PFKFLIYGRTGWIGGLLGKLCEKQGIPYEYGKGRLEDRSSLMADLQNVKPTHVFNAAGVTGRPNVDWCESHKTETIRTNV 462
GmRHM4 PFMFLIYGRTGWIGGLLGKLCEKQGIPYEYGKGRLEDRSSLLADLQNVKPTHVFNAAGVTGRPNVDWCESHKTETIRTNV 462
GmRHM1 ALKFLIYGRTGWIGGLLGKLCEKQGIPYEYGKGRLEDRSSLVADIQNVKPTHIFNAAGVTGRPNVDWCESHKTETIRTNV 462
VvRHM1 SLKFLLYGRTGWIGGLLGKLCEKQGIPYEYGRGRLEDRASLLADIQNVKPTHVFNAAGVTGRPNVDWCESHKPETIRANV 466
GhRHM3 ALKFLIYGRTGWIGGLLGQLCDKQGIPFEYGKGRLEDRSSLTADIRNIKPTHVFNAAGVTGRPNVDWCESHKTETIRTNV 461
GhRHM1 ALKFLIYGRTGWIGGLLGQLCEKQGIPFEYGRGRLEDRSSLMADIQNIKPTHVFNAAGVTGRPNVDWCESHKTETIRTNV 461
GhRHM2 SLKFLIYGRTGWIGGLLGQLCEKQGIAFAYGKGRLEDRSSLNADIQNIKPTHVFNAAGVTGRPNVDWCESHKTETIRANV 461
PeRHM1 SLKFLIYGRTGWIGGLLGKLCEKQGISYEYGKGRLEDRSSLLSDIQNVRPTHVFNAAGVTGRPNVDWCESHKTETIRTNV 464
RcRHM1 SLKFLIYGRTGWIGGLLGKLCEKQGIPFEYGRGRLEDRSSILADIQNVRPTHVFNAAGVTGRPNVDWCESHKTETIRTNV 441
AtRHM1 SLKFLIYGKTGWIGGLLGKICDKQGIAYEYGKGRLEDRSSLLQDIQSVKPTHVFNSAGVTGRPNVDWCESHKTETIRANV 463
CxcRHM2 PMKFLIYGRTGWIGGLLGKICEKQGIPYEYGKGRLEERSSLLQDIQAVKPTHVFNAAGVTGRPNVDWCEFHKPETIRTNV 463
CxcRHM1 PLKFLIYGRTGWIGGLLGKICEKQGIPFEYGKGRLQERSSLLQDIHTVKPTHVFNAAGVTGRPNVDWCEFHKPETIRTNV 463
CxcRHM3 SMKFLIYGRTGWIGGLLGKICEKQGISFEYGKGRLQDRSQLLLDIQNVKPTHVFNAAGVTGRPNVDWCESHKAETIRTNV 466
CxcRHM4 SMKFLIYGRTGWIGGLLGKICEKQGISFEYGKGRLQDRSQLLLDIQNVKPTHVFNAAGVTGRPNVDWCESHKAETIRTNV 466
CxcRHM5 SLKFLIYGRTGWIGGLLGKICEKQGISFEYGKGRLQDRSQLVSDIQNVKPTHVFNAAGVTGRPNVDWCEFHKPETIRTNV 466
SbRHM1 AFKFLVYGRTGWIGGLLGKICEKKGIPYEYGKGRLQERSSLNLDIQTIKPTHVFNAAGVTGRPNVDWCESHKPDTIRTNV 457
BdRHM1 SLKFLIYGRTGWIGGLLGKICEKQGIPHEYGKGRLEERSSLILDIQTVKPTHVFNAAGVTGRPNVDWCESHKPDTIRTNV 458
HvRHM1 SLKFLIYGRTGWIGGLLGKICEKQGIPYEYGKGRLEERFSIVLDIQTVKPTHVFNAAGVTGRPNVDWCESHKPDTIRTNV 458
AtRHM2 SLKFLIYGKTGWLGGLLGKLCEKQGITYEYGKGRLEDRASLVADIRSIKPTHVFNAAGLTGRPNVDWCESHKPETIRVNV 461
AtRHM3 SLKFLIYGKTGWLGGLLGKLCEKQGIPYEYGKGRLEDRASLIADIRSIKPSHVFNAAGLTGRPNVDWCESHKTETIRVNV 461
SbRHM3 LYKFLIYGRTGWIGGLLGKICEKQGILYEYGKGRLEERSQLLEDIRNVKPTHVFNAAGVTGRPNVDWCETHKQDTVRTNV 463
ZmRHM2 LYKFLIYGRTGWIGGLLGKICDKQGIPYEYGKGRLEERSQLLEDIRNVKPTQVFNAAGVTGRPNVDWCETHKQDTIRTNV 463
ZmRHM1 SYRFLIYGRTGWIGGLLGKICEKQGIPYEYGKGRLEERSQLLEDIRNVKPTHVFNAAGVTGRPNVDWCETHKQDTIRTNV 463
OsRHM1 SYKFLIYGRTGWIGGLLGKICEKQGIPYEYGKGRLEERSQLLQDIRNVKPTHVFNAAGVTGRPNVDWCETHKQDTIRTNV 466
GhRHM4 SLKFLIYGRTGWIGGLLGKLCEKQGIPFEYGKGRLEQRSQLLDDIQTVKPTHVFNAAGVTGRPNVDWCETHKPETIRTNV 468
SmRHM1 AMKLLIYGRTGWIGGLLGKICEREGIPYEYGSGRLENRLQLEADILTVKPTHIFNAAGVTGRPNVDWCESHKVETIRANV 465
VvRHM2 GLKFLIYGRTGWIGGLLGKLCKDGGIEFEYGKGRLQDRKTLMEDIRRVQPTHVFNAAGVTGRPNVDWCESHKVETIRTNV 451
PeRHM2 GLKFLIYGRTGWIGGLLGKLCKDGGIAFEYGQGRLEDRKSILKDIKKVNPTHVFNAAGVTGRPNVDWCESHKVETIRTNV 454
GmRHM3 SLKFLIYGGAGWIGGLIGNICEKQGIPFEYGMARLDDRSQILFDFRTIKPTHVFNASGVIGALNVKWFEAHKPETIRAVV 469


GmRHM2 VGTLTMADVCREHGILMVNYATGCIFEYNATHPEGSGIGFKEEDKPNFIGSFYSKTKAMVEELLREYDNVCTLRVRMPIS 542
GmRHM4 AGTLTMADVCREHGILMINYATGCIFEYNATHPEGSGIGFKEEDKPNFIGSFYSKTKAMVEELLRDYDNVCTLRVRMPIS 542
GmRHM1 AGTLTLADVSREHGLLMINYATGCIFEYDAAHPEGSGIGFKEEDRPNFFGSFYSKTKAMVEELLKEYDNVCTLRVRMPIS 542
VvRHM1 AGTLTLADVCREHGLLMMNFATGCIFEYDAAHPEGSGIGFKEEDTPNFAGSFYSKTKAMVEELLKEFDNVCTLRVRMPIS 546
GhRHM3 AGTLTLADVCREHGLLMMNFATGCIFEYDAGHPQGSGIGFKEEDKPNFTGSFYSKTKAMVEELLKEYNNVCTLRVRMPIS 541
GhRHM1 AGTLTLADVCREHGLLMMNFATGCIFEYDAGHPEGSGIGYKEEDKPNFTGSFYSKTKAMVEELLKEYDNVCTLRVRMPIS 541
GhRHM2 AGTLTLADVCRDHGLLMMNFATGCIFEYDAAHPQGSGIGFKEEDKPNFIGSFYSKTKAMVEELFKEYDNVCTLRVRMPIS 541
PeRHM1 TGTLTLADLCREHNLLMMNFATGCIFEYNASHPEGSGIGFTEEDKPNFIGSYYSKTKAMVEELLKEYDNVCTLRVRMPIS 544
RcRHM1 SGTLTLADVCREHNLLMMNFATGCIFEYDAAHPEGSGIGYTEDDKPNFTGSFYSKTKAMVEELLKEYDNVCTLRVRMPIS 521
AtRHM1 AGTLTLADVCREHGLLMMNFATGCIFEYDDKHPEGSGIGFKEEDTPNFTGSFYSKTKAMVEELLKEYDNVCTLRVRMPIS 543
CxcRHM2 VGTLTLADVCRENGLLMMNYATGCIFEYDAQHLEGSGIGFKEEDTPNFAGSFYSKTKAMVEELLKEYDNVCTLRVRMPIS 543
CxcRHM1 VGTLTLADVCRENGLLMMNYATGCIFEYDANHLEGSGIGFKEEDTPNFAGSFYSKTKAMVEELLKEYDNVCTLRVRMPIS 543
CxcRHM3 AGTLTLADVCREHNLLMMNYATGCIFEYDDKHREGSGIGFKEEDKPNFTGSFYSKTKAMVEELMKEFDNVCTLRVRMPIS 546
CxcRHM4 AGTLTLADVCREHNLLMMNYATGCIFEYDDKHREGSGIGFKEEDKPNFTGSFYSKTKAMVEELMKEFDNVCTLRVRMPIS 546
CxcRHM5 VGTLTLADVCREHNLLVINYATGCIFEYDDKHPEGSGIGFKEEDGPNFTGSFYSKTKAMVEELLKEYDNVCTLRVRMPIT 546
SbRHM1 VGTLNLADVCREHGLLMMNYATGCIFEYDADHPEGSGIGFKEEDTPNFTGSFYSKTKAMVEELLREYENVCTLRVRMPIS 537
BdRHM1 AGTLTLADVCREHGLLVMNYATGCIFEYDANHPEGSGIGFKEEDKPNFTGSFYSKTKAMVEELLKEYDNVCTLRVRMPIS 538
HvRHM1 VGTLTLADVCREHGLLVINYATGCIFEYDANHPEGSGIGFKEEDKPNFTGSFYSKTKAMVEELLKEYENVCTLRVRMPIS 538
AtRHM2 AGTLTLADVCRENDLLMMNFATGCIFEYDATHPEGSGIGFKEEDKPNFFGSFYSKTKAMVEELLREFDNVCTLRVRMPIS 541
AtRHM3 AGTLTLADVCRENDLLMMNFATGCIFEYDAAHPEGSGIGFKEEDKPNFTGSFYSKTKAMVEELLREFDNVCTLRVRMPIS 541
SbRHM3 VGTLNLADVCREQGLLMVNYATGCIFEYDAKHPEGSGIGFKEEDTPNFTGSFYSKTKAMVEELLKEYDNVCTLRVRMPIS 543
ZmRHM2 VGILNLADVCREQGLLMINYATGCIFEYDAKHPERSGIGFKEEDTPNFTGSFYSKTKAMVEELLKEYDNVCTLRVRMPIS 543
ZmRHM1 VGTLNLADVCRKQGLLMINYATGCIFEYDAKHPEGSGIGFKEEDTPNFIGSFYSKTKAMVEELLKEYDNVCTLRVRMPIS 543
OsRHM1 VGTLNLADVCREQGLLMINYATGCIFEYDAKHPEGSGIGFKEEDKPNFTGSYYSKTKAMVEELLQEYDNVCTLRVRMPIS 546
GhRHM4 VGTLTLADICREHDLLMINYATGCIFEYDATHPLGTGVGFKEEDKPNFTGSFYSKTKAVVEELLREFDNVCTLRVRMPIS 548
SmRHM1 VGTLTLADVCKQHGLLLMNFATGCIFEYDDSHPINSGIGFLEDDKPNFTGSFYSKTKAMVEELLKCYDNVCTLRVRMPIS 545
VvRHM2 LGTLTLADVCKEQGLLMMNFATGCIFEYDEEHPEGSGVGFKEEDKPNFVGSFYSKTKAMVEELLREYENVCTLRVRMPIS 531
PeRHM2 VGTLTLADVCKEHNLLMMSFATGCIFEYDQDHQEGSGIGFKEEDKPNFTGSFYSKTKAMVEELLREYENVCTLRVRMPIS 534
GmRHM3 VGVLTLADVCRDHGLLMMNYAFG----------GN------LEDKAYSTDAFYFRTQAKVEELLKEYENVCTLRIQLPVS 533
GmRHM2 SDLSNPRNFITKISRYNKVVNIPNSMTILDELLPISIEMAKRNLRGIWNFTNPGVVSHNEILEMYRDYIDPNFKWANFTL 622
GmRHM4 SDLSNPRNFITKISRYNKVVNIPNSMTILDELLPISIEMAKRNLKGIWNFTNPGVVSHNEILEMYRDYIDPNFKWSNFTL 622
GmRHM1 SDLSNPRNFITKISRYNKVVNIPNSMTILDELLPISIEMAKRNLRGIWNFTNPGAVSHNEILEMYRDYIDPSFKWANFNL 622
VvRHM1 SDLNNPRNFITKISRYNKVVNIPNSMTVLDELLPISIEMAKRNCRGIWNFTNPGVVSHNEILEMYKSYIDPNFKWANFTL 626
GhRHM3 SDLNNPRNFITKISRYSKVVNIPNSMTILDELLPISIEMAKRNLTGIWNFTNPGVVSHNEILEMYKKYIDPKFQWANFTL 621
GhRHM1 SDLNNPRNFITKIARYNKVVNIPNSMTILDELLPISIEMAKRNLNGIWNFTNPGVVSHNEILEMYKAYIDPKFQWVNFTL 621
GhRHM2 SDLNNPRNFITKISRYNKVVNIPNSMTVLDELLPISIEMAKRNLRGIWNFTNPGVVSHNEILEMYKTYIDPKFKWENFTL 621
PeRHM1 SDLNNPRNFITKISRYNKVVNIPNSMTILDELLPISIEMAKRNLRGIWNFTNPGVVSHNEILEMYKNYINPDFTWVNFDL 624
RcRHM1 SDLNNPRNFITKISRYNKVVNIPNSMTVLDELLPISIEMAKRNLRGIWNFTNPGVVSHNEILEMYKSCIDPDFKWVNFTL 601
AtRHM1 SDLNNPRNFITKISRYNKVVNIPNSMTVLDELLPISIEMAKRNLKGIWNFTNPGVVSHNEILEMYRDYINPEFKWANFTL 623
CxcRHM2 SDLNNPRNFITKISRYNKVVNIPNSMTVLDELLPISVEMAKRNLTGIWNFTNPGVVSHNEILEMYKKYIDPTFQWANFTL 623
CxcRHM1 SDLNNPRNFITKISRYNKVVNIPNSMTVLDELLPISVEMAKRNCRGIWNFTNPGAVSHNEILEMYKKYIDPTFKWENFTL 623
CxcRHM3 SDLNNPRNFITKISRYNKVVNIPNSMTILDELLPISIDMAKRNCRGIWNFTNPGVVSHNEILEMYKAYIDPRFKWANFTL 626
CxcRHM4 SDLNNPRNFITKISRYNKVVNIPNSMTILDELLPISIDMAKRNCRGIWNFTNPGVVSHNEILEMYKAYIDPNFKWTNFTL 626
CxcRHM5 SDLNNPRNFITKISRYNKVVNIPNSMTILDELLPISVEMAKRNCRGIWNFTNPGVVSHNEILEMYKAYIDPSFKWANFTL 626
SbRHM1 SDLNNPRNFVTKISRYNKVVNIPNSMTMLDELLPISVEMAKRNLRGIYNFTNPGVVSHNEILEMYKQYIDPSFKWTNFTL 617
BdRHM1 SDLNNPRNFITKISRYNKVVNIPNSMTVLDELLPISVEMAKRNLRGIWNFTNPGVVSHNEILEMYKKYLDPSYKWTNFTL 618
HvRHM1 SDLNNPRNFITKISRYDKVVNIPNSMTMLDELLPISVEMAKRNLRGIWNFTNPGVVSHNEILEMYKKYMDPSYKWTNFTL 618
AtRHM2 SDLNNPRNFITKISRYNKVVDIPNSMTVLDELLPISIEMAKRNLRGIWNFTNPGVVSHNEILEMYKNYIEPGFKWSNFTV 621
AtRHM3 SDLNNPRNFITKISRYNKVVNIPNSMTILDELLPISIEMAKRNLRGIWNFTNPGVVSHNEILEMYKSYIEPDFKWSNFNL 621
SbRHM3 SDLSNPRNFITKIARYDKVVNIPNSMTILDELLPISIEMAKQDCRGIWNFTNPGVVSHNEILEMYKKYINPDFKWTNFTL 623
ZmRHM2 SDLNNPRNFITKIARYDKVVNIPNSMTILDELLPISIEMAKRDCRGIWNFTNPGVVSHNEILEMYKKYINPDFKWTNFTL 623
ZmRHM1 SDLSNPRNFITKIARYDKVVNIPNSMTILDELLPISIEMAKRDCRGIWNFTNPGAVSHNEILEMYKKYINPDFKWTNFTL 623
OsRHM1 SDLSNPRNFITKIARYDKVVNIPNSMTILDELLPISIEMAKRDCRGIWNFTNPGVVSHNEILEMYKKYLNPDFKWTNFTL 626
GhRHM4 SDLSNPRNFITKITRYNKVVDISNSMTILDELLPVSIEMAKRNLRGIWNFTNPGVVSHNEILQMYKDYIDPNFNWVNFSL 628
SmRHM1 SDLSNPRNFITKITRYEKVVDIPNSMTVLDELLPLSIEMARRNCTGIWNFTNPGTVSHNEILEMYREYIDQGFKWSNFTL 625
VvRHM2 SDLSNPRNFITKIARYNKVVNIPNSMTVLDELLPISIEMAKRNCRGIWNFTNPGVVSHNEILEMYRDYIDPGFEWVNFNL 611
PeRHM2 SDLSNPRNFITKITRYDKVVNIPNSMTVLDELLPISIEMAKRNCRGIWNFTNPGVVSHNEILEMYRDYIDPEFKWMNFDL 614
GmRHM3 SDLSNPHNLITMITRSDKVANIPNSITVLDELVPISIEMAKRNCRGIWNFTNPGTVTCNEILEMYKDYVDPTFKWVNFTL 613

```
GmRHM2  EEQAKVIVAPRSNNEMDASKLKTEFPELLSIKESLIKYVFEPNKKT------- 668
GmRHM4  EEQAKVIVAPRSNNEMDASKLKTEFPELLSIKESLIKYVFEPNKKT------- 668
GmRHM1  EEQAKVIIAARSNNEMDASKLKNEFPELLSIKESLIKYVFEPNKKTA------ 669
VvRHM1  EEQAKVIVAARSNNEMDASKLKNEFPELLPIKDSLIKYVFEPNQKSLAA---- 675
GhRHM3  EEQAKVIVAPRSNNEMDASKLKKEFPELLPIKESLIKYVFEPNKRT------- 667
GhRHM1  EEQAKVIVAPRSNNEMDASKLKNEFPDLLPIKESLIKYVFEPNKRT------- 667
GhRHM2  EEQAKVIVAPRSNNEMDASKLKKEFPELLSIKESLIKYAFEPNRRT------- 667
PeRHM1  EEQAKVIVAPRSNNEMDASKLKNEFPELLPIKGSLIKYVFEPNKRT------- 670
RcRHM1  EEQAKVIVAPRSNNEMDAAKLKKEFPELLSIKESLIKYVFEANKKT------- 647
AtRHM1  EEQAKVIVAPRSNNEMDASKLKKEFPELLSIKESLIKYAYGPNKKT------- 669
CxcRHM2 EEQAKVIVAARSNNEMDATKLKTEFPELLSIKDSLIQYVFEPNRKVPAN---- 672
CxcRHM1 EEQAQVIVAARSNNEMDATKLKTEFPELLSIKDSLIKYVFEPNRKVPAN---- 672
CxcRHM3 EEQAKVIVAARSNNELDASKLKAEFPELLSIKDSLIKYVFEPNKKV------- 672
CxcRHM4 EEQAKVIVAPRSNNELDASKLKKEFPELLSIKDSLIKYVFEPNKKV------- 672
CxcRHM5 EEQAKVIVAARSNNELDASKLKAEFPELLSIKDSLIKYVFEPNKKV------- 672
SbRHM1  EEQAKVIVAPRSNNEMDATKLKKEFPELLSIKDSLIKFVFEPNRKVPIN---- 666
BdRHM1  EEQAKVIVAPRSNNEMDATKLKKEFPELLSIKDSLVKYVFEPNRKVPAN---- 667
HvRHM1  EEQAKVIVAPRSNNEMDAAKLKREFPELLSIKDSLIKYVFEPNRKVPAT---- 667
AtRHM2  EEQAKVIVAARSNNEMDGSKLSKEFPEMLSIKESLLKYVFEPNKRT------- 667
AtRHM3  EEQAKVIVAPRSNNEMDGAKLSKEFPEMLSIKDSLIKYVFEPNKRT------- 667
SbRHM3  EEQAKVIVAPRSNNEMDASKLKAEFPQLLSIKDSLIKYVFEPNRKVPSI---- 672
ZmRHM2  EEQSKVIVAPRSNNEMDASKLKAEFSQLMSIKDSLIKYVFEPNRKVPAN---- 672
ZmRHM1  EEQAKVIVAPRSNNEMDTSKLKAEFPQLLSIKDSLIKYVFEPNRKVPVN---- 672
OsRHM1  EEQAKVIVAPRSNNEMDASKLKSEFPELLSIKDSLVKYVFEPNRKVPAN---- 675
GhRHM4  QEQAKVIVAPRSNNELDASKLKNEFPELLSIKDSLIKYVFEPNRKTIAWEHSN 681
SmRHM1  EEQAKVIVAPRSNNELDTKKMEKEFPQLMHIKDSLVKLVFVPNSSSNGDKK-- 676
VvRHM2  EEQAKVIVAPRSNNELDASKLKKEFPELLSIKESIIKYVFEPNKKS------- 657
PeRHM2  EEQAKVIVAPRSNNELDGTKLKNEFPEMLSIKESILEYVFKPNKKT------- 660
GmRHM3  EQQAHVPSP--STNEMDSSKLKKEFPELLPVKDSLIKYVFEPEKKSLGF---- 660
```

Figure S3.6: Amino acid sequence alignment of NCBI-nr identified putative and characterized plant RHM. The alignment includes protein sequences of putative and characterized plant RHMs found in the NCBI-nr database. Amino acids with highlighted with blue background colour are those different from the consensus.

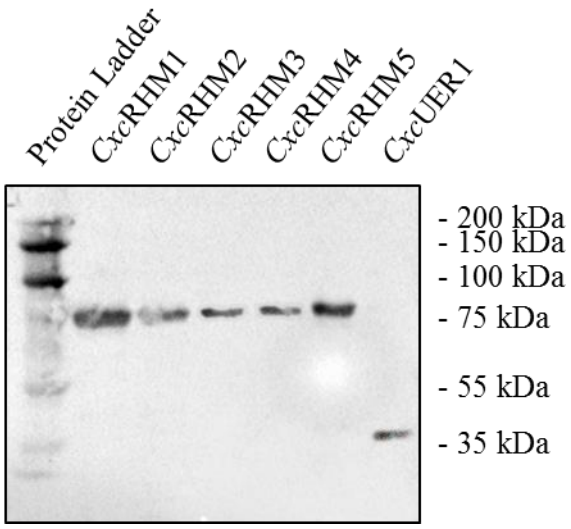


Figure S3.7: Western blot of purified *Cxc*RHM and *Cxc*UER1 enzymes after purification. Protein from *E. coli* expressing the full-length recombinant enzymes was blotted using antibody specific for the His6-tag.

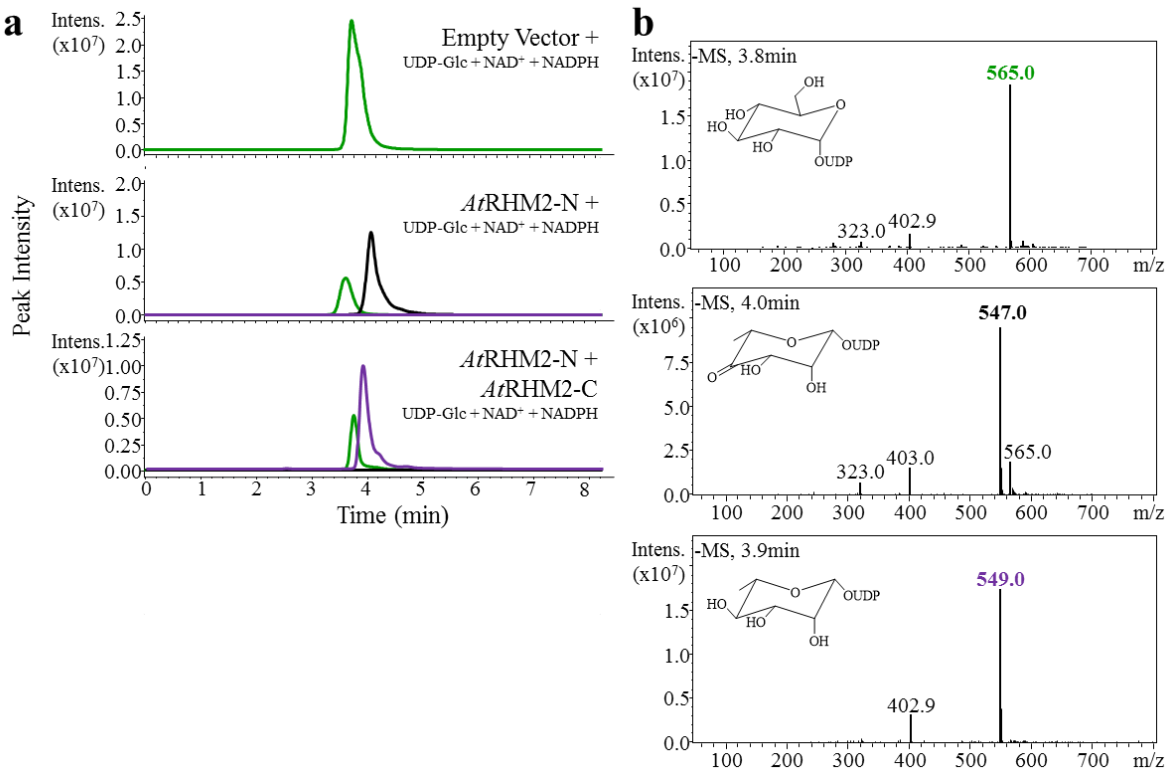


Figure S3.8: Select regions of extracted ion LC-MS traces and corresponding mass spectra of products formed by enzymatic reaction with *At*RHM2-N and *At*RHM2-C. (a) Purified protein derived from *E. coli* expressing a control vector, *At*RHM2-N, and *At*RHM2-C (*Oka Jigami 2007REF*) were incubated overnight with 1 mM UDP-glucose, 1mM NAD^+, and 1 mM NADPH. Green traces were confirmed by comparison of mass spectra and retention time against analytical standard to be UDP-glucose. Based on previously reported work by Oka *et al.* (2007), NMR has shown the black and purple traces correspond to UDP-4-keto-6-deoxy-glucose and UDP-rhamnose respectively. (b) Mass spectra analysis of UDP-sugars. Peak with negative ionization m/z of 565.0, 547.0, and 549.0 correspond to UDP-glucose, UDP-4-keto-6-deoxy-glucose, and UDP-rhamnose (the theoretical molecular weight of each is 565.30, 549.30, and 547.29), respectively. UDP-4-keto-6-deoxy-glucose is predicted to exist as both a keto and gem-diol pentose in aqueous solution, as is suggested by the presence of an ion with m/z of 565.0 [M-H](superscript -). Within reach spectra, the ion with m/z of 403.0 [M-H](superscript -) corresponds to UDP.

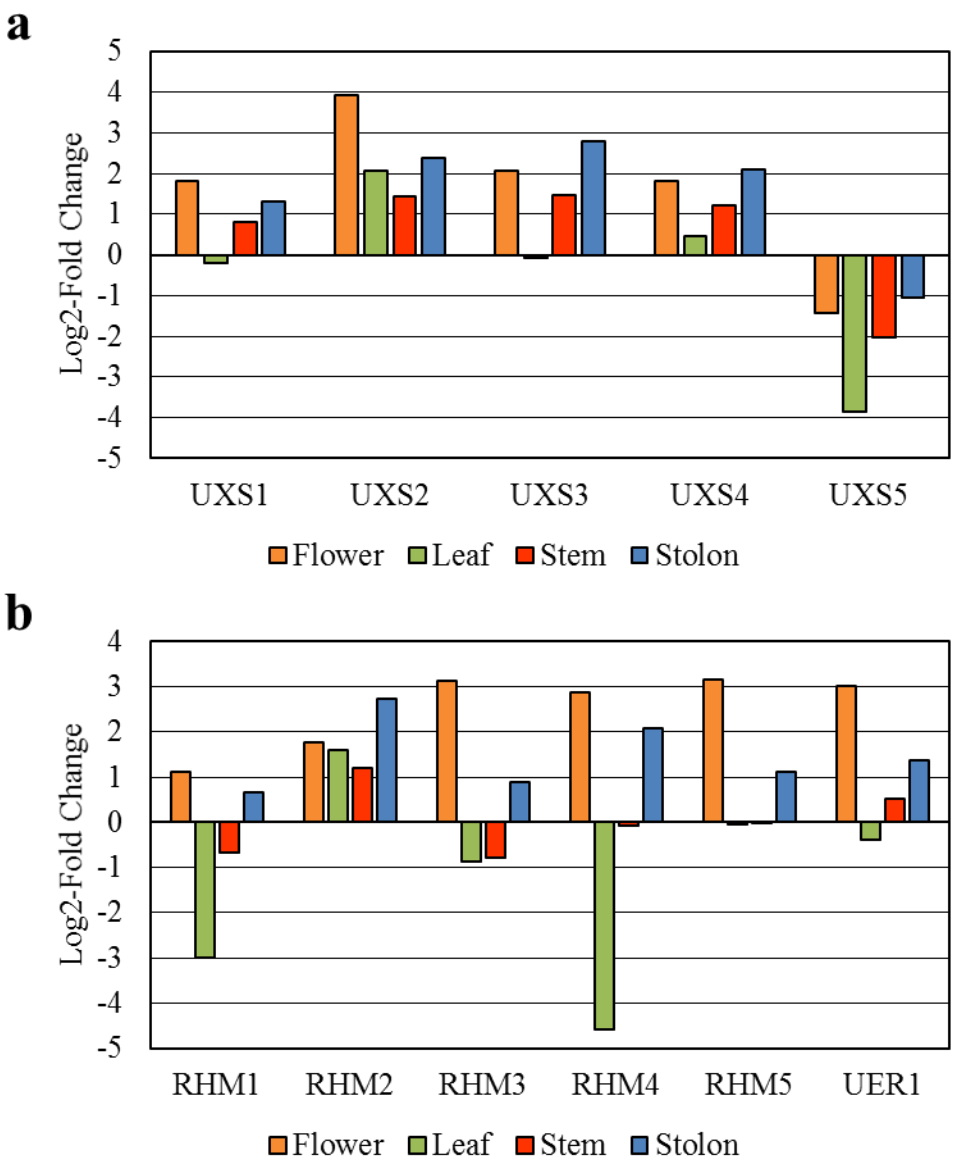


Figure S3.9: Differential Expression Analysis of *Cxc*UXS and *Cxc*RHM families. Estimated gene expression profiles for the *Cxc*UXS and *Cxc*RHM families were analyzed by in silico differential expression analysis of transcriptomic data to further explore potential roles they may play in C. x crocosmiiflora organs. To compare gene expression across multiple biological samples with an abundance estimation normalization method, a reference gene with constant expression across all biological samples was needed. Lacking such a gene, the expression levels for each unigene in corm samples were used as a reference, allowing the expression levels in the other organs to be compared to one another.

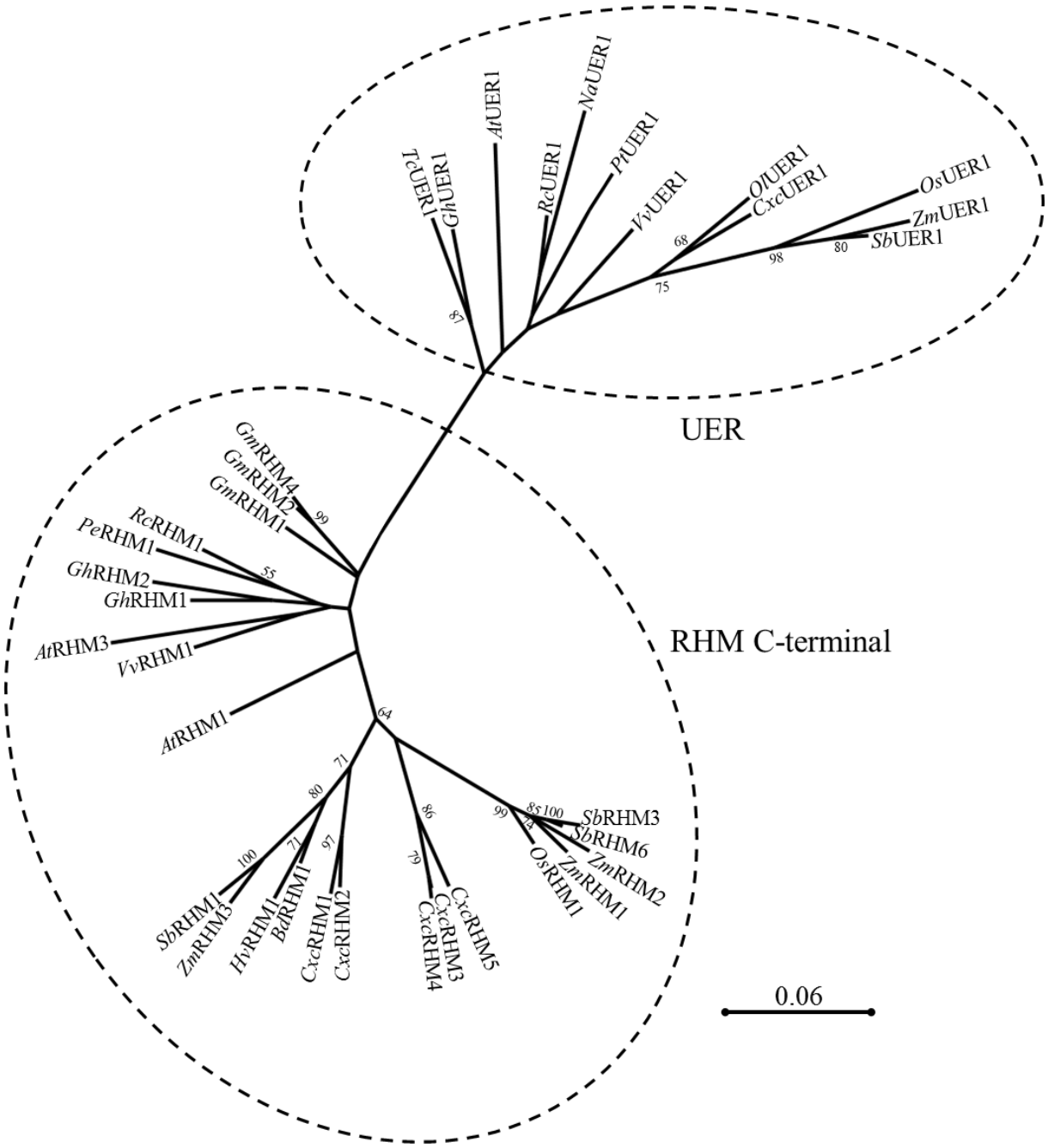
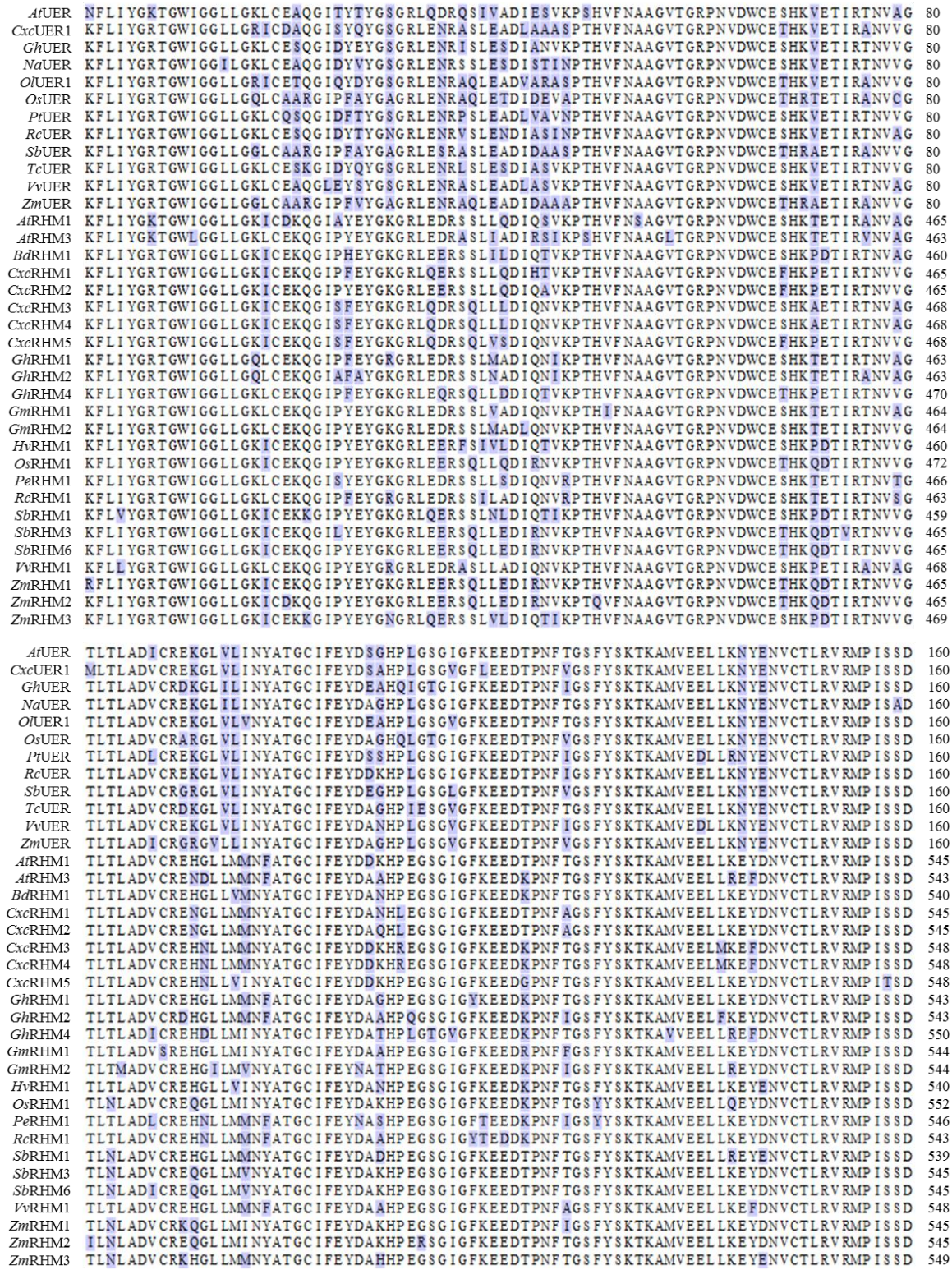
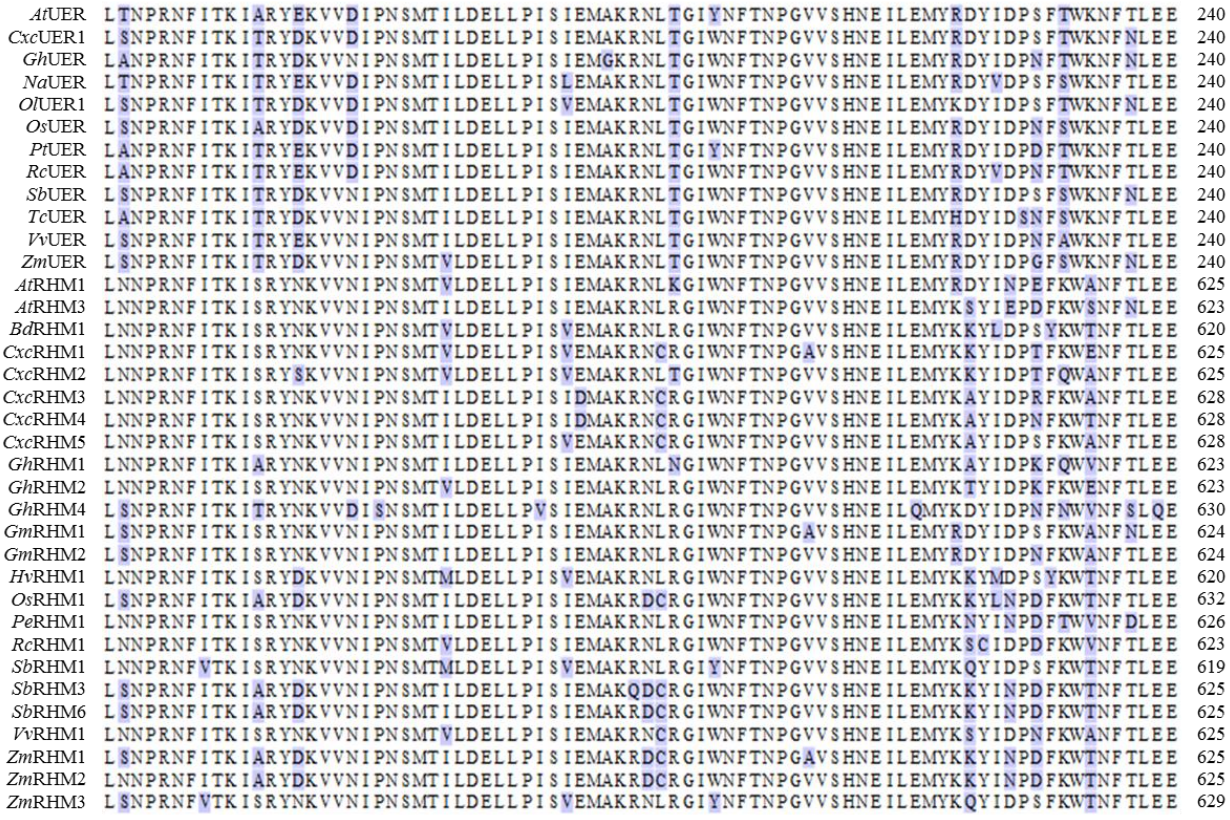


Figure S3.10: Phylogenetic analyses of plant UDP-rhamnose synthase C-terminal and UDP-4-keto-6-deoxy-glucose 3,5-epimerase/UDP-4-keto-rhamnose reductase. The maximum-likelihood tree was produced with characterized and putative plant RHM C-terminal and UER sequences obtained from the NCBI NR database using the MEGA 7.0 program (bootstrap value set at 1,000). Bootstrap values over 50% are indicated above the nodes. The black bar represents 0.06 amino acid substitutions per site. Protein alignment and sequences are given in Fig. S3.11 and Table S3.6.


AtUER NFLIYGKTGWIGGLLGKLCEAQGITYTYGSGRLQDRQSIVADIESVKPSHVFNAAGVTGRPNVDWCESHKVETIRTNVAG 80
CxcUER1 KFLIYGRTGWIGGLLGRICDAQGISYQYGSGRLENRASLEADLAAASPTHVFNAAGVTGRPNVDWCETHKVETIRANVVG 80
GhUER KFLIYGRTGWIGGLLGKLCESQGIDYEYGSGRLENRISLESDIANVKPTHVFNAAGVTGRPNVDWCESHKVETIRTNVVG 80
NaUER KFLIYGRTGWIGGILGKLCEAQGIDYVYGSGRLENRSSLESDISTINPTHVFNAAGVTGRPNVDWCESHKVETIRTNVVG 80
OlUER1 KFLIYGRTGWIGGLLGRICETQGIQYDYGSGRLENRAQLEADVARASPTHVFNAAGVTGRPNVDWCETHKVETIRANVVG 80
OsUER KFLIYGRTGWIGGLLGQLCAARGIPFAYGAGRLENRAQLETDIDEVAPTHVFNAAGVTGRPNVDWCETHRTETIRANVCG 80
PtUER KFLIYGRTGWIGGLLGKLCQSQGIDFTYGSGRLENRPSLEADLVAVNPTHVFNAAGVTGRPNVDWCESHKVETIRTNVVG 80
RcUER KFLIYGRTGWIGGLLGKLCESQGIDYTYGNGRLENRVSLENDIASINPTHVFNAAGVTGRPNVDWCESHKVETIRTNVAG 80
SbUER KFLIYGRTGWIGGLLGGLCAARGIPFAYGAGRLESRASLEADIDAASPTHVFNAAGVTGRPNVDWCETHRAETIRANVVG 80
TcUER KFLIYGRTGWIGGLLGKLCESKGIDYQYGSGRLENRLSLESDIASVKPTHVFNAAGVTGRPNVDWCESHKVETIRTNVVG 80
VvUER KFLIYGRTGWIGGLLGKLCEAQGLEYSYGSGRLENRASLEADLASVKPTHVFNAAGVTGRPNVDWCESHKVETIRTNVAG 80
ZmUER KFLIYGRTGWIGGLLGGLCAARGIPFVYGAGRLENRAQLEADIDAAAPTHVFNAAGVTGRPNVDWCETHRAETIRANVVG 80
AtRHM1 KFLIYGKTGWIGGLLGKICDKQGIAYEYGKGRLEDRSSLLQDIQSVKPTHVFNSAGVTGRPNVDWCESHKTETIRANVAG 465
AtRHM3 KFLIYGKTGWLGGLLGKLCEKQGIPYEYGKGRLEDRASLIADIRSIKPSHVFNAAGLTGRPNVDWCESHKTETIRVNVAG 463
BdRHM1 KFLIYGRTGWIGGLLGKICEKQGIPHEYGKGRLEERSSLILDIQTVKPTHVFNAAGVTGRPNVDWCESHKPDTIRTNVAG 460
CxcRHM1 KFLIYGRTGWIGGLLGKICEKQGIPFEYGKGRLQERSSLLQDIHTVKPTHVFNAAGVTGRPNVDWCEFHKPETIRTNVVG 465
CxcRHM2 KFLIYGRTGWIGGLLGKICEKQGIPYEYGKGRLEERSSLLQDIQAVKPTHVFNAAGVTGRPNVDWCEFHKPETIRTNVVG 465
CxcRHM3 KFLIYGRTGWIGGLLGKICEKQGISFEYGKGRLQDRSQLLLDIQNVKPTHVFNAAGVTGRPNVDWCESHKAETIRTNVAG 468
CxcRHM4 KFLIYGRTGWIGGLLGKICEKQGISFEYGKGRLQDRSQLLLDIQNVKPTHVFNAAGVTGRPNVDWCESHKAETIRTNVAG 468
CxcRHM5 KFLIYGRTGWIGGLLGKICEKQGISFEYGKGRLQDRSQLVSDIQNVKPTHVFNAAGVTGRPNVDWCEFHKPETIRTNVVG 468
GhRHM1 KFLIYGRTGWIGGLLGQLCEKQGIPFEYGRGRLEDRSSLMADIQNIKPTHVFNAAGVTGRPNVDWCESHKTETIRTNVAG 463
GhRHM2 KFLIYGRTGWIGGLLGQLCEKQGIAFAYGKGRLEDRSSLNADIQNIKPTHVFNAAGVTGRPNVDWCESHKTETIRANVAG 463
GhRHM4 KFLIYGRTGWIGGLLGKLCEKQGIPFEYGKGRLEQRSQLLDDIQTVKPTHVFNAAGVTGRPNVDWCETHKPETIRTNVVG 470
GmRHM1 KFLIYGRTGWIGGLLGKLCEKQGIPYEYGKGRLEDRSSLVADIQNVKPTHIFNAAGVTGRPNVDWCESHKTETIRTNVAG 464
GmRHM2 KFLIYGRTGWIGGLLGKLCEKQGIPYEYGKGRLEDRSSLMADLQNVKPTHVFNAAGVTGRPNVDWCESHKTETIRTNVVG 464
HvRHM1 KFLIYGRTGWIGGLLGKICEKQGIPYEYGKGRLEERFSIVLDIQTVKPTHVFNAAGVTGRPNVDWCESHKPDTIRTNVVG 460
OsRHM1 KFLIYGRTGWIGGLLGKICEKQGIPYEYGKGRLEERSQLLQDIRNVKPTHVFNAAGVTGRPNVDWCETHKQDTIRTNVVG 472
PeRHM1 KFLIYGRTGWIGGLLGKLCEKQGISYEYGKGRLEDRSSLLSDIQNVRPTHVFNAAGVTGRPNVDWCESHKTETIRTNVTG 466
RcRHM1 KFLIYGRTGWIGGLLGKLCEKQGIPFEYGRGRLEDRSSILADIQNVRPTHVFNAAGVTGRPNVDWCESHKTETIRTNVSG 463
SbRHM1 KFLVYGRTGWIGGLLGKICEKKGIPYEYGKGRLQERSSLNLDIQTIKPTHVFNAAGVTGRPNVDWCESHKPDTIRTNVVG 459
SbRHM3 KFLIYGRTGWIGGLLGKICEKQGILYEYGKGRLEERSQLLEDIRNVKPTHVFNAAGVTGRPNVDWCETHKQDTVRTNVVG 465
SbRHM6 KFLIYGRTGWIGGLLGKICEKQGIPYEYGKGRLEERSQLLEDIRNVKPTHVFNAAGVTGRPNVDWCETHKQDTIRTNVVG 465
VvRHM1 KFLLYGRTGWIGGLLGKLCEKQGIPYEYGRGRLEDRASLLADIQNVKPTHVFNAAGVTGRPNVDWCESHKPETIRANVAG 468
ZmRHM1 RFLIYGRTGWIGGLLGKICEKQGIPYEYGKGRLEERSQLLEDIRNVKPTHVFNAAGVTGRPNVDWCETHKQDTIRTNVVG 465
ZmRHM2 KFLIYGRTGWIGGLLGKICDKQGIPYEYGKGRLEERSQLLEDIRNVKPTQVFNAAGVTGRPNVDWCETHKQDTIRTNVVG 465
ZmRHM3 KFLIYGRTGWIGGLLGKICEKKGIPYEYGNGRLQERSSLVLDIQTIKPTHVFNAAGVTGRPNVDWCESHKPDTIRTNVVG 469
AtUER TLTLADICREKGLVLINYATGCIFEYDSGHPLGSGIGFKEEDTPNFTGSFYSKTKAMVEELLKNYENVCTLRVRMPISSD 160
CxcUER1 MLTLADVCREKGLVLINYATGCIFEYDSAHPLGSGVGFLEEDTPNFVGSFYSKTKAMVEELLKNYENVCTLRVRMPISSD 160
GhUER TLTLADVCRDKGLILINYATGCIFEYDEAHQIGTGIGFKEEDTPNFIGSFYSKTKAMVEELLKNYENVCTLRVRMPISSD 160
NaUER TLTLADVCREKGLILINYATGCIFEYDAGHPLGSGIGFKEEDTPNFTGSFYSKTKAMVEELLKNYENVCTLRVRMPISAD 160
OlUER1 TLTLADVCREKGLVLVNYATGCIFEYDEAHPLGSGVGFKEEDTPNFTGSFYSKTKAMVEELLKNYENVCTLRVRMPISSD 160
OsUER TLTLADVCRARGLVLINYATGCIFEYDAGHQLGTGIGFKEEDTPNFVGSFYSKTKAMVEELLKNYENVCTLRVRMPISSD 160
PtUER TLTLADLCREKGLVLINYATGCIFEYDSSHPLGSGIGFKEEDTPNFIGSFYSKTKAMVEDLLRNYENVCTLRVRMPISSD 160
RcUER TLTLADVCREKGLVLINYATGCIFEYDDKHPLGSGIGFKEEDTPNFIGSFYSKTKAMVEELLKNYENVCTLRVRMPISSD 160
SbUER TLTLADVCRGRGLVLINYATGCIFEYDEGHPLGSGLGFKEEDTPNFVGSFYSKTKAMVEELLKNYENVCTLRVRMPISSD 160
TcUER TLTLADVCRDKGLVLINYATGCIFEYDAGHPIESGVGFKEEDTPNFTGSFYSKTKAMVEELLKNYENVCTLRVRMPISSD 160
VvUER TLTLADVCREKGLVLINYATGCIFEYDANHPLGSGVGFKEEDTPNFIGSFYSKTKAMVEDLLKNYENVCTLRVRMPISSD 160
ZmUER TLTLADICRGRGVLLINYATGCIFEYDAGHPLGSGVGFKEEDTPNFVGSFYSKTKAMVEELLKNYENVCTLRVRMPISSD 160
AtRHM1 TLTLADVCREHGLLMMNFATGCIFEYDDKHPEGSGIGFKEEDTPNFTGSFYSKTKAMVEELLKEYDNVCTLRVRMPISSD 545
AtRHM3 TLTLADVCRENDLLMMNFATGCIFEYDAAHPEGSGIGFKEEDKPNFTGSFYSKTKAMVEELLREFDNVCTLRVRMPISSD 543
BdRHM1 TLTLADVCREHGLLVMNYATGCIFEYDANHPEGSGIGFKEEDKPNFTGSFYSKTKAMVEELLKEYDNVCTLRVRMPISSD 540
CxcRHM1 TLTLADVCRENGLLMMNYATGCIFEYDANHLEGSGIGFKEEDTPNFAGSFYSKTKAMVEELLKEYDNVCTLRVRMPISSD 545
CxcRHM2 TLTLADVCRENGLLMMNYATGCIFEYDAQHLEGSGIGFKEEDTPNFAGSFYSKTKAMVEELLKEYDNVCTLRVRMPISSD 545
CxcRHM3 TLTLADVCREHNLLMMNYATGCIFEYDDKHREGSGIGFKEEDKPNFTGSFYSKTKAMVEELMKEFDNVCTLRVRMPISSD 548
CxcRHM4 TLTLADVCREHNLLMMNYATGCIFEYDDKHREGSGIGFKEEDKPNFTGSFYSKTKAMVEELMKEFDNVCTLRVRMPISSD 548
CxcRHM5 TLTLADVCREHNLLVINYATGCIFEYDDKHPEGSGIGFKEEDGPNFTGSFYSKTKAMVEELLKEYDNVCTLRVRMPITSD 548
GhRHM1 TLTLADVCREHGLLMMNFATGCIFEYDAGHPEGSGIGYKEEDKPNFTGSFYSKTKAMVEELLKEYDNVCTLRVRMPISSD 543
GhRHM2 TLTLADVCRDHGLLMMNFATGCIFEYDAAHPQGSGIGFKEEDKPNFIGSFYSKTKAMVEELFKEYDNVCTLRVRMPISSD 543
GhRHM4 TLTLADICREHDLLMINYATGCIFEYDATHPLGTGVGFKEEDKPNFTGSFYSKTKAVVEELLREFDNVCTLRVRMPISSD 550
GmRHM1 TLTLADVSREHGLLMINYATGCIFEYDAAHPEGSGIGFKEEDRPNFFGSFYSKTKAMVEELLKEYDNVCTLRVRMPISSD 544
GmRHM2 TLTMADVCREHGILMVNYATGCIFEYNATHPEGSGIGFKEEDKPNFIGSFYSKTKAMVEELLREYDNVCTLRVRMPISSD 544
HvRHM1 TLTLADVCREHGLLVINYATGCIFEYDANHPEGSGIGFKEEDKPNFTGSFYSKTKAMVEELLKEYENVCTLRVRMPISSD 540
OsRHM1 TLNLADVCREQGLLMINYATGCIFEYDAKHPEGSGIGFKEEDKPNFTGSYYSKTKAMVEELLQEYDNVCTLRVRMPISSD 552
PeRHM1 TLTLADLCREHNLLMMNFATGCIFEYNASHPEGSGIGFTEEDKPNFIGSYYSKTKAMVEELLKEYDNVCTLRVRMPISSD 546
RcRHM1 TLTLADVCREHNLLMMNFATGCIFEYDAAHPEGSGIGYTEDDKPNFTGSFYSKTKAMVEELLKEYDNVCTLRVRMPISSD 543
SbRHM1 TLNLADVCREHGLLMMNYATGCIFEYDADHPEGSGIGFKEEDTPNFTGSFYSKTKAMVEELLREYENVCTLRVRMPISSD 539
SbRHM3 TLNLADVCREQGLLMVNYATGCIFEYDAKHPEGSGIGFKEEDTPNFTGSFYSKTKAMVEELLKEYDNVCTLRVRMPISSD 545
SbRHM6 TLNLADICREQGLLMVNYATGCIFEYDAKHPEGSGIGFKEEDTPNFTGSFYSKTKAMVEELLKEYDNVCTLRVRMPISSD 545
VvRHM1 TLTLADVCREHGLLMMNFATGCIFEYDAAHPEGSGIGFKEEDTPNFAGSFYSKTKAMVEELLKEFDNVCTLRVRMPISSD 548
ZmRHM1 TLNLADVCRKQGLLMINYATGCIFEYDAKHPEGSGIGFKEEDTPNFIGSFYSKTKAMVEELLKEYDNVCTLRVRMPISSD 545
ZmRHM2 ILNLADVCREQGLLMINYATGCIFEYDAKHPERSGIGFKEEDTPNFTGSFYSKTKAMVEELLKEYDNVCTLRVRMPISSD 545
ZmRHM3 TLNLADVCRKHGLLMMNYATGCIFEYDAHHPEGSGIGFKEEDTPNFTGSFYSKTKAMVEELLKEYENVCTLRVRMPISSD 549


AtUER LTNPRNFITKIARYEKVVDIPNSMTILDELLPISIEMAKRNLTGIYNFTNPGVVSHNEILEMYRDYIDPSFTWKNFTLEE 240
CxcUER1 LSNPRNFITKITRYDKVVDIPNSMTILDELLPISIEMAKRNLTGIWNFTNPGVVSHNEILEMYRDYIDPSFTWKNFNLEE 240
GhUER LANPRNFITKITRYDKVVNIPNSMTILDELLPISIEMGKRNLTGIWNFTNPGVVSHNEILEMYRDYIDPNFTWKNFNLEE 240
NaUER LTNPRNFITKITRYEKVVDIPNSMTILDELLPISLEMAKRNLTGIWNFTNPGVVSHNEILEMYRDYVDPSFSWKNFTLEE 240
OlUER1 LSNPRNFITKITRYDKVVDIPNSMTILDELLPISVEMAKRNLTGIWNFTNPGVVSHNEILEMYKDYIDPSFTWKNFNLEE 240
OsUER LSNPRNFITKIARYDKVVDIPNSMTILDELLPISIEMAKRNLTGIWNFTNPGVVSHNEILEMYRDYIDPNFSWKNFTLEE 240
PtUER LANPRNFITKITRYEKVVDIPNSMTILDELLPISIEMAKRNLTGIYNFTNPGVVSHNEILEMYRDYIDPDFTWKNFTLEE 240
RcUER LANPRNFITKITRYEKVVDIPNSMTILDELLPISIEMAKRNLTGIWNFTNPGVVSHNEILEMYRDYVDPNFTWKNFTLEE 240
SbUER LSNPRNFITKITRYDKVVNIPNSMTILDELLPISIEMAKRNLTGIWNFTNPGVVSHNEILEMYRDYIDPSFSWKNFNLEE 240
TcUER LANPRNFITKITRYDKVVNIPNSMTILDELLPISIEMAKRNLTGIWNFTNPGVVSHNEILEMYHDYIDSNFSWKNFTLEE 240
VvUER LSNPRNFITKITRYEKVVNIPNSMTILDELLPISIEMAKRNLTGIWNFTNPGVVSHNEILEMYRDYIDPNFAWKNFTLEE 240
ZmUER LSNPRNFITKITRYDKVVNIPNSMTVLDELLPISIEMAKRNLTGIWNFTNPGVVSHNEILEMYRDYIDPGFSWKNFNLEE 240
AtRHM1 LNNPRNFITKISRYNKVVNIPNSMTVLDELLPISIEMAKRNLKGIWNFTNPGVVSHNEILEMYRDYINPEFKWANFTLEE 625
AtRHM3 LNNPRNFITKISRYNKVVNIPNSMTILDELLPISIEMAKRNLRGIWNFTNPGVVSHNEILEMYKSYIEPDFKWSNFNLEE 623
BdRHM1 LNNPRNFITKISRYNKVVNIPNSMTVLDELLPISVEMAKRNLRGIWNFTNPGVVSHNEILEMYKKYLDPSYKWTNFTLEE 620
CxcRHM1 LNNPRNFITKISRYNKVVNIPNSMTVLDELLPISVEMAKRNCRGIWNFTNPGAVSHNEILEMYKKYIDPTFKWENFTLEE 625
CxcRHM2 LNNPRNFITKISRYSKVVNIPNSMTVLDELLPISVEMAKRNLTGIWNFTNPGVVSHNEILEMYKKYIDPTFQWANFTLEE 625
CxcRHM3 LNNPRNFITKISRYNKVVNIPNSMTILDELLPISIDMAKRNCRGIWNFTNPGVVSHNEILEMYKAYIDPRFKWANFTLEE 628
CxcRHM4 LNNPRNFITKISRYNKVVNIPNSMTILDELLPISIDMAKRNCRGIWNFTNPGVVSHNEILEMYKAYIDPNFKWTNFTLEE 628
CxcRHM5 LNNPRNFITKISRYNKVVNIPNSMTILDELLPISVEMAKRNCRGIWNFTNPGVVSHNEILEMYKAYIDPSFKWANFTLEE 628
GhRHM1 LNNPRNFITKIARYNKVVNIPNSMTILDELLPISIEMAKRNLNGIWNFTNPGVVSHNEILEMYKAYIDPKFQWVNFTLEE 623
GhRHM2 LNNPRNFITKISRYNKVVNIPNSMTVLDELLPISIEMAKRNLRGIWNFTNPGVVSHNEILEMYKTYIDPKFKWENFTLEE 623
GhRHM4 LSNPRNFITKITRYNKVVDISNSMTILDELLPVSIEMAKRNLRGIWNFTNPGVVSHNEILQMYKDYIDPNFNWVNFSLQE 630
GmRHM1 LSNPRNFITKISRYNKVVNIPNSMTILDELLPISIEMAKRNLRGIWNFTNPGAVSHNEILEMYRDYIDPSFKWANFNLEE 624
GmRHM2 LSNPRNFITKISRYNKVVNIPNSMTILDELLPISIEMAKRNLRGIWNFTNPGVVSHNEILEMYRDYIDPNFKWANFTLEE 624
HvRHM1 LNNPRNFITKISRYDKVVNIPNSMTMLDELLPISVEMAKRNLRGIWNFTNPGVVSHNEILEMYKKYMDPSYKWTNFTLEE 620
OsRHM1 LSNPRNFITKIARYDKVVNIPNSMTILDELLPISIEMAKRDCRGIWNFTNPGVVSHNEILEMYKKYLNPDFKWTNFTLEE 632
PeRHM1 LNNPRNFITKISRYNKVVNIPNSMTILDELLPISIEMAKRNLRGIWNFTNPGVVSHNEILEMYKNYINPDFTWVNFDLEE 626
RcRHM1 LNNPRNFITKISRYNKVVNIPNSMTVLDELLPISIEMAKRNLRGIWNFTNPGVVSHNEILEMYKSCIDPDFKWVNFTLEE 623
SbRHM1 LNNPRNFVTKISRYNKVVNIPNSMTMLDELLPISVEMAKRNLRGIYNFTNPGVVSHNEILEMYKQYIDPSFKWTNFTLEE 619
SbRHM3 LSNPRNFITKIARYDKVVNIPNSMTILDELLPISIEMAKQDCRGIWNFTNPGVVSHNEILEMYKKYINPDFKWTNFTLEE 625
SbRHM6 LSNPRNFITKIARYDKVVNIPNSMTILDELLPISIEMAKRDCRGIWNFTNPGVVSHNEILEMYKKYINPDFKWTNFTLEE 625
VvRHM1 LNNPRNFITKISRYNKVVNIPNSMTVLDELLPISIEMAKRNCRGIWNFTNPGVVSHNEILEMYKSYIDPNFKWANFTLEE 625
ZmRHM1 LSNPRNFITKIARYDKVVNIPNSMTILDELLPISIEMAKRDCRGIWNFTNPGAVSHNEILEMYKKYINPDFKWTNFTLEE 625
ZmRHM2 LNNPRNFITKIARYDKVVNIPNSMTILDELLPISIEMAKRDCRGIWNFTNPGVVSHNEILEMYKKYINPDFKWTNFTLEE 625
ZmRHM3 LSNPRNFVTKISRYNKVVNIPNSMTILDELLPISVEMAKRNLRGIYNFTNPGVVSHNEILEMYKQYIDPSFKWTNFTLEE 629

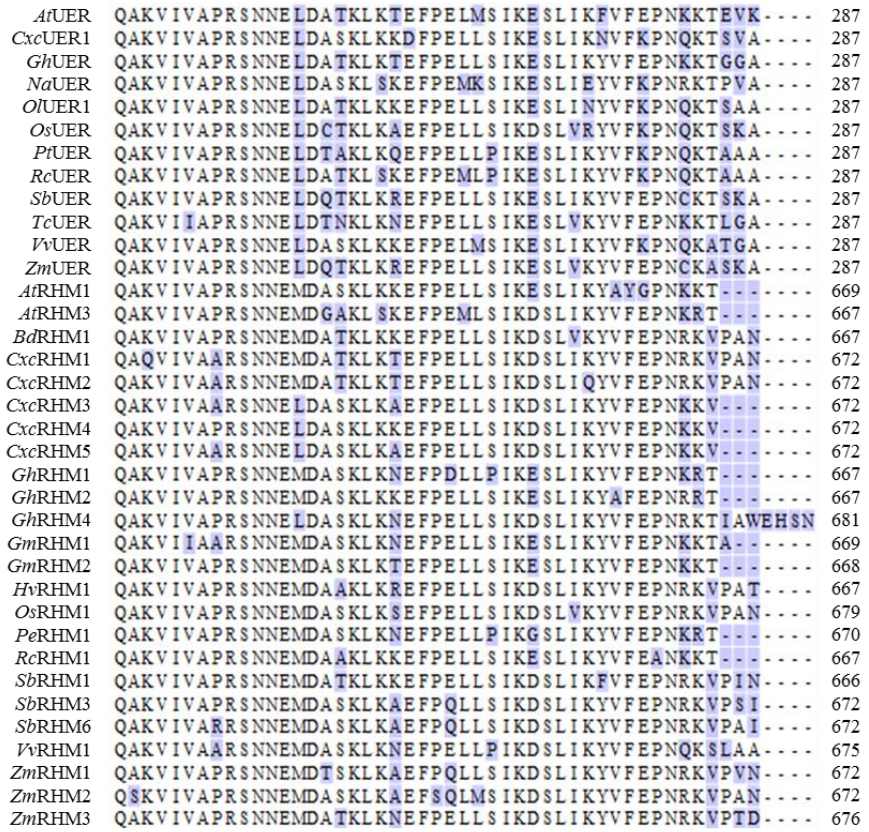

AtUER QAKVIVAPRSNNELDATKLKTEFPELMSIKESLIKFVFEPNKKTEVK---- 287
CxcUER1 QAKVIVAPRSNNELDASKLKKDFPELLSIKESLIKNVFKPNQKTSVA---- 287
GhUER QAKVIVAPRSNNELDATKLKTEFPELLSIKESLIKYVFEPNKKTGGA---- 287
NaUER QAKVIVAPRSNNELDASKLSKEFPEMKSIKESLIEYVFKPNRKTPVA---- 287
OlUER1 QAKVIVAPRSNNELDATKLKKEFPELLSIKESLINYVFKPNQKTSAA---- 287
OsUER QAKVIVAPRSNNELDCTKLKAEFPELLSIKDSLVRYVFKPNQKTSKA---- 287
PtUER QAKVIVAPRSNNELDTAKLKQEFPELLPIKESLIKYVFKPNQKTAAA---- 287
RcUER QAKVIVAPRSNNELDATKLSKEFPEMLPIKESLIKYVFKPNQKTAAA---- 287
SbUER QAKVIVAPRSNNELDQTKLKREFPELLSIKESLIKYVFEPNCKTSKA---- 287
TcUER QAKVIIAPRSNNELDTNKLKNEFPELLSIKESLVKYVFEPNKKTLGA---- 287
VvUER QAKVIVAPRSNNELDASKLKKEFPELMSIKESLIKYVFKPNQKATGA---- 287
ZmUER QAKVIVAPRSNNELDQTKLKREFPELLSIKESLVKYVFEPNCKASKA---- 287
AtRHM1 QAKVIVAPRSNNEMDASKLKKEFPELLSIKESLIKYAYGPNKKT-------- 669
AtRHM3 QAKVIVAPRSNNEMDGAKLSKEFPEMLSIKDSLIKYVFEPNKRT-------- 667
BdRHM1 QAKVIVAPRSNNEMDATKLKKEFPELLSIKDSLVKYVFEPNRKVPAN---- 667
CxcRHM1 QAQVIVAARSNNEMDATKLKTEFPELLSIKDSLIKYVFEPNRKVPAN---- 672
CxcRHM2 QAKVIVAARSNNEMDATKLKTEFPELLSIKDSLIQYVFEPNRKVPAN---- 672
CxcRHM3 QAKVIVAARSNNELDASKLKAEFPELLSIKDSLIKYVFEPNKKV-------- 672
CxcRHM4 QAKVIVAPRSNNELDASKLKKEFPELLSIKDSLIKYVFEPNKKV-------- 672
CxcRHM5 QAKVIVAARSNNELDASKLKAEFPELLSIKDSLIKYVFEPNKKV-------- 672
GhRHM1 QAKVIVAPRSNNEMDASKLKNEFPDLLPIKESLIKYVFEPNKRT-------- 667
GhRHM2 QAKVIVAPRSNNEMDASKLKKEFPELLSIKESLIKYAFEPNRRT-------- 667
GhRHM4 QAKVIVAPRSNNELDASKLKNEFPELLSIKDSLIKYVFEPNRKTIAWEHSN 681
GmRHM1 QAKVIIAARSNNEMDASKLKNEFPELLSIKESLIKYVFEPNKKTA------ 669
GmRHM2 QAKVIVAPRSNNEMDASKLKTEFPELLSIKESLIKYVFEPNKKT-------- 668
HvRHM1 QAKVIVAPRSNNEMDAAKLKREFPELLSIKDSLIKYVFEPNRKVPAT---- 667
OsRHM1 QAKVIVAPRSNNEMDASKLKSEFPELLSIKDSLVKYVFEPNRKVPAN---- 679
PeRHM1 QAKVIVAPRSNNEMDASKLKNEFPELLPIKGSLIKYVFEPNKRT-------- 670
RcRHM1 QAKVIVAPRSNNEMDAAKLKKEFPELLSIKESLIKYVFEANKKT-------- 667
SbRHM1 QAKVIVAPRSNNEMDATKLKKEFPELLSIKDSLIKFVFEPNRKVPIN---- 666
SbRHM3 QAKVIVAPRSNNEMDASKLKAEFPQLLSIKDSLIKYVFEPNRKVPSI---- 672
SbRHM6 QAKVIVARRSNNEMDASKLKAEFPQLLSIKDSLIKYVFEPNRKVPAI---- 672
VvRHM1 QAKVIVAARSNNEMDASKLKNEFPELLPIKDSLIKYVFEPNQKSLAA---- 675
ZmRHM1 QAKVIVAPRSNNEMDTSKLKAEFPQLLSIKDSLIKYVFEPNRKVPVN---- 672
ZmRHM2 QSKVIVAPRSNNEMDASKLKAEFSQLMSIKDSLIKYVFEPNRKVPAN---- 672
ZmRHM3 QAKVIVAPRSNNEMDATKLKNEFPELLSIKDSLIKYVFEPNRKVPTD---- 676

Figure S3.11: Amino acid sequence alignment of NCBI-nr identified putative and characterized plant RHM and UER. The alignment includes protein sequences of putative and characterized plant UXSs found in the NCBI-nr database. Amino acids with highlighted with blue background colour are those different from the consensus.

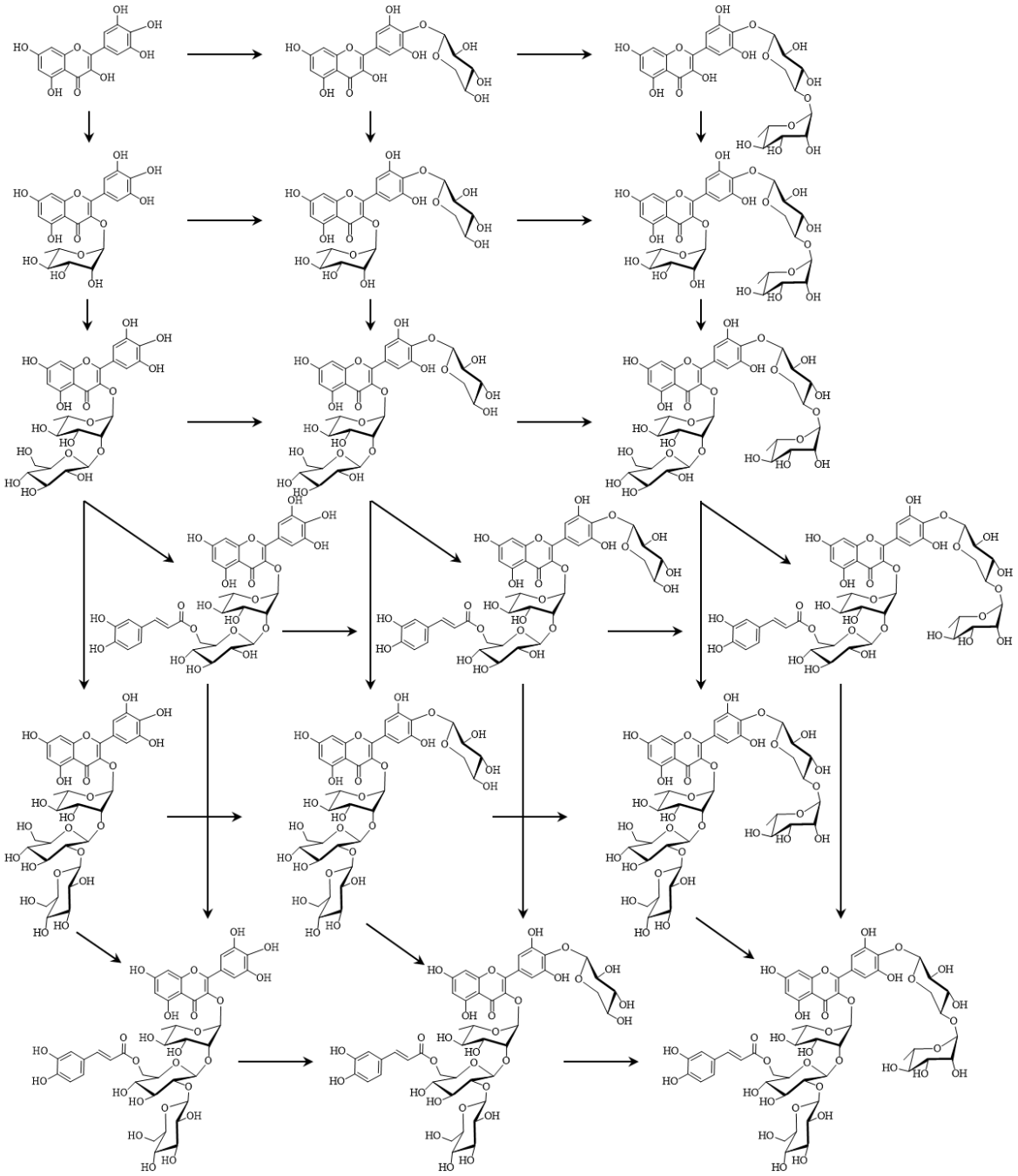


Figure S4.1: All possible montbretin A biosynthetic routes. Considering all potential reactions needed to form montbretin A and lacking any knowledge on intermediates of the pathway, the biosynthetic pathway is better represented by a multi-dimensional matrix than linear pathway.

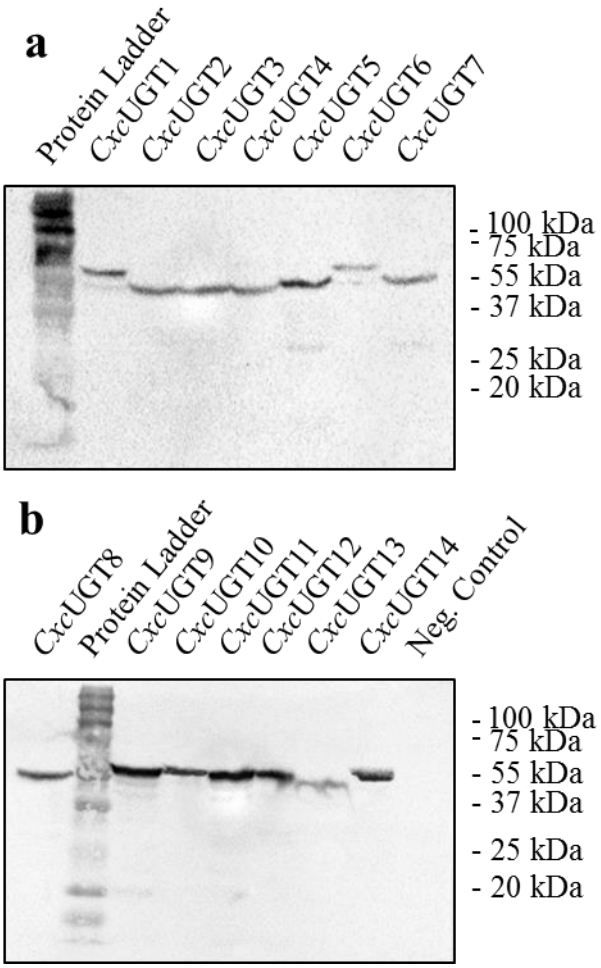


Figure S4.2: Western blot of purified *Cxc*UGTs. Protein from *E. coli* expressing the full-length recombinant enzymes was blotted using antibody specific for the His_6-tag.

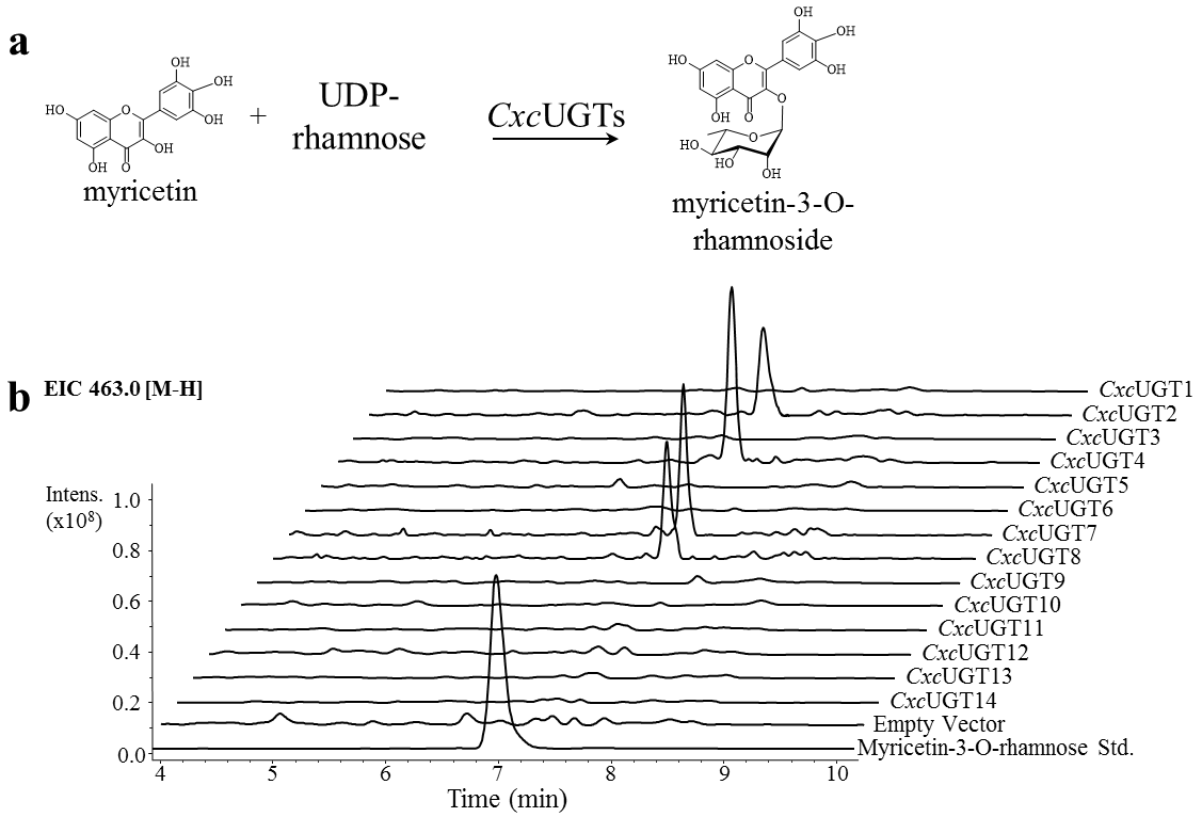


Figure S4.3: Select regions of extracted ion LC-MS chromatograph for *C. x crocosmiiflora* GT1 UGT enzyme assays. (a) Substrate acceptor and UDP-sugar donor being tested for activity against *Cxc*UGTs. (b) Protein derived from *E. coli* expressing a control vector and the 14 GT1 UGTs of interest were incubated overnight with 1 mM UDP-glucose, 1 mM NAD^+, 1 mM NADPH, 6.5 μM purified *Cxc*RHM1, and 100 μM myricetin and assessed for their ability to form myricetin-rhamnoside (theoretical molecular weight of 464.38). Traces shown for each sample is the extracted ion chromatograph for m/z of 463.0 [M-H](superscript -).

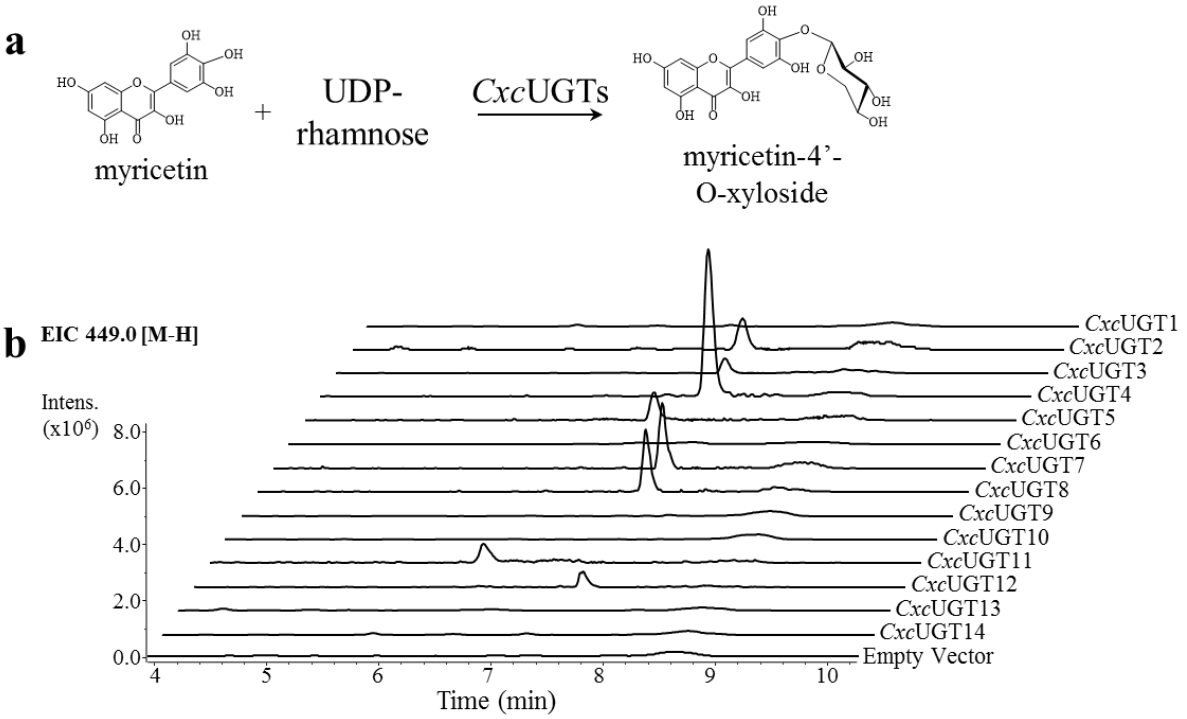


Figure S4.4: Representative regions of extracted ion LC-MS chromatograph for *C. x crocosmiiflora* GT1 UGT enzyme assays. (a) Substrate acceptor and UDP-sugar donor being tested for activity against *Cxc*UGTs. (b) Protein derived from *E. coli* expressing a control vector and the 14 GT1 UGTs of interest were incubated overnight with 1 mM UDP-glucuronic acid, 1 mM NAD^+, 1 μM purified *Cxc*UXS4, and 100 μM myricetin and assessed for their ability to produce myricetin-xyloside (theoretical molecular weight of 450.35). Traces shown for each sample is the extracted ion chromatograph for m/z of 449.0 [M-H](superscript -).

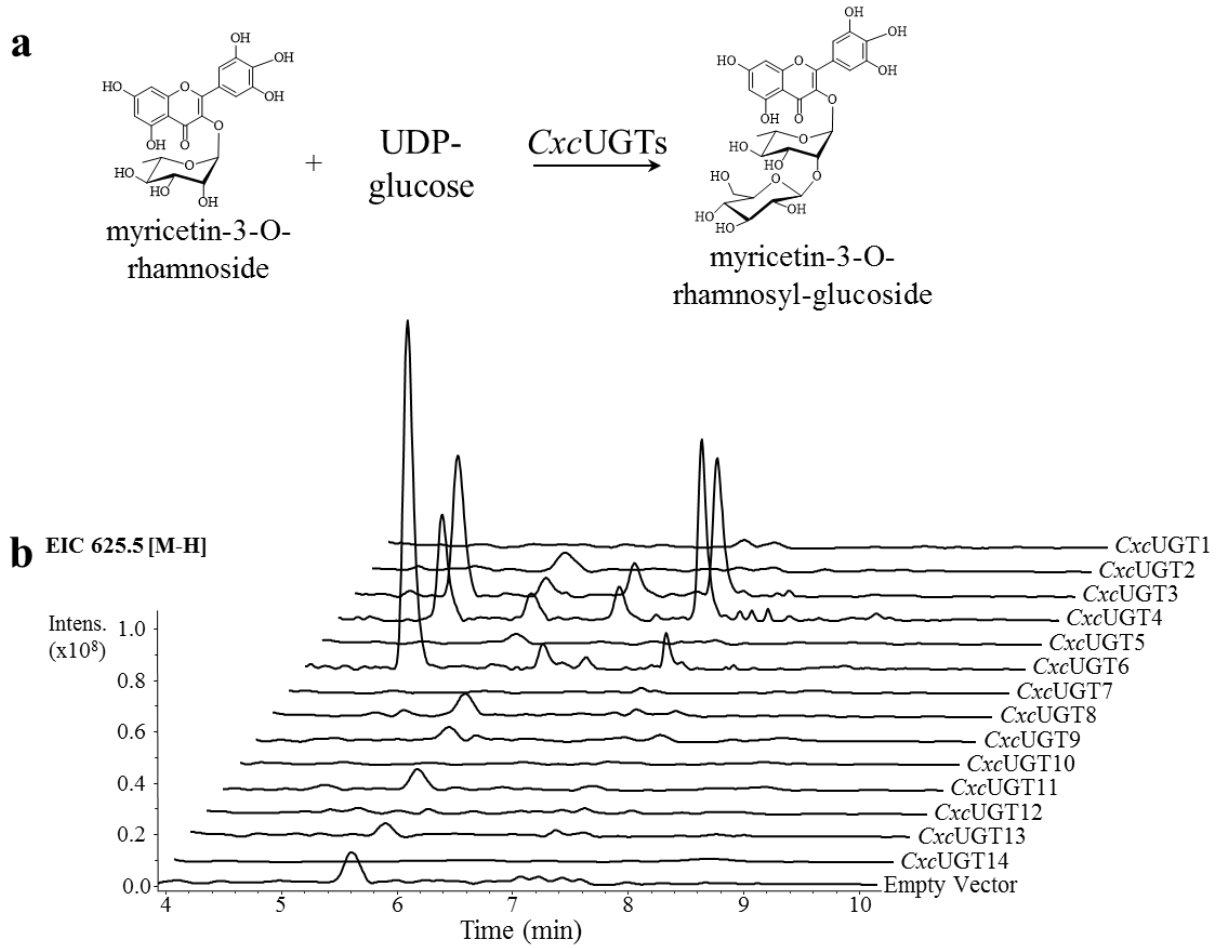


Figure S4.5: Representative regions of extracted ion LC-MS chromatograph for *C. x crocosmiiflora* GT1 UGT enzyme assays. (a) Substrate acceptor and UDP-sugar donor being tested for activity against *Cxc*UGTs. (b) Protein derived from *E. coli* expressing a control vector and the 14 GT1 UGTs of interest were incubated overnight with 1 mM UDP-glucose and 100 μM myricetin-3-O-rhamnose for their ability to produce a myricetin-3-O-rhamnose glucoside (theoretical molecular weight of 626.52). Traces shown for each sample is the extracted ion chromatograph for m/z of 625.5 [M-H](superscript -).

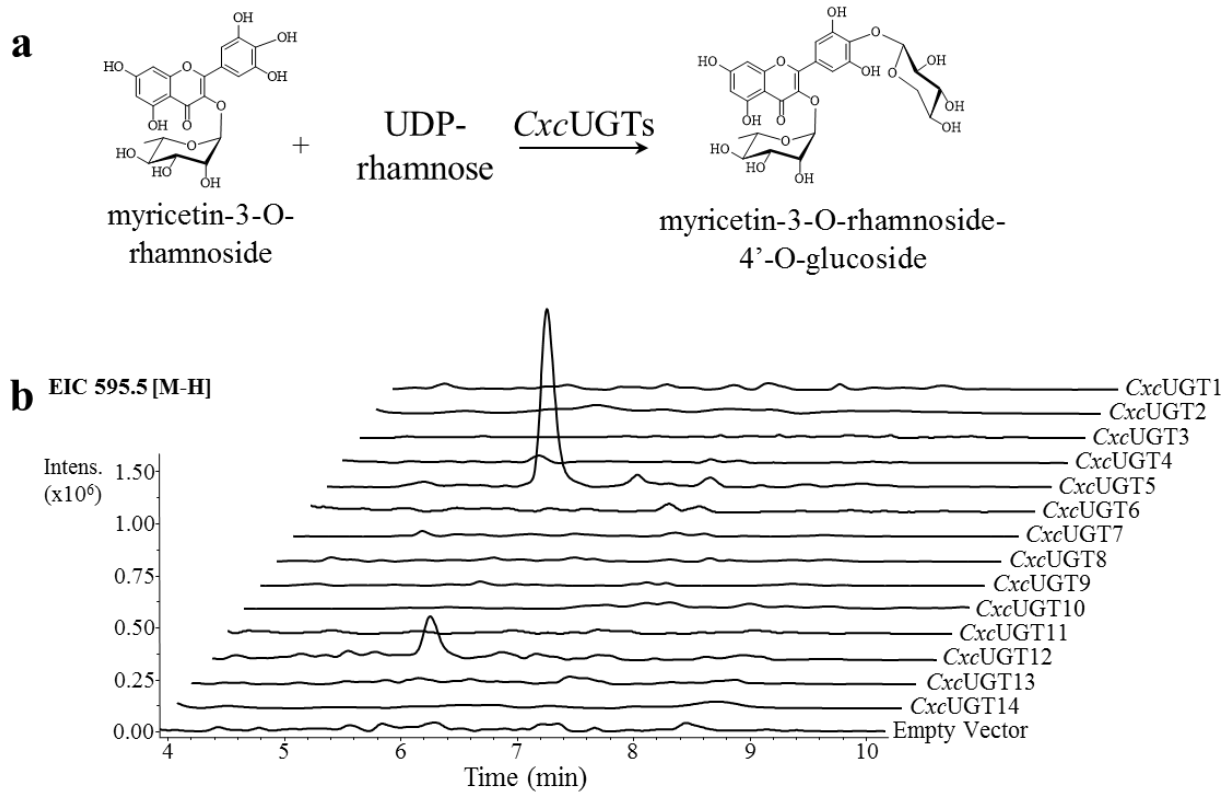


Figure S4.6: Representative regions of extracted ion LC-MS chromatograph for *C. x crocosmiiflora* GT1 UGT enzyme assays. (a) Substrate acceptor and UDP-sugar donor being tested for activity against *Cxc*UGTs. (b) Protein derived from *E. coli* expressing a control vector and the 14 GT1 UGTs of interest were incubated overnight with 1 mM UDP-glucuronic acid, 1 mM NAD^+, 1 μM purified *Cxc*UXS4, and 100 μM myricetin-3-O-rhamnose for their ability to produce a myricetin-3-O-rhamnose xyloside (theoretical molecular weight of 596.49). Traces shown for each sample is the extracted ion chromatograph for m/z of 595.5 [M-H](superscript -).

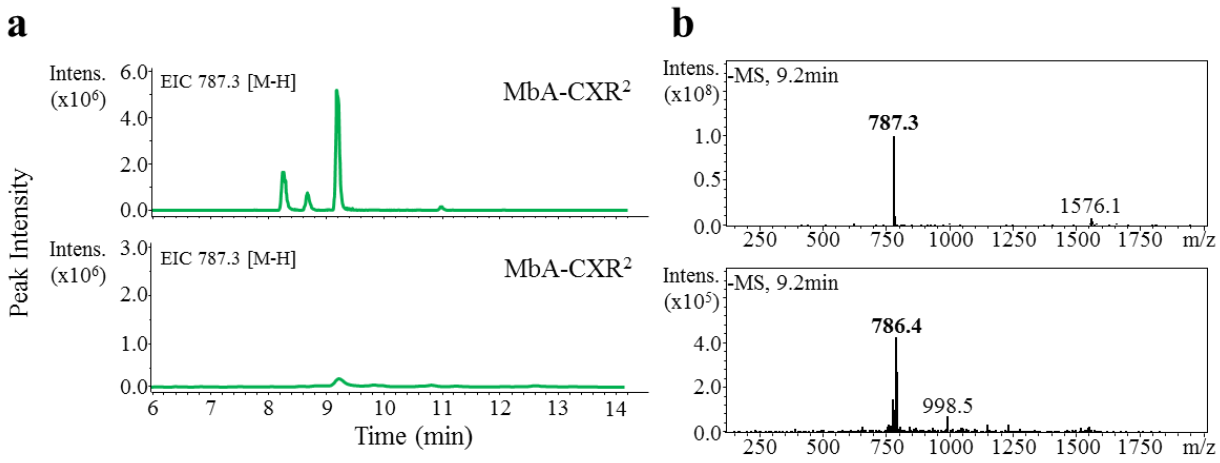


Figure S4.7: Representative regions of extracted ion LC-MS chromatograph and corresponding mass spectra of potential montbretin A intermediates. (a) Extracted ion chromatographs of purified MbA-CXR′ run (i) directly after purification (April 2016) and (ii) during enzyme assays of the 14 candidate GT1 UGTs (January 2017). Traces shown for each sample is the extracted ion chromatograph for m/z of 787.3 [M-H](superscript -). (b) Mass spectra of 787.3 peaks. Peaks with m/z of 787.3 [M-H](superscript -) is believed to correspond to MbA-CXR^2 (theoretical molecular weight of 788.66).

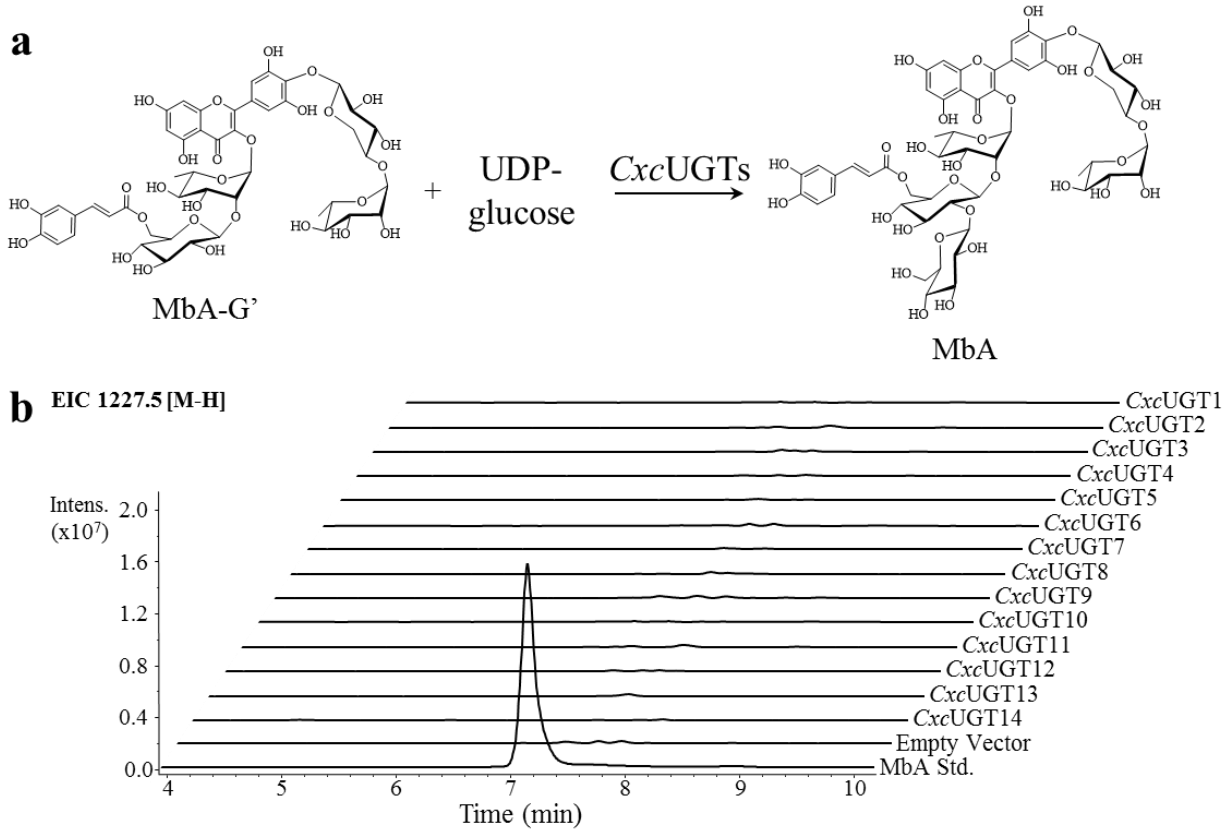


Figure S4.8: Representative regions of extracted ion LC-MS chromatograph for *C. x crocosmiiflora* GT1 UGT enzyme assays. (a) Substrate acceptor and UDP-sugar donor being tested for activity against *Cxc*UGTs. (b) Protein derived from *E. coli* expressing a control vector and the 14 GT1 UGTs of interest were incubated overnight with 1 mM UDP-glucose, and 100 μM MbA-G′ and assessed for their ability to produced MbA (theoretical molecular weight of 1229.03). Traces shown for each sample is the extracted ion chromatograph for m/z of 1227.5 [M-H](superscript -).

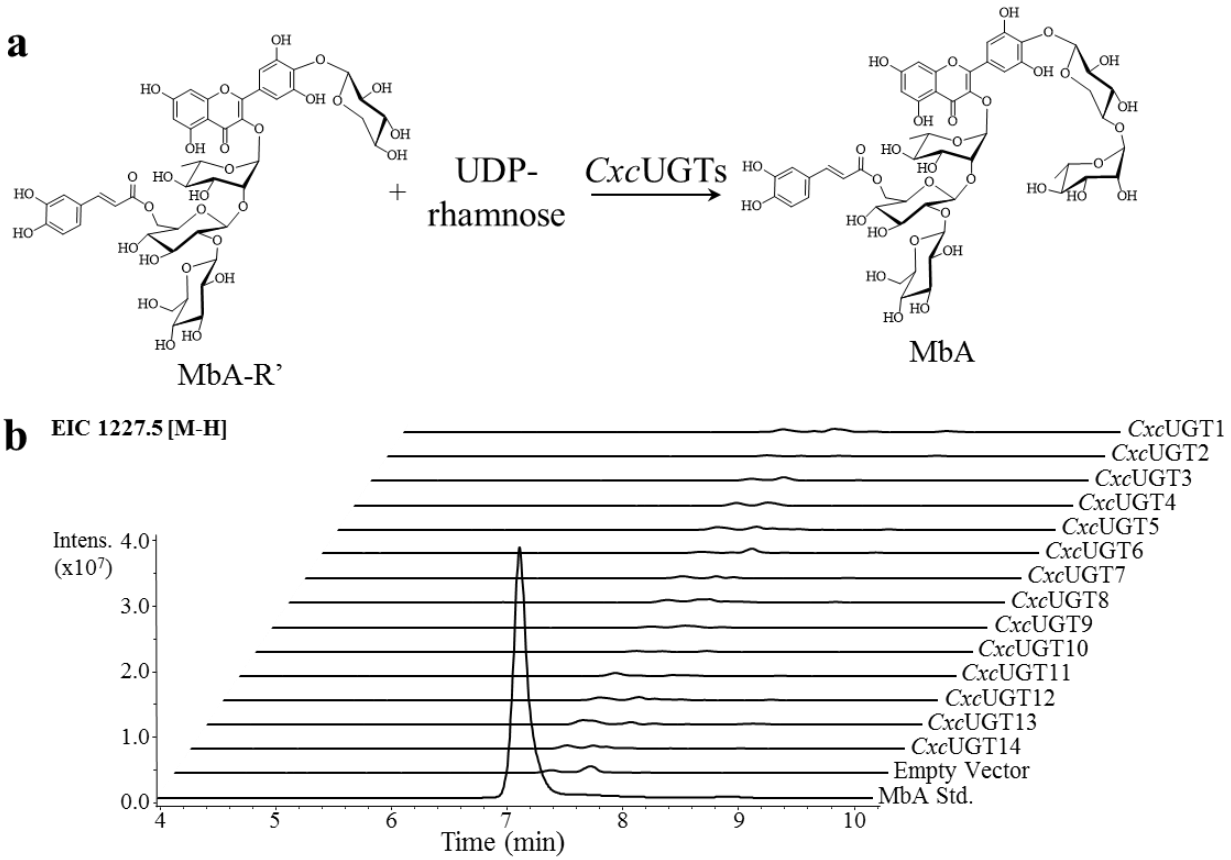


Figure S4.9: Representative regions of extracted ion LC-MS chromatograph for *C. x crocosmiiflora* GT1 UGT enzyme assays. (a) Substrate acceptor and UDP-sugar donor being tested for activity against *Cxc*UGTs. (b) Protein derived from *E. coli* expressing a control vector and the 14 GT1 UGTs of interest were incubated overnight with 1 mM UDP-glucose, 1 mM NAD^+, 1 mM NADPH, 6.5 μM purified *Cxc*RHM1, and 100 μM MbA-R′ and assessed for their ability to produce MbA (theoretical molecular weight of 1229.03). Traces shown for each sample is the extracted ion chromatograph for m/z of 1227.5 [M-H](superscript -).

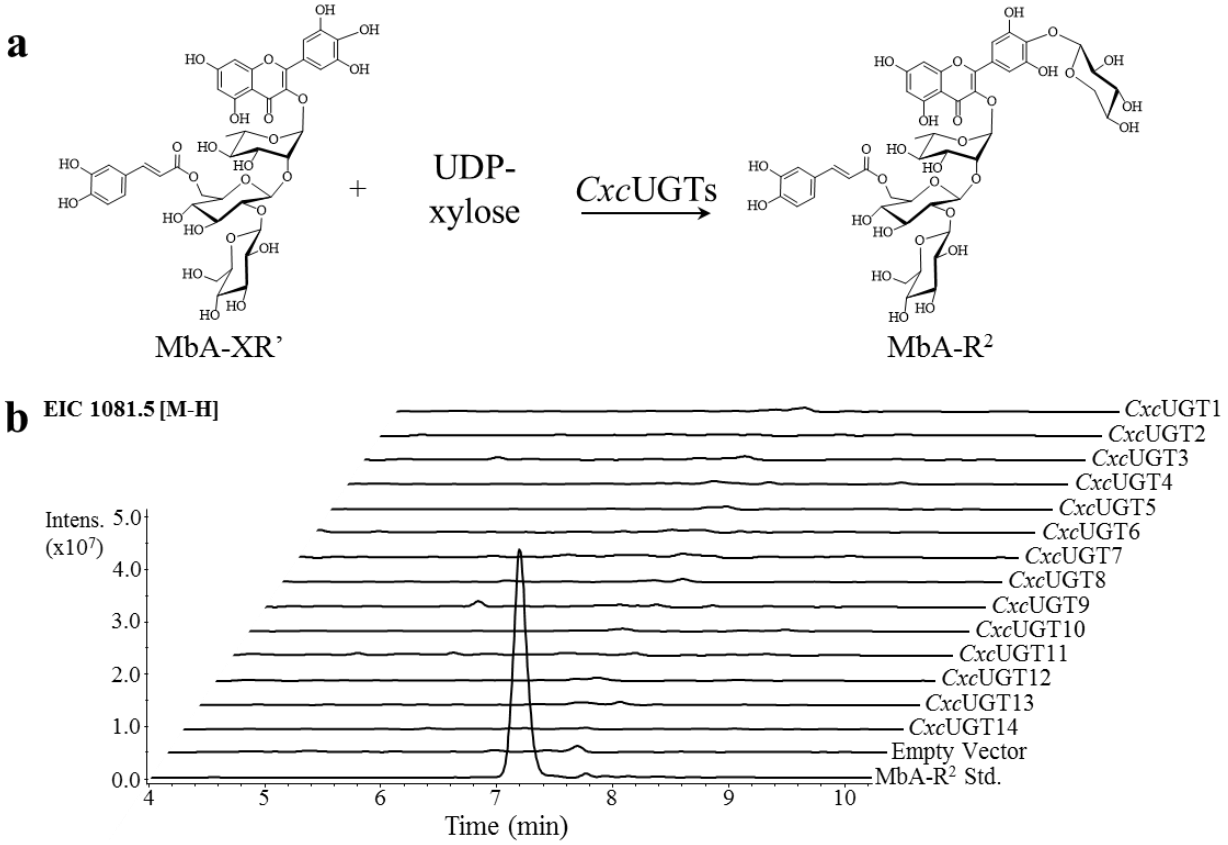


Figure S4.10: Representative regions of extracted ion LC-MS chromatograph for *C. x crocosmiiflora* GT1 UGT enzyme assays. (a) Substrate acceptor and UDP-sugar donor being tested for activity against *Cxc*UGTs. (b) Protein derived from *E. coli* expressing a control vector and the 14 GT1 UGTs of interest were incubated overnight with 1 mM UDP-glucuronic acid, 1 mM NAD^+, 1 μM purified *Cxc*UXS4, and 100 μM MbA-XR′ and assessed for their ability to produce MbA-R′ (theoretical molecular weight of 1082.93). Traces shown for each sample is the extracted ion chromatograph for m/z of 1081.5 [M-H](superscript -).

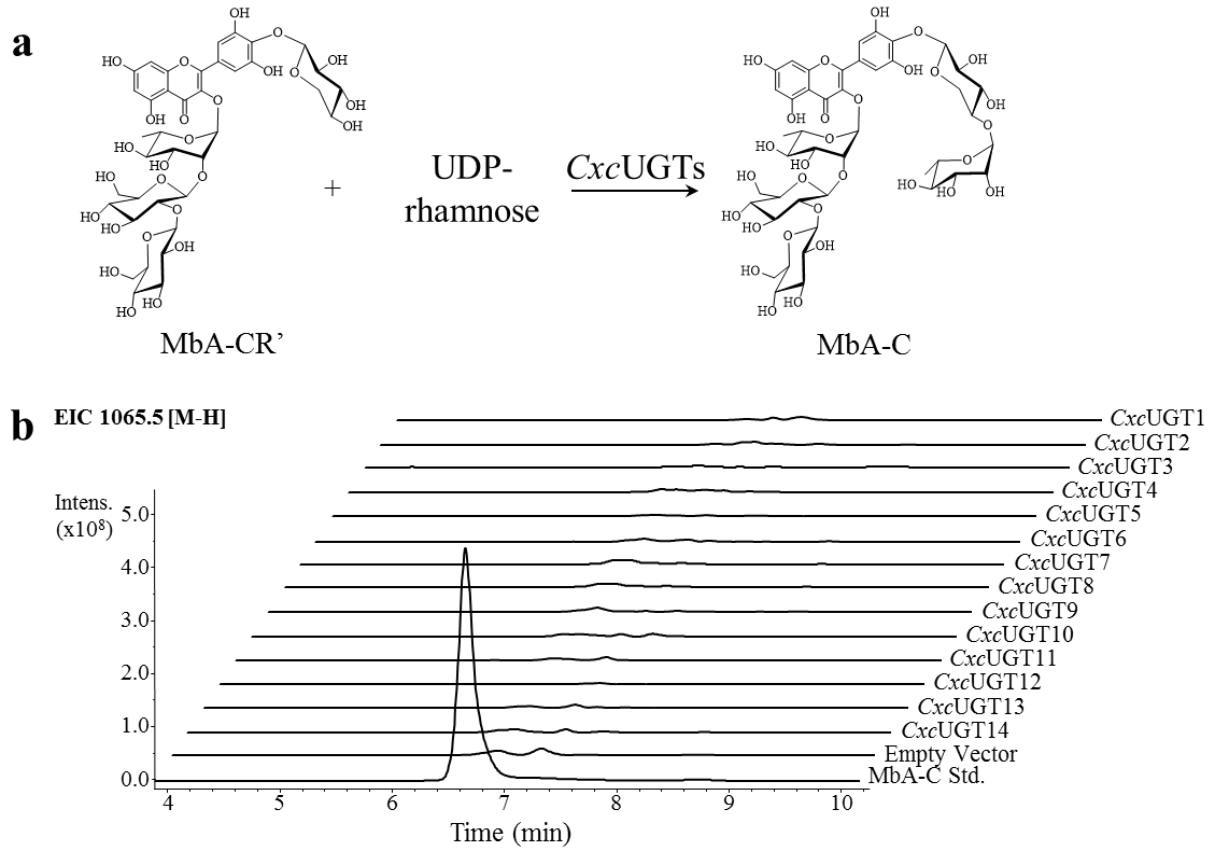


Figure S4.11: Representative regions of extracted ion LC-MS chromatograph for *C. x crocosmiiflora* GT1 UGT enzyme assays. (a) Substrate acceptor and UDP-sugar donor being tested for activity against *Cxc*UGTs. (b) Protein derived from *E. coli* expressing a control vector and the 14 GT1 UGTs of interest were incubated overnight with 1 mM UDP-glucose, 1 mM NAD^+, 1 mM NADPH, 6.5 μM purified *Cxc*RHM1, and 100 μM MbA-CR′ and assessed for their ability to produce MbA-C (theoretical molecular weight of 1066.90). Traces shown for each sample is the extracted ion chromatograph for m/z of 1065.5 [M-H](superscript -).

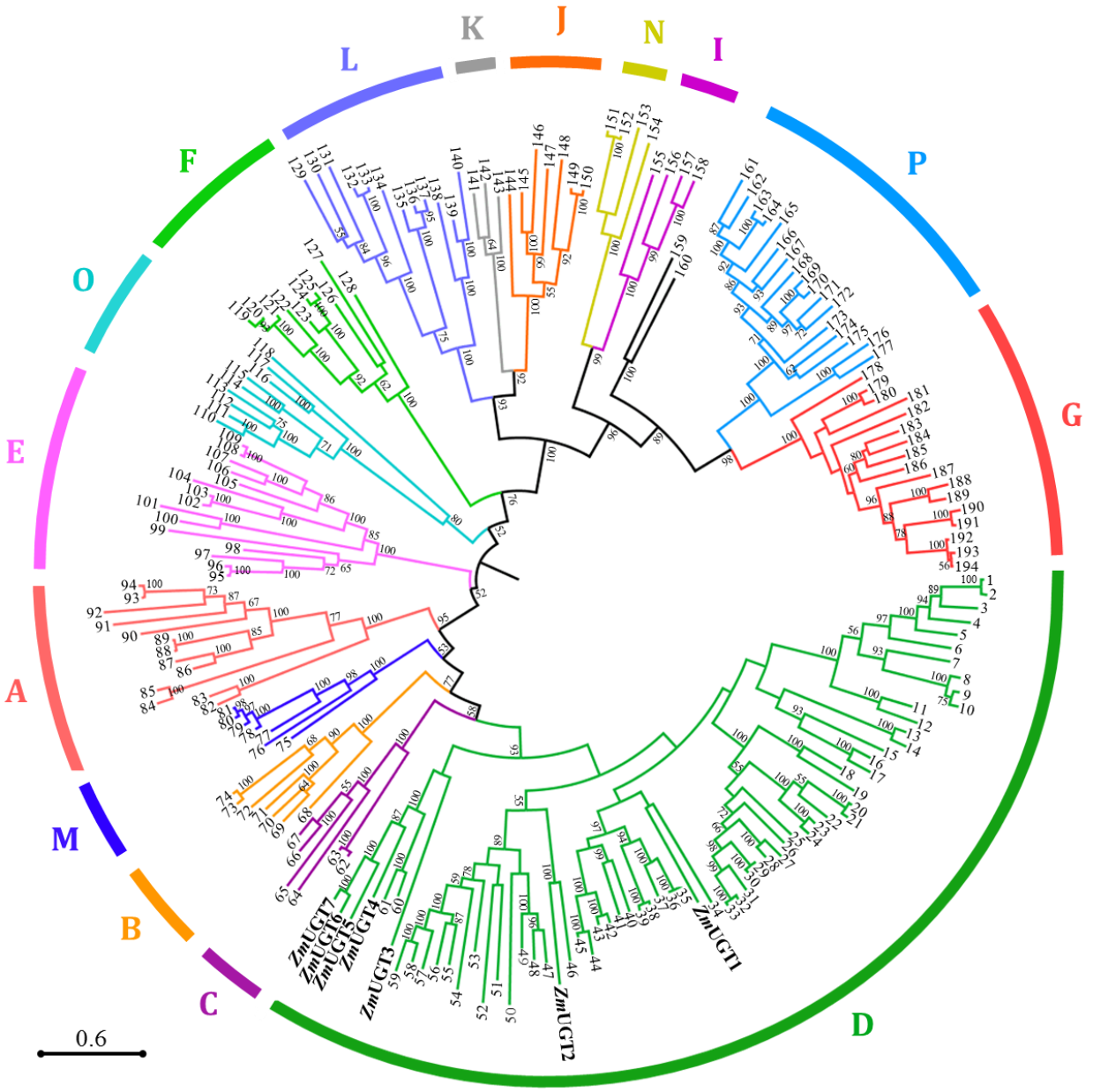


Figure S4.12: Phylogenetic analysis of GT1 UGTs from *C. x crocosmiiflora*. The maximum-likelihood tree was produced using the MEGA 7.0 program (bootstrap value set at 1,000) with sequences of 160 *C. x crocosmiiflora*, 15 *Arabidopsis thaliana*, 24 *Zea mays*, and 2 *Oryza sativa* GT1 UGT with amino acid sequences greater than 300. Bootstrap values over 50% are indicated above the nodes. The black bar represents 0.6 amino acid substitutions per site. The GT1 UGTs previously identified as group Q UGTs are indicated on tree. The remaining sequences are numbered 1 – 194 and correspond to the legend found in Table S4.5).

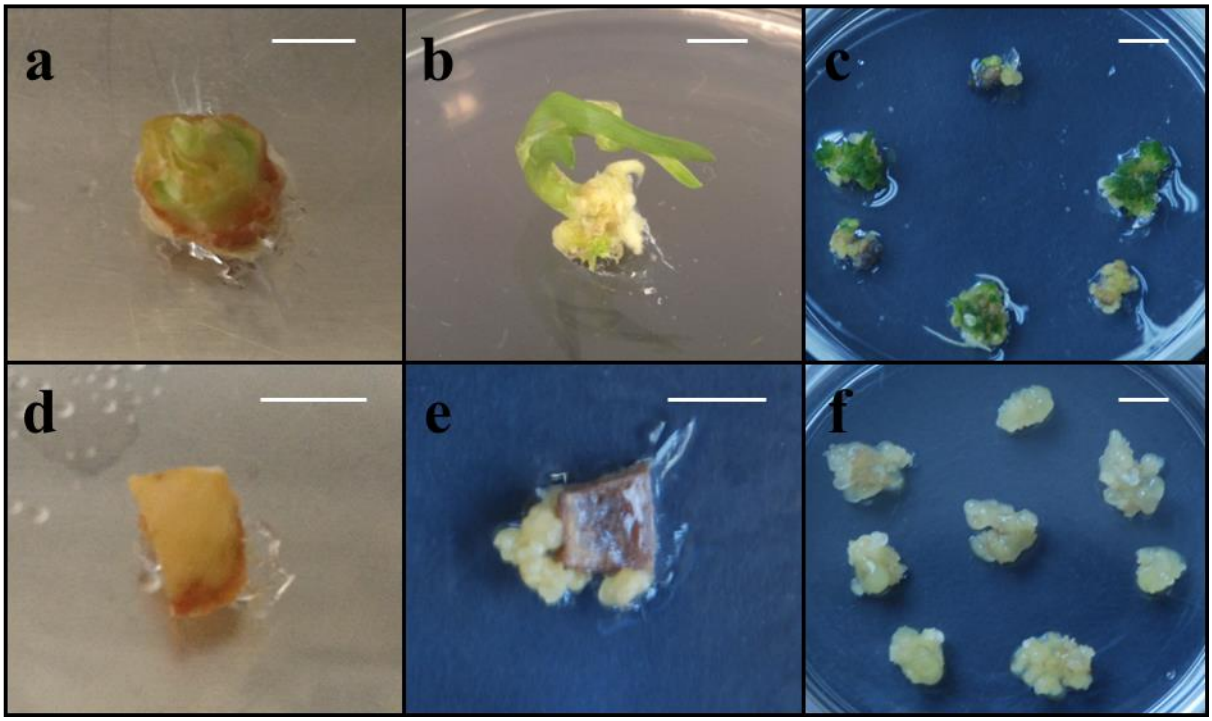


Figure S6.1: *C.* x *crocosmiiflora* callus cultures. (a-b) meristem tissue at week 0 and 2 respectively. (c) Meristem-derived callus at week 8. (d-e) corm tissue at week 0 and 5 respectively. (f) corm-derived callus at week 10. White bar represents 1 cm. Cultures were produced by first sterilizing plants in washes of 70% EtOH and 2% NaClO. Under sterile conditions the meristem and a ~1 cm^3 corm section were removed and plated on specific media. Meristem-based tissue was grown under 16 hours of light per day on minimal plant culture media with the hormones kinetin (1.0 mg/L) and 1-napthalenacetic acid (0.1 mg/L). Corm-based tissue was grown under 0 hours of light per day on minimal plant culture media with the hormones 2,4-dichlorophenoxyacetic acid (0.1 mg/L) and 6-benzylaminopurine (1.0 mg/L). Meristem- and corm-derived callus were first seen, respectively, at weeks 2 and 4 post explant. Subsequently, callus growth was removed and propagated in identical media.