12-2016

Identification and characterization of genes involved in metabolism of n5 monoene precursors to n5 anacardic acids in the trichomes of Pelargonium x hortorum.

Richa A. Singhal

Follow this and additional works at: https://ir.library.louisville.edu/etd

Part of the Biology Commons, Biotechnology Commons, Molecular Genetics Commons, and the Plant Biology Commons

IDENTIFICATION AND CHARACTERIZATION OF GENES INVOLVED IN
METABOLISM OF N5 MONOENE PRECURSORS TO N5 ANACARDIC ACIDS
IN THE TRICHOMES OF *PELARGONIUM × HORTORUM*

By

RICHA A. SINGHAL
M.Sc., University of Mumbai, India 2007
B. Sc., University of Mumbai, India 2005

A Dissertation
Submitted to the Faculty of the
College of Arts and Sciences of the University of Louisville in
Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy
In
Biology

Department of Biology
University of Louisville
Louisville, Kentucky

December 2016

IDENTIFICATION AND CHARACTERIZATION OF GENES INVOLVED IN
METABOLISM OF N5 MONOENE PRECURSORS TO N5 ANACARDIC ACIDS
IN THE TRICHOMES OF *PELARGONIUM × HORTORUM*

By
Richa A. Singhal
M.Sc., University of Mumbai, India 2007
B. Sc., University of Mumbai, India 2005

A Dissertation Approved on
November 21st, 2016
by the following Dissertation Committee:

_____
Dissertation Director
Dr. David Schultz

_____
Dr. Michael Perlin

_____
Dr. Joseph Steffen

_____
Dr. Kenneth Palmer

_____
Dr. Pradeep Kachroo

DEDICATION

This dissertation is dedicated to my family. My mother Mrs. Lalita

Singhal, my father Mr. Arunkumar Singhal and my sister Ruma

Singhal. I thank them for their unconditional love, support, motivation

and prayers. They are the source of my strength and have taught me

to never give up and always work hard to achieve my goals.

ACKNOWLEDGMENTS

I would like to express my heartfelt gratitude towards my mentor Dr. David J. Schultz for his gudiance and patience all these years. I thank him a lot for his insight and expertise which has greatly helped in developing my career as a future scientist and a critical thinker. His support and mentoring has made it possible for me to achieve my academic goals and complete my dissertation. I would also like to thank my committee members, Drs. Michael Perlin, Joseph Steffen, Kenneth Palmer and Pradeep Kachroo for their support, guidance and motivation. I thank all of them for always being available for my scientific queries and helping me develop my skills as a scientist. Their unified inputs along with their encouragement made this dissertation possible.

I would especially like to thank Sabine Waigel and Dr. Wolfgang Zacharias from the genomics core facility for allowing me to use their lab facilities and teaching me the nitty gritty of the experiments needed for completion of this degree. I also thank them for providing me with neccesary supplies needed to perform my experiments. Without their support, it wouldn't have been possible for me to complete the work needed for my dissertation.

like a mother and my guide ever since I landed in the United States of America. I want to thank the mentors for my life Dr. Daisaku Ikeda and Dr. Shashishekhar Narayan Sinha for guiding me towards being a good human being, good scientist and maintaing a never give up spirit in life. I would like to thank my previous emplyoers in India, Mr. Rakesh Shah and late Mr. Ravindra Prajapati who supported me for pursuing Ph.D in United States. I would like to thank my extended family members, my uncles Mr. Vasant Salian, Mr. Jayant Salian, Mr. Yatendra Goyal and my cousing sister Garima Goyal for their never ending support during the course of my study at the United States. I finally conclude the acknowledgments by thanking my parents Mrs. Lalita Singhal, Mr. Arunkumar Singhal and my sister Ruma Singhal who are my source of strength and happiness.

ABSTRACT

IDENTIFICATION AND CHARACTERIZATION OF GENES INVOLVED IN METABOLISM OF N5 MONOENE PRECURSORS TO N5 ANACARDIC ACIDS IN THE TRICHOMES OF *PELARGONIUM × HORTORUM*

Richa Singhal
November 21st 2016

Unusual monoenoic fatty acids (UMFA's) and specialized metabolites called anacardic acids (AnAc) are produced in glandular trichomes of *Pelargonium × hortorum* (geranium). The UMFA's, $16:1^{\Delta 11}$ and $18:1^{\Delta 13}$ are precursors for the synthesis of unsaturated AnAc 22:1n5 and 24:1n5 that contribute to pest resistance in geraniums. UMFAs and their derived AnAc metabolites not only provide a useful biological marker that differentiates the biosynthetic pathway for unusual mononenes from the common fatty acids (i.e. stearic, palmitic, oleic, linoleic and linolenic) but also have industrial, medical and agricultural applications. Fatty acid biosynthesis enzymes like acyl carrier proteins (ACPs); thioesterases (TEs) and β-ketoacyl-ACP synthases (KASs) are required for common fatty acid as well as the UMFA biosynthesis. Based on this, it is hypothesized that the specific isoforms of the fatty acid biosynthesis enzymes are highly expressed in trichomes and are involved specifically in metabolic channeling of UMFAs to anacardic acid synthesis within trichomes of geranium.

This hypothesis is based on the knowledge that there is a novel $\Delta 9$ myristoyl-ACP desaturase (MAD) that directs acyl-ACP into UMFA biosynthesis and the products of MAD are correlated with the dominant congeners of AnAc (22:1n5 and 24:1n5). Transcription of MAD as well as production of $16:1^{\Delta 11}$ and $18:1^{\Delta 13}$ and AnAc 22:1n5 and AnAc 24:1n5 has been found to be highly trichome specific.

This dissertation reports the identification of the complete nucleotide and protein sequences of genes for 2 ACPs, 3 FAT-As, 3 FAT-Bs, 4 KAS Is, 1 KAS II and 1 KAS III from a geranium EST database. Quantitative real-time PCR (qRT-PCR) was used to analyze tissue-specific expression patterns of the target genes, which indicated that ACP 1, ACP 2, KAS I-a/b, KAS Ic, FAT-A1, and FAT-A2 are highly expressed in trichomes. To further this research, a *de novo* RNA and micro-RNA transcriptome was generated from trichomes and bald pedicle of geranium, which helped in identification of several genetic components involved in UMFA synthesis. Bioinformatics analysis of RNA-transcriptome along with qRT-PCR and biochemical assays (HPLC and GC) were used to correlate the effect of temperature (18°C, 23°C and 28°C) on gene expression (ACPs, KASs, FAT-As) and changes in production of $16:1^{\Delta 11}$ and $18:1^{\Delta 13}$ UMFAs and 22:1n5 and 24:1n5 AnAc. Results of this work show that expression of ACP 1, ACP 2, KAS I-c, KAS I-a/b were correlated with changes in UMFAs and AnAc production with temperature, thus indicating their potential role in UMFA metabolism. We also determined that 23°C is an optimal temperature for production of UMFAs

and AnAcs as compared to 18°C and 28°C. To determine and verify the function of ACP 1 and ACP 2, we co-expressed these genes in conjunction with a Δ9 myristoyl-ACP (MAD) desaturase in both *E. coli* and tobacco. *E. coli* assay results show that expression of ACP 2 with MAD increased the production of UMFAs significantly, thus validating the novel role of ACP 2 in UMFA production. This work, in addition to the generation of a *de novo* transcriptome, provides a platform for further defining UMFA metabolism within trichomes of geranium.

TABLE OF CONTENTS

5. CO-EXPRESSION OF *PELARGONIUM × HORTORUM* ACYL CARRIER PROTEIN WITH Δ^9 14:0-ACP-DESATURASE.

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Summary

Fatty acids and other lipids obtained from plants are commonly utilized in agricultural, pharmacological, chemical and food industries. However, the utility of plant lipids as chemical feedstocks can be limited due to a lack of chemical diversity. Some plants are known to produce unusual fatty acids (UFAs) that have distinct chemical structures as compared to common fatty acids. UFAs are potential valuable feedstocks for the chemical industry and have a wide variety of applications as polymers, fuels and renewable sources of energy. Plants that accumulate UFAs are often not amenable to agriculture and little is known about the biosynthesis of these compounds, thus limiting the utilization of this renewable resource. Therefore, there is an enormous interest within plant biotechnology communities to identify genetic components (specialized fatty acid biosynthesis enzymes) involved in UFA synthesis, especially to generate transgenic crop plants engineered to accumulate high levels of specific UFAs. The focus of this dissertation is to study the biosynthesis of specific unusual monoenoic fatty acids (UMFAs) in trichomes (hair like structures) of *Pelargonium × hortorum* (garden geranium).

Introduction

Plant Unusual Fatty Acids

Plant lipids play a vital role in maintaining the structural organization of cells and organelles and are also required for storage of energy and signal transduction pathways[1]. Stearic, palmitic, oleic, linoleic and linolenic acid are five common fatty acids found in plant lipids. These common fatty acids have 16 to 18 carbons and contain up to three non-conjugated double bonds[2]. Fatty acids (FAs) that differ in structure from the common fatty acids can be referred to as "unusual fatty acids" (UFAs)[2]. There are about 200 such different fatty acid structures produced by plants. This wide variety of UFAs occurs in plants due to the difference in arrangement, placement and number of double and triple bonds in the structure along with different length of acyl chains which can vary from 8 to 24 carbons. Additionally the difference is also attributed to the type of novel functional group attached to the fatty acid chain such as hydroxy, epoxy, methoxy, acetylenic, furanoids, cyclopentenyl and cyclopropyl groups [3, 4] (Fig1.1). Plant species that contain UFAs tend to synthesize these structures in specific tissues (like seed endosperm and trichomes) or at specific stages of development. Examples of these plant species include, *Coriander sativum* (i.e. petroselinic acid $18:1^{\Delta6}$), *Cocos nucifera* (i.e. lauric acid 12:0)*, Ricinus communis* (i.e. ricinoleicic acid $18:1^{\Delta6}$), *Crepis palaestina* (i.e. vernoilc acid 18:1=O), *Pelargonium × hortorum* (i.e. $16:1^{\Delta11}$ and $18:1^{\Delta13}$) and *Brassica napus* (i.e. erucic acid $22:1^{\Delta13}$)[5].

1. Ricinoleic Acid (Hydroxy)
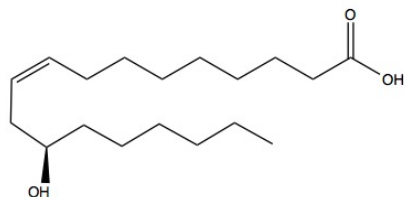
2. Vernolic Acid (epoxy)

3. Lauric Acid (Medium chain)
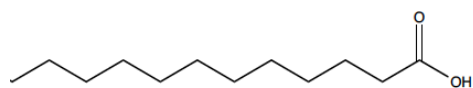
4. Erucic Acid (Very Long Chain)
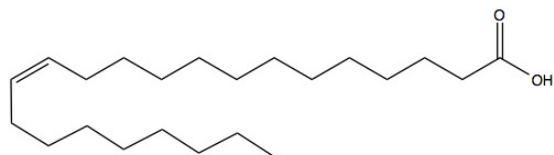
5. Petroselinic Acid (Specific double bond position)
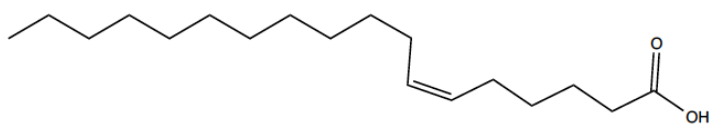
6. Chaulmoogric Acid (cyclopentenyl group)
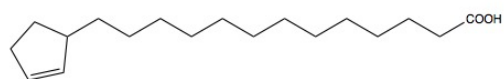
Figure 1.1. Plant unusual fatty acid structures

<u>Applications</u>

UFAs and industrial oils share similar chemical and physical properties and therefore have a wide variety of applications as polymers (such as paints, lubricants, nylons, plastics, and cosmetics), renewable sources of energy (fuels) and potential petroleum replacement products[5, 6]. Monounsaturated fatty acids like oleate ($18:1^{\Delta 9}$) and palmitoleate ($16:1^{\Delta 9}$) have optimal properties for biodiesel production because the single bond in these fatty acids helps in refining the cold-temperature flow properties of biodiesel and also provides better ignition quality and fuel stability[6]. Thus, UMFAs are ideal targets for biodiesel production. Most UFAs are excluded from membrane lipids of plant cells and are sequestered into oil bodies as part of storage lipids reaching up to 60% of the dry weight of seeds[3, 6]. Some UFAs can exist in cytosolic or epidermal membranes (specifically long chain UFAs)[7] that are used as substrates by plants for the production of plant specialized/secondary metabolites and other lipids that prevent water loss on plant surfaces by forming a hydrophobic barrier and additionally provide protection against plant pathogens and environmental stress[6]. Some of these metabolites (like anacardic acids) not only protect plants from biotic and abiotic challenges but also have clinical applications[8]. If the production of UFAs can be enhanced in seeds and other UFA specific tissues, then they can serve as a petroleum substitute and have various industrial, chemical and nutraceuticals appliacations [3, 8-10]. This fundamental and applied application of UFAs in plant biology has led to considerable interest in studying

4

the biosynthesis of UFAs and identifying specific enzymes involved in their production.

Primary metabolism leading to biosynthesis of Unusual Fatty Acids.

UFAs are hypothesized to be produced through a distinct "metabolic channel" of the primary metabolic pathways that results in production of common fatty acids found in all plants[11, 12]. In plants, fatty acid biosynthesis occurs in plastids through the activity of fatty acid synthase[13].

Fatty acid synthase

There are two basic types of fatty acid synthase (FAS) complexes consisting of enzymes needed for fatty acid synthesis (Figure 1.2). The type I system is found in mammals, lower eukaryotes and fungi. This system has 1 or 2 separate domains of a large multifunctional polypeptide and integrated enzymes for fatty acid synthesis[13]. The type II system is found in bacteria and plants. This system is a complex with monofunctional enzymes that are encoded by a separate genes[13].
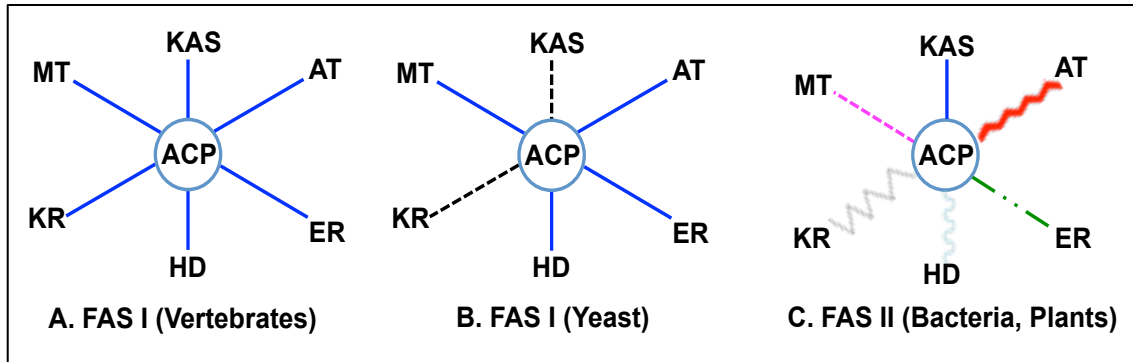
Figure 1.2. Fatty Acid Synthase complex. A) FAS I, found in vertebrates with one large polypeptide (blue color). B) FAS I, found in yeast with two separate polypeptides (blue and black colors). C) FAS II, found in plants and bacteria with seven separate polypeptides (various colors). Acyl carrier protein (ACP), β-ketoacyl- ACP synthase (KAS), acetyl transferase (AT), enoyl reductase (ER), dehydratase (HD), ketoreductase (KR) and malonyl transferase (MT).

Enzymes and key steps of primary metabolism

All of the enzymes of FAS II along with acetyl-CoA carboxylase (biotin-containing enzyme), carry out reactions of fatty acid synthesis (Figure 1.3). In plants acetyl-CoA carboxylase (ACCase) exists in two molecular forms, multiprotein complexes (in plastid) and multifunctional proteins (in cytosol) with expections in the Poaceae and Geraniaceae families [14,15]. The plastid localized AACase catalyzes the first committed and rate limiting step in FAS and produces malonyl-CoA that is subsequently converted to malonyl-ACP during *de novo* fatty acid synthesis while cytosolic ACCase provides malonyl-CoA for fatty acid elongation and other cellular processes. There is experimental evidence of posttranslational inhibition of acetyl–CoA carboxylase activity in the plastid that inhibits fatty acid synthesis and total oil accumulation. This post translation inihibition occurs due to inefficient utilization and excess of unusual fatty acids within the endoplasmic

reticulum[16]. To date, a novel isoform/ gene equivalent of ACCase, specific to unusual fatty acids has not been identified.

Amongst the various enzymes of fatty acid synthase complex, acyl carrier proteins (ACPs) are central to the process of fatty acid synthesis since they are conserved carriers of acyl intermediates throughout this process of synthesis[17]. In plants ACP is a small (9kD) separate polypeptide[18] that is synthesized in the cytoplasm and post-translationally imported into plastids[17]. Plants encode multiple ACP genes located in the nuclear genome that are expressed in specific tissues. Novel isoforms of ACPs have been identified for tissue specific production of UUFAs[9]. ACPs are indispensable metabolic cofactors and signaling molecules because they are not only involved in primary fatty acid synthesis but also in the synthesis of phospholipids, oligosaccharides, endotoxins and glycolipids. Given the known functions of ACPs discussed above, over-expression or altering the expression of ACP in plants can help in manipulating and altering the fatty acid profile, nutritional quality, and quantity of seed oils[17].

Other enzymes of the fatty acid synthase complex include, ACP transacylase that catalyzes the transfers of acetyl group or malonyl group from CoA to ACP to form acetyl ACP and malonyl-ACP[14] (Figure 1.3). The elongation reactions are catalyzed by three types of β-ketoacyl-ACP synthase (KAS)[13]. KAS III catalyzes the initial condensation reaction to form butyryl ACP; KAS I catalyzes the

condensation of C_4 to C_{14}-ACPs, whereas KAS II catalyzes the condensation of

C_{14} and C_{16}-ACPs with malonyl-ACP. KAS II is also responsible for determining

the $C_{16}:C_{18}$ ratio in the products (Figure 1.3). The following reduction-

dehydration-reduction cycles are catalyzed by fatty acid reductases and fatty acid

dehydrase (Figure 1.3). For the successive addition of two carbon units to the

growing fatty acid chain, β-ketoacyl-ACP reductase catalyzes the first reductive

step and forms β-hydroxyacyl-ACP by reducing the keto group to a hydroxyl

group (Figure 1.3). β-hydroxyacyl-ACP dehydrase then catalyzes the removal of

a water molecule, which forms an enoyl group[14] (Figure 1.3). Further, enoyl-ACP

reductase catalyzes the second reductive step and forms a four-carbon

compound called butyryl ACP (Figure 1.3). This elongation reaction is repeated

7 times to produce a palmitoyl ACP [14](16C).

Finally, two types of fatty acid thioesterases catalyze the chain termination

reaction by hydrolyzing the thioester linkage between ACP and the acyl group.

FAT A preferentially hydrolysis unsaturated acyl-ACPs and FAT B preferentially

hydrolysis saturated acyl-ACPs[14] (Figure 1.3). Usually, the *in vivo* products of

plant fatty acid synthases are palmitoyl-ACP and stearoyl-ACP in the ratio of

approximately (1:4)[14]. However, the type of products formed by plant fatty acid

synthase depends on the specificity of the thioesterase with the substrate. For

example, in some plants like California Bay and *Cuphea*, acyl-ACP thioesterase

causes premature termination of the chain-lengthening cycle, which leads to

formation of medium-chain products of fatty acids[19]. After chain termination, free

fatty acids are exported to the cytoplasm and are activated to CoA ester by Acyl

ACP synthetase (Figure 1.3). These products are then processed through either

the eukaryotic pathway or the prokaryotic pathway. The eukaryotic pathway is

located at the endoplasmic reticulum (ER) where acyl-CoAs are assembled for

glycerolipids synthesis and the prokaryotic pathway the acyl-ACPs are used for

lipid synthesis within the plastid (Figure 3.1)[14, 20]. Additionally for the formation

of unsaturated fatty acids, acyl-ACP desaturases catalyze the introduction of cis

double bonds into the acyl-ACPs at specific positions.[19,21] (Figure 1.3 and 1.4).

Mostly Δ9 stearoyl ACP desaturase acts on a majority of C18:0 in plants and

additional desaturases are present in a few species that produce unusual C18:1

congeners. Isoforms of desaturases vary in specificities for acyl chain length and

position of double bond insertion. The Δ9 desaturases are responsible for

synthesis of oleic acid, linoleic acid and α-linolenic acids that are major plant fatty

acids and they represent 85% of total membrane acids [14, 21].
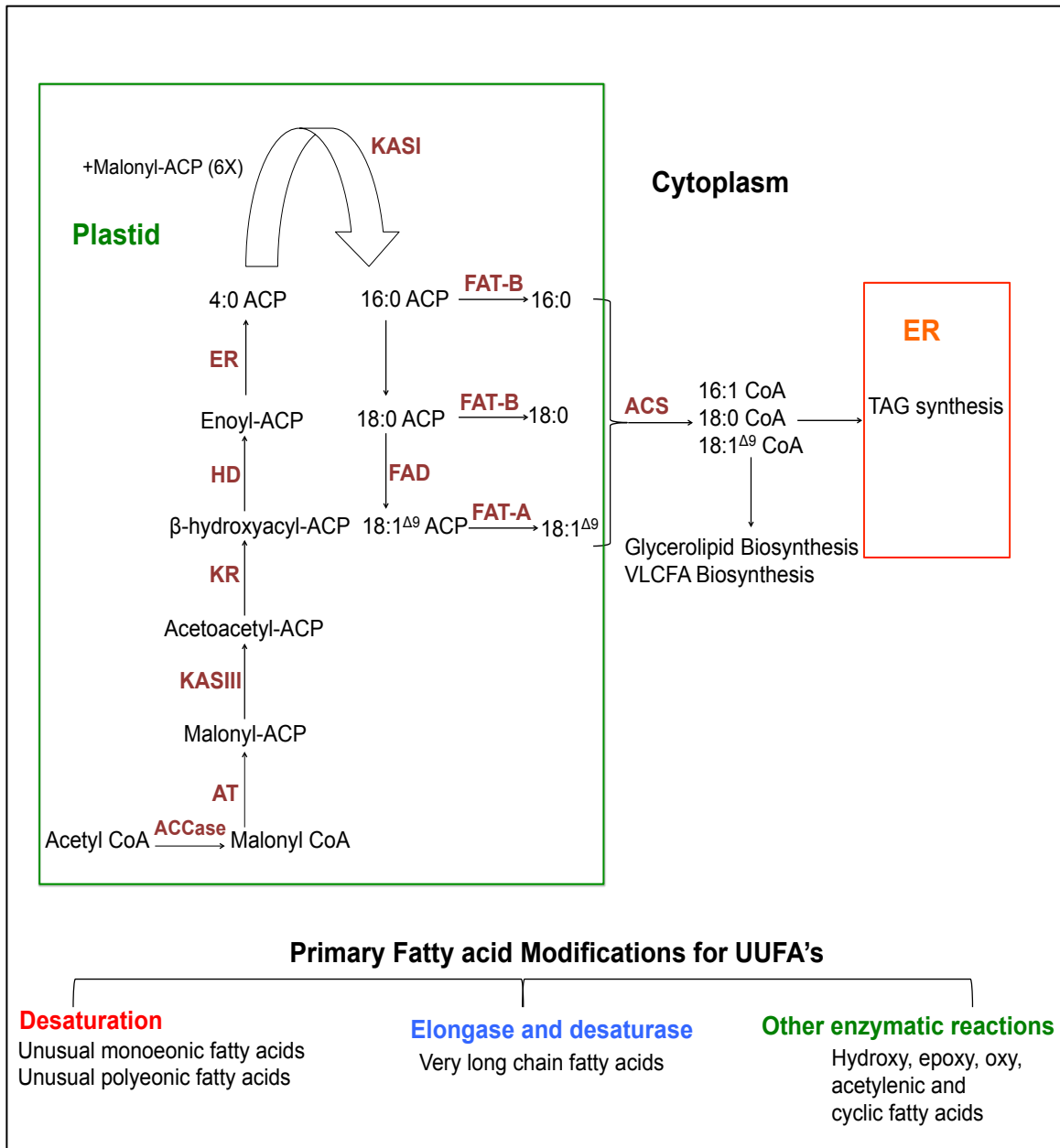
Figure 1.3. Primary Fatty acid biosynthesis leading to production of unusual fatty acids. ACCase (acetyl-CoA carboxylase), MT (malonyl transferase), KAS (β-ketoacyl-ACP synthase), AT (acetyl transferase), ER (enoyl reductase), HD (dehydratase), FAT (fatty acid acyl-ACP thioesterase), ACS (acetyl-CoA synthase), FAD (fatty acid desaturase) and ER (endoplasmic reticulum).

Enzymatic Modifications of UFAs production

It is the process of gene duplication, alternative splicing and specialization of

housekeeping genes has given rise to various isoforms/paralogs of enzymes

involved in fatty acid synthesis (enzymes discussed above) [5]. Since these enzymes are variants of fatty acid synthesis enzymes, they have closely related catalytic functions and altered substrate specificity, which enables them to carry out an additional or alternative enzymatic reaction to produce UFA's (Figure 1.4)[3]. Novel fatty acid biosynthesis enzyme isoforms have been discovered that are tissue-specific and involve synthesis and accumulation of UFAs[11, 22, 23]. Examples of these genes include the discovery of novel acyl carrier proteins, thioesterases, desaturases, acyltransferases and novel β-ketoacyl-ACP synthases[22, 23]. Very long chain fatty acids or glycerolipids are produced via substrate channeling of acyl-ACPs in the cytosol[3, 11] whereas unusual monoenes that are produced in the plastid that can either get incoporated into triacylglycerol or undergo enzymatic modifications that lead to production of specialized metabolites in specific tissues [3, 11]. Most of the UFAs become part of storage lipids whereas very few readily enter plastidial or cytosolic membrane lipids[3]. However the exact sorting mechanism of UFA's is still unknown and may be different for different plant species[3].

Types of Unusual Fatty Acids

Specific type of UFAsare produced by modified fatty acid reactions like desaturation, methylation, acetylation, hydroxylation, expoxidation and elongation (Figure 1.4)
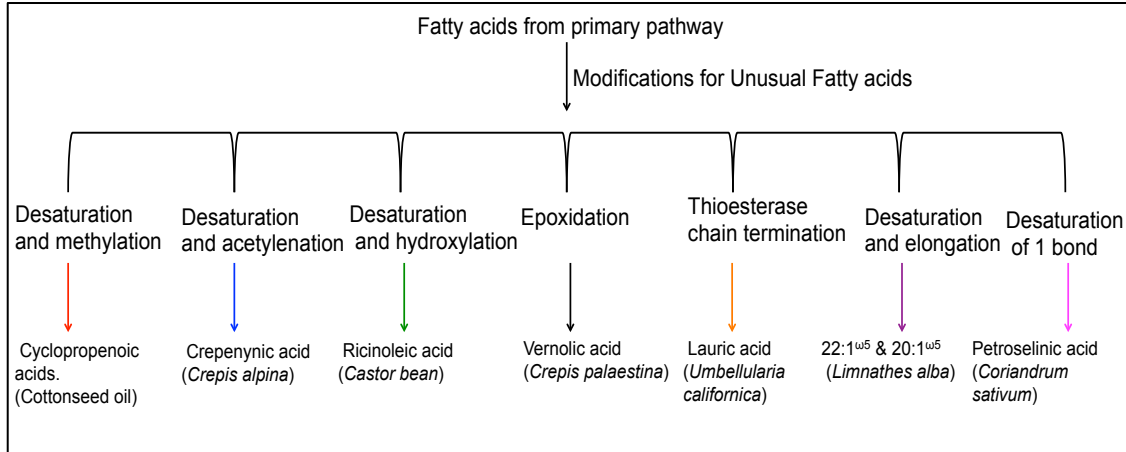
Figure 1.4. Production of specific unusual fatty acids through enzymatic modifications

Medium Chain fatty acid (MCFA)

MCFA chain length ranges from 8 to 14 carbons. Synthesis of MCFA's requires a dedicated acyl-ACP thioesterase that cleaves the acyl chain prematurely to produce C8-C14 fatty acids (Figure 1.4)[24]. As an example, seeds of *Umbellularia californica* produce a thiosesterase that is specific for lauroyl-ACP and it terminates the chain length at 12C instead of 16C or 18C as seen in most plants[25]. A novel thioesterase from *Umbellularia californica* with substrate preference for C12:0 was expressed in rapeseed, this led to alteration of enlogation pathway such that up to 40% of short chain lauric acid was produced in the seed oils [5, 26, 27]. The amount of laurate in the transgenic rapeseed plant was lower as compared to the amounts found in coconut and *C. lanceolata* because of restricted accumulation of laurate at sn-1 and sn-3 positions of triacylglycerol. This was due to lack of a specific lysophosphatidic acid acyl-glycerol-3-phosphate acyl transferase (LPAAT) enzyme, which utilizes medium chain fatty acid as a substrate and causes accumulation of laurate at sn-1,2 and

3 positions of triacylglycerol.   Thus, when a specific LPAAT from the coconut

and a novel thioesterase from *Umbellularia californica* was expressed in

rapeseed, the production of MCFA increased by an additional 40%. Apart from

thiosestrases, a novel KAS I has been identified in coconut, which is specific for

MCFAs[28]. These findings suggest that more than one enzyme or group of

enzymes are needed for the production of UFAs[5, 26].

Very Long Chain Fatty acid (VLCFA)

Successive rounds of elongation occur extrapastidially for production of

VLCFA's. The chain length for VLCFAs is greater than 18 carbons (Figure 1.4).

These VLCFAs can be both mono-unsaturated (e.g eruic acid) and

polyunsaturated (e.g., arachidic acid). Each elongation reaction cycle is

controlled by an acyl-CoA enlongase complex involving four reactions -

condensation of two carbons of malonyl-CoA, followed by reduction, dehydration

and a second reduction reaction[5]. Novel fatty acid elongase enzymes involved in

production of VLCFA have been identified in seeds of *Arabidopsis* and *Brassica*

napus[5 29].

Hydroxyl (OH), Epoxy (=O), Acetylenic (−C≡C−) Fatty Acids

Synthesis of hydroxyl, epoxy, methoxy and acetylenic fatty acids takes place on

the endoplasmic reticulum as opposed to plastids for most UFAs. It requires

specialized enzymes (desaturase, epoxgenase, acetylenase, hydroxylase) that

are structurally related to the extraplastidial Δ12 desaturase (FAD2) family (Figure 1.4)[30].

Unusual Monoeonic fatty acid (UMFA)

UMFAs have one double at specific positions in the fatty acid structure that are not commonly found in all plants. They have a wide variety of functions in industry and medicine (Table 1.1). Five distinct acyl-acyl carrier protein desaturase (AAD) genes have been isolated, each capable of producing distinct unusual monoenoic fatty acids. Additionally, four proteins - ferredoxin (Fd), acyl carrier protein, 3-ketoacyl-ACP synthase and thioesterase, have specific roles in the production of UMFA in their native plants and thus play an important role in understanding production of UMFAs[9, 31-33]. For example, in *Coriandrum sativum*, a specialized soluble fatty acid desaturase that is closely related to Δ9 desaturase, introduces a double bond between 4C and 5C of a 16 carbon acyl chain, which is then extended by a novel condensing KAS enzyme and esterified by a novel FAT- A enzyme to produce UMFA petroselinic acid $(18:1^{\triangle 6})$[25, 33-35].

Table 1.1 Details of Unusual Monoeonic fatty acids

Unusual Monoeonic Fatty Acids	Plant	Novel enzyme	Use/Application
Erucic ($22:1^{\omega 9}$)	*Arabidopsis thaliana*	β-ketoacyl-ACP synthase	Oil paints, precursor to biodiesel fuel.
Ricinoleic (18:1-OH)	*Ricinus Communis*	Diiron Hydroxylase	Medicine (anti-inflammatory, analgesic), biodiesel and nylon precursor.
Vernolic (18:1=O)	*Crepis palaestina*	FAD2-related expoxygenase	Paints, adhesives.
Petroselinic ($18:1^{\Delta 6}$)	*Coriandrum sativum*	Δ6 desaturase, β-ketoacyl-ACP synthase, fatty acid acyl-ACP thioesterase	Surfactants, cosmetics, food industry.
$14:1^{\Delta 9}$, $16:1^{\Delta 11}$ and $18:1^{\Delta 13}$	*Pelargonium × hortorum*	Δ9 desaturase	Substrates for pest resistant anacardic acid production.
Palmitoleic acid ($16:1^{\Delta 9}$)	*Hippophae rhamnoides*	Δ9 desaturase	Detergents, medicine.

Includes plant source, novel enzymes and applications[3, 4, 11, 25, 36, 37]

Derivatives of Unusual Fatty Acids

Derivatives of UFAs include a class of secondary/specialized metabolites containing phenolic lipids (polyketides), essential omega fatty acids and various other aliphatic compounds like cuticular lipids[6, 38, 39]. Phenolic compounds (polyketides) are of interest since they have a wide variety of biological activities like antimicrobial, anti-cholesterol and anticancer and are also used in agriculture and animal husbandry[38]. Enzymes of the type III polyketide synthases (PKS) superfamily (including well studied chalcone/stilbene synthases involved in flavonoid biosynthesis) are involved in the synthesis of these phenolic lipids[36, 38]. Type III PKS enzymes are found in plants and bacteria and recent genome and biochemical studies show their presence in filamentous fungi[38]. Products of type III PKSs include chalcones, pyrones, stilbenes, acridones and resorcinolic lipids like Cardol, Cardanol, anacardic acid, oliveotolic acid and urishiol[36, 38]. The

common biosynthetic theme of these enzymes includes the utilization of fatty

acid CoA esters as substrates, followed by condensation reactions to generate a

polyketide intermediate. This intermediate then undergoes subsequent

cyclization followed by aromatization to produce phenolic lipids, (an example of

this is illustrated in Figure 1.5)[36, 38]. It has been observed that distinct PKS's exist

and contribute to the production of specific lipids[38]. For example, distinct

alkyresorcinol synthase uses unusual $\Delta^{9,12,15}$- 16:3-CoA precursors to produce a

pentadecatrienylresorcinol intermediate, which then undergoes additional

modifications to produce sorgoleone[40]. Similarly, type III PKS with specific

substrate specificty can lead to formation of related compounds like, cardol,

cardanol, urishiol anacardic acids and alkyl resorcinols (such as olivetol)[36].

Additionally, stilbene synthase, which is a divergent chalcone synthase type III

PKS, forms a tetraketide intermediate, like chalcone synthase and then

subsequently undergoes C2/C7 cyclization instead of C1/C6 cyclization to form

resveratrol[41]. Furthermore, in trichomes of *Pelargonium × hortorum* unusual

16:1$^{\Delta 11}$-CoA and 18:1$^{\Delta 13}$-CoA are substrates for production of anacardic acids.

An enzyme keto-acyl CoA synthase 2 has been identified which is structurally

similar to Type III PKS's and is potentially involved in a final condensation

reaction that leads to the production of anacardic acid (unpublished), (Figure 1.5)
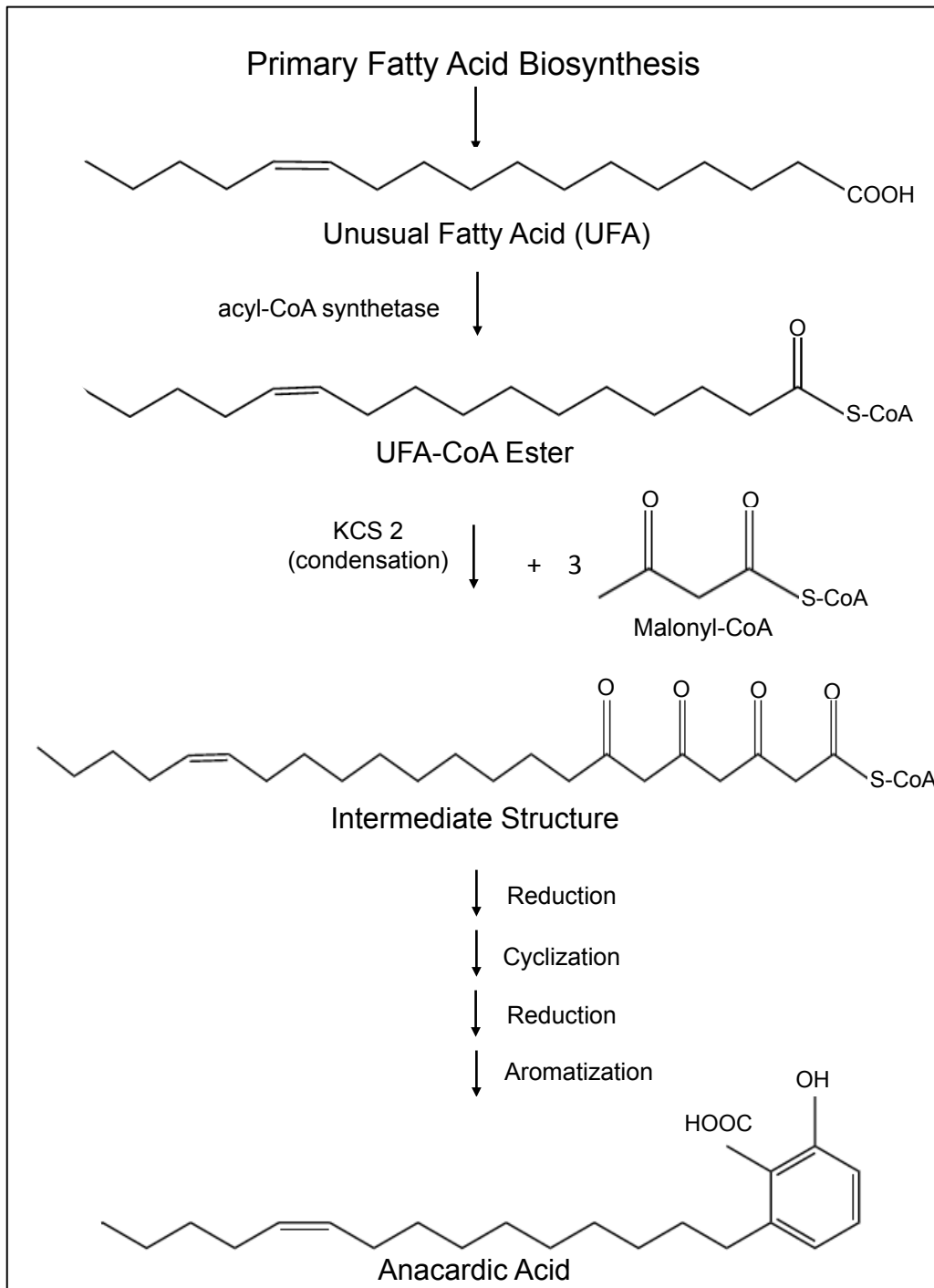
Figure 1.5. Anacardic acid - a derivative of unusual fatty acid.

Production of UFAs in transgenic plants

Researchers are trying to cultivate plants that accumulate unusual fatty acids since enormous interest in the diversity, functions and commercial application of UFAs has gained momentum in recent years [4]. Such plants can reduce the cost of production of UFAs, which can serve as petroleum substitutes along with various industrial and medicinal products[4]. Unfortunately plants that produce UFAs are not suitable for agriculture and little is known about the biosynthesis of UFAs, which limits their utility as a renewable resource. Thus, plant biotechnology research focuses on identifying genetic components (specialized fatty acid biosynthesis enzymes, cofactors or enzyme complexes) for generation of transgenic plants that accumulate high levels of specific UUFAs. Furthermore it is also important to identify the site, time and level of expression of UUFA related genes along with an investigation of the substrate availability, level and enhancement[4].

Researchers in this area have observed that when specific fatty acid genes from plants that naturally accumulate UFAs are introduced into transgenic plants, the yield of UFAs obtained in the seed oil is very low[4]. For example, expression of fatty acid hydroxylase 12 (obtained from castor) under a strong seed-specific promoter leads to very low accumulation of ricinoleic acid in transgenic plants (<1% transgenic tobacco and <17% in *Arabidopsis* as compared to 90% in storage glycerides of the castor plant)[42, 43]. The exact reason for minimal accumulation of UFAs in transgenic plants is still being evaluated and there are

few possible explanations. It is possible that the expression of a single gene to obtain a higher yield may not be sufficient in some cases and it may require an assembly of compatible multicomponent enzymes that are functional in a transgenic system or have additional secondary activities[4, 11]. This hypothesis was supported by experimental evidence where the expression of acyl ACP desaturase under a strong seed-specific promoter led to the accumulation of 1-15% in seeds of *Arabidopsis* as compared to 80% accumulation in the seed of the parent plant (*Thunbergia alata, Coriandrum sativum*)[11]. Experiments were done to confirm that there was no degradation of UFAs during seed development, no enzyme barriers were identified that could lead to poor incorporation of UFA into triacylglycerol. Even the competition with endogenous protein for substrates and cofactors was not the reason for limited accumulation of UMFAs. It was further evaluated that even though the expression of acyl ACP desaturase was higher than the endogenous Δ9-18:0 ACP desaturase, its activity was 10-100 fold lower than the Δ9-18:0 ACP desaturase, indicating that the expression level of desaturase alone is not the limiting factor for accumulation of UMFAs[11].

Another possible explanation is beta-oxidation of UFAs to acetyl-CoA that causes poor channeling of UFAs leading to inefficient incoporation into triacylglycerols, limited availability of substrate or additional activity/ interference from enzymes present in the host [5, 44]. This was demonstrated in an experiment involving production of VLCFA erucic acid in rapeseed. The erucic acid specific

LPAAT gene from *Limnanthes* (which incorporates VLCFAs at all three sn positions of TAG) was expressed in rapeseed[5]. There wasn't a significant increase in total erucic acid content (40%) in transgenic lines, suggesting that its production was limited by the activity of enlongase[5].   To test this possibility, the LPAAT gene from *Limnanthes* along with the fatty acid elongase gene from *Brassica napus* was co-expressed in rapeseed and the resulting transgenic lines expressing these genes accumulated slightly more erucic acid (60%) as compared to lines expressing only LPAAT gene (40%). Limited accumulation of erucic acid in rapeseed can be due to limited availability of the substrate or beta oxidation unusual VLCFA. Evaluating the mechanism of efficient incorporation and stabilization of UFAsinto triacylglycerols can help in accumulating high levels of UFAsin the seeds of transgenic plants.

Geranium as a model system for studying UMFA synthesis

Pelargonium × hortorum (geranium) consists of glandular trichomes that can be used to study the synthesis of UMFAs. Trichomes are small hair-like structures present on the surface of leaves, stems and pedicles of plants[45]. Glandular trichomes are metabolically specialized cells that produce a wide array of specialized metabolites[45]. UMFAs ($16:1^{\Delta11}$ and $18:1^{\Delta13}$) are present in glandular trichomes of geranium and are the direct precursors for production of $22:1^{\omega5}$ and $24:1^{\omega5}$ anacardic acids[36]. These unsaturated anacardic acids confer pest resistance to plants and are found to inhibit the growth of breast cancer cells [36, 46]. They also have antibacterial, antifungal and molluscicidal properties[45].

Glandular trichomes of geranium are highly specialized for production of specific UMFAs - $16:1^{\Delta11}$ and $18:1^{\Delta13}$, which as substrates account for more than 80% of the anacardic acid profile[45]. Furthermore, glandular trichomes are readily isolated as pure cell preparations, thus making them an ideal model tissue for exploration of the underlying genetics and biochemistry of UMFA synthesis. A novel Δ9 14:0-acyl carrier protein (ACP) desaturase (MAD) has already been identified within trichomes of *Pelargonium × hortorum*. The MAD gene is responsible for producing myristoleic acid ($14:1^{\Delta9}$) that is elongated into $16:1^{\Delta11}$ and $18:1^{\Delta13}$ UMFAs[47]. To facilitate a more complete understanding of this system, the goal of this project is to identify and characterize various genetic componets including distinct isoforms of fatty acid biosynthesis enzymes like ACPs, KASs and FATs that are highly expressed in the trichomes and potentially involved specifically in metabolic channeling of UMFAs to anacardic acid synthesis within trichomes of geranium.

CHAPTER 2

IDENTIFICATION AND EXPRESSION ANALYSIS OF FATTY ACID
BIOSYNTHESIS GENES FROM *PELARGONIUM × HORTORUM*.

Summary

Unusual monoenoic fatty acids (UMFA's) and derived specialized metabolites

called anacardic acids (AnAc) are produced in glandular trichomes of

Pelargonium × hortorum (geranium). The UMFA's, $16:1^{\Delta 11}$ and $18:1^{\Delta 13}$ are

precursors for the synthesis of AnAc 22:1n5 and 24:1n5 that confer pest

resistance in geranium. UMFAs and their AnAc metabolites provide a useful

biological marker to differentiate the biosynthetic pathway for unusual

mononenes from the common fatty acids and they have industrial, medical and

agricultural applications. Acyl carrier proteins (ACPs), β-ketoacyl-ACP synthases

(KASs) and acyl-ACP thioesterases (FATs) are required for common fatty acids

as well as the UMFA biosynthesis. Thus, we hypothesized that specific

isoforms/paralogs of these fatty acid biosynthesis enzymes will be highly

expressed in trichomes and will be involved specifically in synthesis of UMFAs by

an alternate metabolic channeling. Subsequently, complete sequences

(nucleotide and amino acids) of 2 ACPs, 3 FAT-As, 3 FAT-Bs, 4 KAS Is, 1 KAS II

and 1 KAS III were identified. Phylogenetic analysis of these target sequences

indicated that ACP 2, KAS Ic, KAS Ia, KAS Ib and FAT-A1 protein sequences

were in the same clade as that of *Coriandrum sativum* ACP, KAS and FAT-A

enzymes which are associated with the biosynthesis of petroselinic acid ($18:1^{\Delta 6}$,

a specific UMFA). Expression analysis of target sequences indicated that ACP

1, ACP 2, KAS Ia/b, KAS Ic, FAT-A1, and FAT-A2 are highly expressed in

trichomes compared to the bald pedicle, suggesting their potential role in

trichome UMFA metabolism.

Introduction

Unusual monoenoic fatty acids (UMFAs) are characterized by presence of a

single double bond at specific positions in the fatty acid structure that are not

commonly found in all plants. UMFAs and industrial oils share similar chemical

and physical properties due to which they have a wide variety of applications as

polymers, fuels, nutraceuticals, medicine and renewable sources of energy[4, 48].

Production of UMFAs is typically restricted to specific tissues (seed endosperm,

trichomes) in various plants like *Thunbergia alata* ($16:1^{\Delta 6}$)*, Coriandrum sativum*

($18:1^{\Delta 6}$)· *Asclepia syriaca* ($16:1^{\Delta 9}$,$18:1^{\Delta 11}$) *, Arabidopsis thaliana* ($22:1^{\omega 5}$)*, Ricinus

communis (18:1=O)*, Hedrea helix* ($16:1^{\Delta 6}$) and *Hippophae rhamnoides* ($22:1^{\omega 5}$)[3,
9, 22, 46, 49-55]. Within these plants that produce UMFAs, novel isoforms of a limited

number of fatty acid biosynthesis (FAS) enzymes have been identified that are

involved in UMFA synthesis [3, 22, 33]. Examples of these FAS enzymes include

ACPs that are conserved carrier of acyl intermediates during fatty acid

synthesis[1, 17], KASs that are involved in condensation reactions[1] and FAT-As

that hydrolyze acyl-ACP products of fatty acids[1].

In *Pelargonium × hortorum* L.H. Bailey (zonal geraniums, Figure 2.1), unusual monoenoic fatty acids 16:1$^{\Delta 11}$ and 18:1$^{\Delta 13}$ are found in specialized structures called glandular trichomes[56]. Glandular trichomes are small hair like structures found on the surface of stems and leaves of the plant. They have a base, stalk and globular head (gland) where the metabolites are secreted (Figure 2.2)[57]. Tall glandular trichomes of geranium produce a specialized metabolite called anacardic acid which occurs in both saturated (22:0, 24:0) and unsaturated forms (22:1n5, 24:1n5)[45]. The unsaturated anacardic acids 22:1n5 and 24:1n5 are unique to pest resistant geraniums where they form a sticky viscous trap which provides the plant with a primary defense against pests like spider mites and aphids [45, 58]. The unsaturated AnAc exudate has proven efficacious against spider mite, larval whiteflies, colorado potato beetle and the tomato hornworm, thus making it a potential candidate for use in controlling agriculture pests[36, 45]. AnAc also has antibacterial, antifungal and molluscicidal properties[59-65]. Furthermore, there is experimental evidence of antitumor activities on cancer cell lines (liver, breast, bladder, cervical and pituitary)[36, 66-74]. The UMFA's 16:1$^{\Delta 11}$ and 18:1$^{\Delta 13}$ are biosynthetic precursors to unsaturated anacardic acids 22:1n5 and 24:1n5 (Figure 2.3)[56]. Therefore, studying synthesis of UMFAs in trichomes of geranium not only aids in defining UMFA metabolism but also helps in understanding production of AnAc.

Figure 2.1. *Pelargonium × hortorum* L.H. Bailey (zonal geranium). Plants grown in environmental chambers, maintained at 20°C with 16-hour photoperiod.

To further study the biochemical pathway and identify all the genetic components involved in production of UMFAs and AnAc, a trichome-specific expressed sequence tags (EST) database was constructed for *Pelargonium × hortorum* (trichomes from the pedicles as a source material). This EST database was generated from 3781 random cDNA (complementary DNA) clones and categorized metabolically based on homology against Gen Bank. For all the target enzyme steps, more than one gene sequence is available from the EST database (Figure 2.4). Since the FAS enzymes are common to both fatty acid synthesis and UMFA synthesis, the distinct protein types encoded by each EST of a target gene potentially represents separate metabolic channels for fatty

acids, one for glycerol lipid and cuticular lipids biosynthesis and one for UMFAs

synthesis leading to biosynthesis of anacardic acids (Figure 2.4).
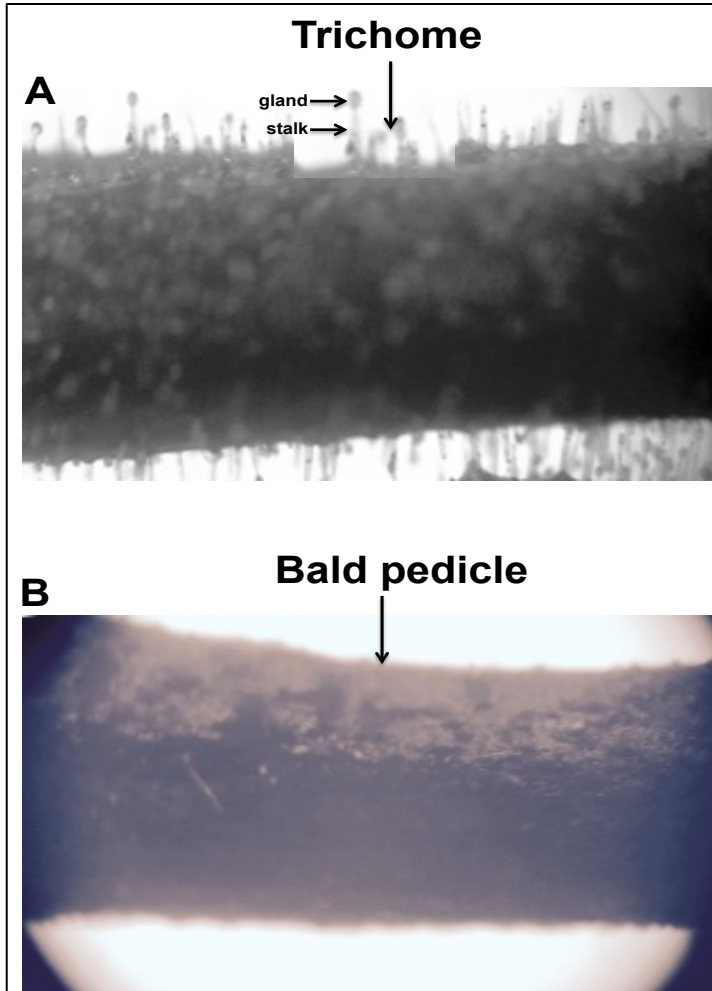
Figure 2.2. Images of trichomes and bald pedicle from geranium. A) Image of pedicle with trichomes (before shearing the trichomes. B) Image of bald pedicle (after shearing the trichomes).

For this study three key fatty acid biosynthesis enzymes, ACPs, KASs and FATs

were evaluated. To elucidate novel functionality of ACPs, KASs and FATs we

propose to differentiate target enzyme isoforms/paralogs/types based of tissue

specificity. Since the UmFAs are only produced in glandular trichomes, we

expect expression of isoforms involved in this process to also be expressed in the same tissue specific pattern. This approach is supported by identification of a novel Δ9 14:0- myristoyl ACP desaturase (MAD), which directs acyl-ACP into UMFA biosynthesis and has been found to be highly trichome specific (Figure 2.4)[75]. In this study complete sequences of FAS genes have been identified using bioinformatics and molecular biology tools, followed by their phylogenetic analysis and expression analysis using quantitative real time PCR (qRT-PCR).
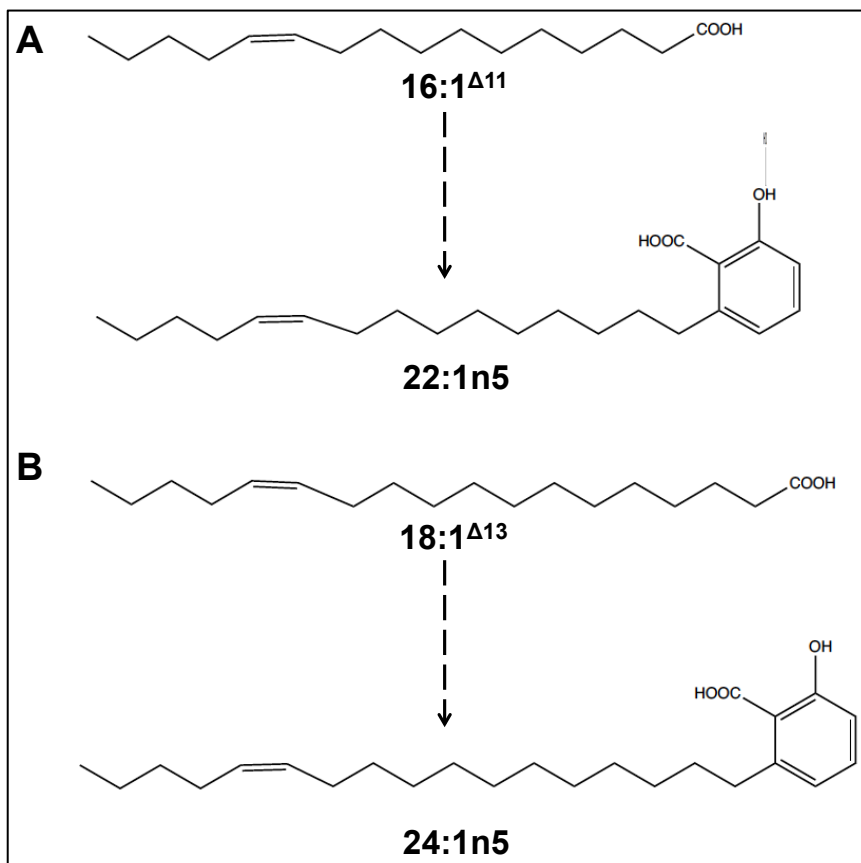
Figure 2.3. Structures of unusual monoenoic fatty acids and derived anacardic acids. Each UMFA, 16:1$^{\Delta 11}$ (A) and 18:1$^{\Delta 13}$ (B) is proposed to be elongated by 6 carbons and cyclized to form the corresponding AnAc, 22:1n5 or 24:1n5, respectively.

Primary Fatty Acid Biosynthesis

Acetyl-CoA

ACCase

Malonyl-CoA

MT

Malonyl-ACP

KAS III

4:0-ACP
(Metabolic Split)

Malonyl-ACP (5X)   Malonyl-ACP (6X)

KAS I

14:0-ACP

*Δ⁹ MAD → $*\Delta^9$ MAD

$14:1^{\Delta 9}$-ACP

KAS

$16:1^{\Delta 11}$-ACP **KAS**→ $18:1^{\Delta 13}$-ACP

FAT   **FAT**

$16:1^{\Delta 11}$ $18:1^{\Delta 13}$

ACS   **ACS**

$16:1^{\Delta 11}$-CoA $18:1^{\Delta 13}$-CoA

KCS 2

Condensation	Condensation
Reduction	Reduction
Cyclization	Cyclization
Reduction	Reduction
Aromatization	Aromatization

$22:1^{\omega 5}$ $24:1^{\omega 5}$

Anacardic acid   **Anacardic acid**

KAS I

16:0-ACP **KAS II**→ 18:0-ACP **AAD**→ $18:1^{\Delta 9}$-ACP

FAT B   **FAT B**   **FAT A**

16:0   18:0 $18:1^{\Delta 9}$

ACS

Acyl-CoA pool

Glycerolipid Biosynthesis
Very Long Chain Fatty Acid Biosynthesis

Enzyme	EST's	Contigs	Singleton
ACCase	5	2	1
MT	0	0	0
ACP	23	2	1
ER	8	1	1
HR	0	0	0
KAR	3	1	1
KAS I	15	4	0
KAS II	3	1	0
KAS III	1	0	1
FAT A	11	3	0
FAT B	11	3	0
ACS	0	0	0

Figure 2.4. A model for primary fatty acid synthesis leading to production of unusual monoenes. This model demonstrates the theory of substrate channeling (metabolic split) of primary fatty acid pathway towards one leading to glycerolipid and very long chain fatty acid synthesis and the other leading to production of unusual monoenes that serves as precursors for anacardic acid synthesis. The box on bottom right indicates the number of contigs and singleton's from the EST database for fatty acid biosynthesis enzymes. ACCase (acetyl-CoA carboxylase), MT (malonyl transferase), ACP (acyl carrier protein), KAS (β ketoacyl-ACP synthase), AT (acetyl transferase), ER (enoyl reductase), HD (dehydratase), AAD (acyl-ACP desaturase), ACS (acyl-CoA synthetase), MAD (myristoyl acp desaturase), FAT (fatty acid acyl-ACP thioesterase) KCS 2 (ketoacyl-CoA synthase). Enzymes marked in red are highly expressed in trichomes.

Material and Methods

Plant Material

Pelargonium × horotrum accession 88-51-10 was a kind gift from Dr. Richard Craig, Pennsylvania State University. The plants were propagated through vegetative cuttings and grown in MetroMix 360 media. Plants were maintained in environmental growth chambers at 20°C, 16-hour photoperiod and light intensity - photosynthetic photon flux = 250 µmol m^{-2} s^{-1}.

RNA extraction and cDNA synthesis

Pedicles were harvested from the flower clusters and transferred into 50 ml polypropylene tubes, placed on ice, then flash frozen in liquid nitrogen and stored at -80°C until used. Dry ice (~1g) was added to a tube containing pedicles and vortexed for 1 min to shear off trichomes from the surface of pedicle (Figure 2.3)[57]. The trichomes adhere to the surface of the tubes while pedicles are transferred to another tube for extraction of bald pedicle (Figure 2.3)[57]. Total RNA was extracted and purified as previously described [76]. Three biological replicates were obtained for all samples. RNA quality and quantity was analyzed using both nanodrop and bioanalyzer. RNA samples were Dnase treated using Ambion Turbo DNA-Free^{TM} kit (ThermoFisher) and cDNA was synthesized using SuperScript® III First-Strand Synthesis System (Invitrogen).

Bioinformatics Analysis

VecScreen server from NCBI (http://www.ncbi.nlm.nih.gov/tools/vecscreen/) was used to eliminate the vector contamination from EST sequences and all other

sequencing work done for this project. Sequences have been assembled into contigs and singletons. Contigs are groups of overlapping DNA reads that have consensus sequences and singletons are DNA reads that do not align with other sequences [77, 78]. Each contig/singleton represents a unique gene, a splice variant or a gene paralog. Contig analysis of EST clones for each gene was done using Vector NTI, Cap3 and Geneious software. Sequence alignment was done using Clustal W2 and AlignX (DNAStar suite software). BlastX server was used to identify target sequences for each contig and singleton. For sequence analysis, restriction digest was performed on 19 KAS clones and 22 FAT clones to identify the longest clone representing each contig. Missing sequence information obtained after molecular analysis (discussed below) was incorporated into the database and re-analyzed. Each of these genes were sequenced multiple times to resolve sequencing errors and the complete sequences were translated in 6 frames using EMBL-EBI six pack translator EMBOSS to identify the coding region.

Molecular Analysis

All restriction enzymes, PCR reagents, T4 DNA ligase, alkaline phosphastase used for molecular work were obtained from Roche Diagnostics, primers from Eurofins, pGEM®-T Easy Vector from Promega, pBluescript SK minus vector and DH5α *E.coli* cells from Novagene.

For primer walking experiments, the primer design and location was selected based on the contig alignment and an overlap of at least 100-150 base pairs with the existing clone sequence was taken into consideration. Every primer (Appendix 2, Table A2.1) was verified to be absent of secondary structures/hairpins and to ensure GC content was 40-60% (Oligonucleotide Properties Calculator). Additional sequence information was obtained by sub-cloning (Appendix 2, table A2.2) target fragments into pGEM®-T Easy Vector (Promega) or pBluescript followed by sequencing. All supplemental sequencing was conducted at the DNA core facility, University Of Louisville. For clones lacking 5' sequence information, Gene Racer[R] core kit with GeneRacer™ III RT Module (Invitrogen) was used for RACE following manufacturers instructions with gene specific primers (Appendix 2, Table A2.3). The cDNA was amplified using touchdown PCR initiated at 94°C (2 minutes) followed by 5 cycles of 94°C (30 seconds) and 72°C (1.5 minutes) then 5 cycles of 94°C (30 seconds) and 70°C (1.5 minutes) followed by 25 cycles of 94°C (30 seconds), 60°C - 68°C (30 seconds), 68°C - 72°C (1.5 minutes) and ended with final elongation of 68°C - 72°C (10 minutes). PCR amplicons were purified using the gene racer kit and sent for sequencing at the DNA core facility.

Phylogenetic Analysis

For phylogenetic analysis, deduced amino acid sequences for target geranium sequences were obtained using EMBL-EBI six pack translator EMBOSS. NCBI GenBank was used to obtain target protein sequences of additional plant species

used to generate the tree and multiple protein sequence alignment was generated by Muscle (EMBL-EBI) with Phylip interleaved output[79]. The phylogenetic trees were constructed from these alignment using RAxml (Trex-online)[80] with the following settings: substitution model: PROTCAT; Matrix name: DAYHOFF; algorithm: (d) Hill-climbing-default; number of bootstraps =1000; Bootstrap random seed (b) 12345 and the percent of replicates shown on the tree nodes.

qRT-PCR assessment of candidate genes

SYBR green and TaqMan assays were conducted in VIIA7[Tm] Real time PCR system from Applied Biosystems, software version 1.2.4. The default ABI PCR conditions were used starting with 50°C (2 minutes), 95°C (10 minutes), followed by 95°C (15 seconds) and 60°C (1 minute) for 40 cycles with temperature increment gradient as 1.6°C/s. RNA extraction and cDNA synthesis was accomplished as discussed above using primers designed for specific sequence regions (Appendix 2, Tables A2.4 and A2.5).

SYBR Green Assay

All-in-One™ qPCR Mix (GeneCopoeia), ROX reference dye at final concentration of 150nM, 2 µl of 1:10 dilution of cDNA and 0.2 µM of final concentration of primers were used for each assay. Primers were designed manually for each contig of interest based on the sequence alignments of the contigs using Oligonucleotide Properties Calculator. Each primer length ranged from 18-30

bases and GC content was between 50-60%. These primers were verified to be absent of self-annealing, complementarity or hairpin formation. Target amplicon lengths ranged between 150-300 bp. Each primer set was tested for specificity using standard end point PCR and qRT-PCR with plasmid preparations of each gene as template (template dilutions ranging from 0.0001 ng to 10 ng). Standard PCR results were viewed via gel electrophoresis and qRT-PCR results were analyzed based on cycle threshold values to verify primer specificity.

TaqMan assay

For all PCR reactions 2 μl of 1:10 dilution of cDNA was added to 5 μl of TaqMan^R gene expression Master Mix (Life Technologies) along with 0.5 μl of gene expression assay (Life Technologies). For each gene expression assay, primers and probe location was selected manually based on sequence alignment of each contig. The selected regions were then incorporated into the program Primer Express™ (Perkin-Elmer, Applied Biosystems, USA) to verify the primer-probe design and efficiency. The 5' end of the probe had the fluorescent reporter dye, 6-carboxy-fluorescein (FAM) and the quencher 6-carboxy-tetramethyl-rhodamine (TAMRA) located at the 3' end of the probe. Each assay was tested for specificity using qRT-PCR with plasmid preparations of each gene as template (template dilutions were 0.0001 ng to 10 ng). The qRT-PCR results were analyzed based on cycle threshold values to verify primer specificity.

Statistical Analysis

The qRT-PCR data were analyzed using paired t-tests to compare the null hypothesis of a 1:1 expression of trichome to bald pedicle ratio. One-way ANOVA with Tukey's multiple comparison and correction test was used to compare gene-to-gene expression levels.

Results and Discussion

ACPs, KASs and FATs of *Pelargonium × horturm*

Contig assembly of EST sequences indicated 3 contigs of FAT-As, 3 contigs of FAT-Bs, 4 contigs of KAS Is, 1 contig KAS II, 1 singleton for KAS III and 2 contigs of ACPs (Table 2.1). Complete gene sequences of ACP 1, ACP 2, KAS Ia, KAS Ib, KAS III, FAT-A1 and FAT-A2 were obtained after contig assembly, primer walking and sub-cloning[81]. Further sequence information was obtained after 5' RACE of KAS Ic, KAS II, FAT-B1, FAT-B2 and FAT-B3 [82]. KAS II, FAT B2 and FAT B3 failed in 5' RACE (Table 2.1). The sequence identity of each contig (appendix 6) was confirmed by BlastX homology with other plant species genes t that were characterized, identified and published (Table 2.1)[9, 28, 29, 32, 49, 50, 83-89].

Table 2.1 Detailed output of EST database genes ACPs, KASs and FATs.

Contigs	EST Clones	Identity	% Match	Plant	Complete	Nucleotide	Protein	GenBank Accession
TE Contig 1	03F05 (FAT-A1)	FAT A	75%	*H.annuus*	Yes	1682 bp	425aa	AAL79361.1
TE Contig 2	07A05 (FAT-A2)	FAT A	70%	*G.mangostana*	Yes	1599 bp	348aa	AAB51524.1
TE Contig 3	03C01 (FAT-A3)	FAT A	71%	*C.sativum*	Yes	1596 bp	334aa	Q42712.1
TE Contig 4	40F05 (FAT-B1)	FAT B	53%	*T.cacao*	Yes	1376bp	363aa	XP_007013277.1
TE Contig 5	21G11(FAT-B2)	FAT B	79%	*J.curcas*	No	1059bp	148aa	NP_001292946.1
TE Contig 6	22A03 (FAT-B3)	FAT B	68%	*J.curcas*	No	1571bp	249aa	NP_001292946.1
KAS Contig 1	02B12 (KAS I-a)	KAS I	78%	*A.thaliana*	Yes	1905bp	482aa	AAC49118.1
KAS Contig 2	09E03 (KAS I-b)	KAS I	81%	*A.thaliana*	Yes	1601bp	464aa	AAC49118.1
KAS Contig 3	42C01 (KAS I-c)	KAS I	82%	*C.nucifera*	Yes	1645bp	379aa	AGJ84410.1
KAS Contig 4	20H06 (KAS I-d)	KAS I	61%	*A.hypogaea*	No	837bp	90aa	ACJ07140.1
KAS Contig 5	01B08 (KAS II)	KAS II	88%	*A.hypogaea*	No	950bp	186aa	ACJ07142.1
KAS Singleton	28B11 (KAS III)	KAS III	84%	*J.curcas*	Yes	1753bp	389aa	NP_001292956.1
ACP Contig 1	06E07 (ACP 1)	ACP	71%	*C.sativum*	Yes	640bp	85aa	AAD46394.1
ACP Contig 2	30E09 (ACP 2)	ACP	66%	*C.sativum*	Yes	754bp	84aa	AAD46394.1

Blue highlights show incomplete sequences. GenBank Accessions[9, 28, 29, 32, 49, 50, 83-86, 89], Blast X version 2.5.1+[90]

Phylogenetic Analysis

Phylogenetic trees were constructed for geranium ACPs, TEs and KASs compared to selected sequences of other plant species using ClustalW, Muscle alignment (EMBL-EBI) and Raxml Trex-online (Figures 2.5, 2.6 and 2.7)[79, 80, 91-93].

Results show the geranium ACP 2, FAT-A1, KAS Ic, share a common phylogenetic node (are in the same clade) with ACP, FAT-A and KAS I respectively from *Coriander sativum*. Importantly, these coriander genes are involved in production of petroselinic acid (an unusual monoene) [9, 33, 34, 94]. These results indicate that the sequences of *Pelargonium × hortorum* ACP 2, FAT-A1, and KAS Ic could share similar function and are potentially involved in production of unusual monoene in trichomes of geranium. Based on this

analysis, transcription levels of these target sequences was hypothesized to be expressed at higher levels in trichome tissue of the geranium, reflecting a potential role in metabolism of UMFA found only in this tissue. All other geranium target sequences aligned with fatty acid biosynthesis transcript sequences from plants that are not specific to synthesis of unusual fatty acids. It is important to note that 1000 bootstrap replicates were used to increase the accuracy of the phylogentic tree along with setting 12345 as bootstrap random seed value to generate repeatable data with high statitsical confidence values for the phylogentic tree. This method is conservative and it lead to lower percentage values for some genes as seen in Figure 2.5, Figure 2.6 and Figure 2.7.
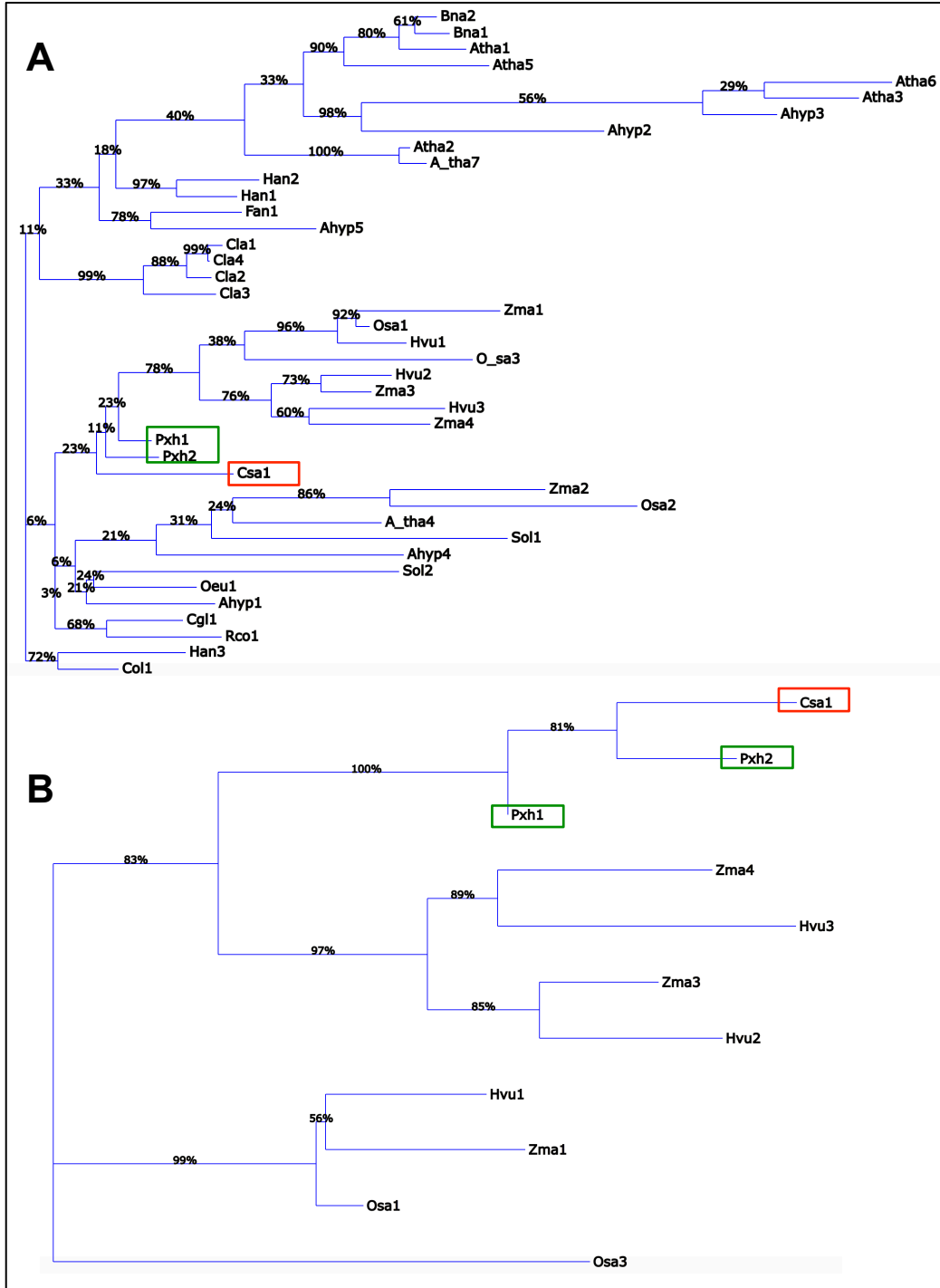
Figure 2.5. Phylogenetic tree of acyl carrier proteins. A) Overall Raxml output. B) Raxml output for genes in same clade as *C.sativum* gene. Green Box - Pxh (*Pelargonium × hortorum*), Red Box - Csa (*Coriandrum sativum*), Ahp (*Arachis hypogaea*), Ath (*Arabidopsis thailana*), Bna (*Brassica napus*), Cg (*Citrus glauca*), Cla (*Cuphea lanceolata*), Col (*Camellia olifera*), Fan (*Fragaria ananassa*), Han (*Helianthus annuus*), Hvu (*Hydra vulgare*), Oeu (*Olea europaea*), O.sa (*Oryza sativa*), Rco (*Ricinus communis*), Sol (*Spinacia oleracea*), Zma (*Zea mays*) and Nta (*Nicotiana tabaccum*). Suffix- Numerals indicates the isoform or paralog of ACPs in their respective plant species.

37
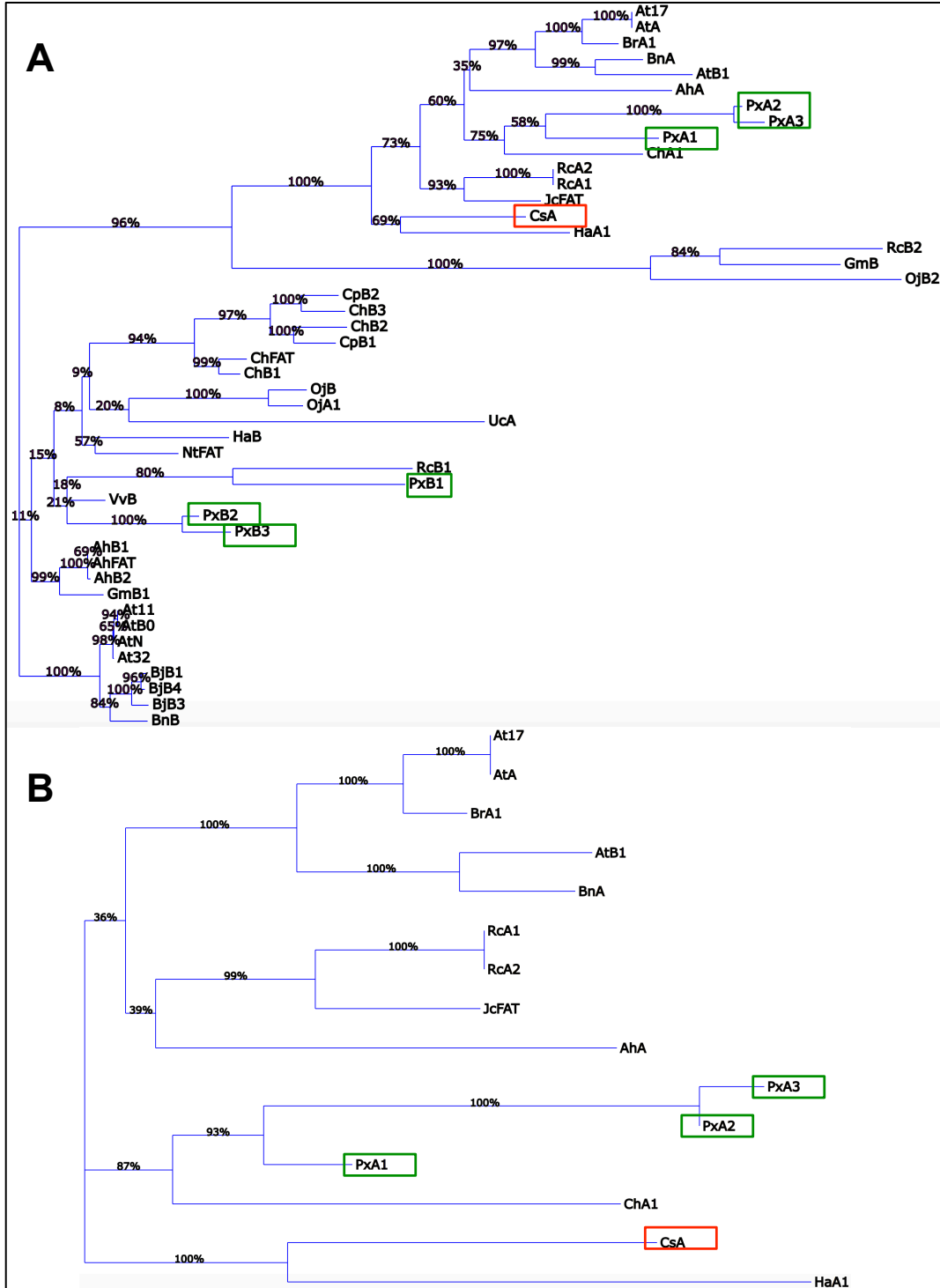
Figure 2.6. Phylogenetic tree of fatty acid thioesterases. A) Overall Raxml output.
B) Raxml output for genes in same clade as *C.sativum* gene. Green Box - Pxh
(*Pelargonium × hortorum*), Red Box - Csa (*Coriandrum.sativum*)*,* Uc (*Umbellularia
californica*), Cp (*Cuphea palustris*), Ath (*Arabidopsis thailana*)*,* Bna (*Brassica napus*), Han-
(*Helianthus annuus*)*,* Oj (*Oryza sativa Japonica*)*,* Jc (*Jatropha curcas*), Bj (*Brassica
Juneca*), Br (*Brassica rapa*), Gm (*Glycine Max*), Nta- (*Nicotiana tabaccum*)*,* Ch (*Cuphea
hookeriana*), Vv (*Vitis vinifera*), Rco (*Ricinus communis*)*,*Ah (*Arachis hypogaea*).
Suffix - A (FAT-A)*,* B (FAT-B).
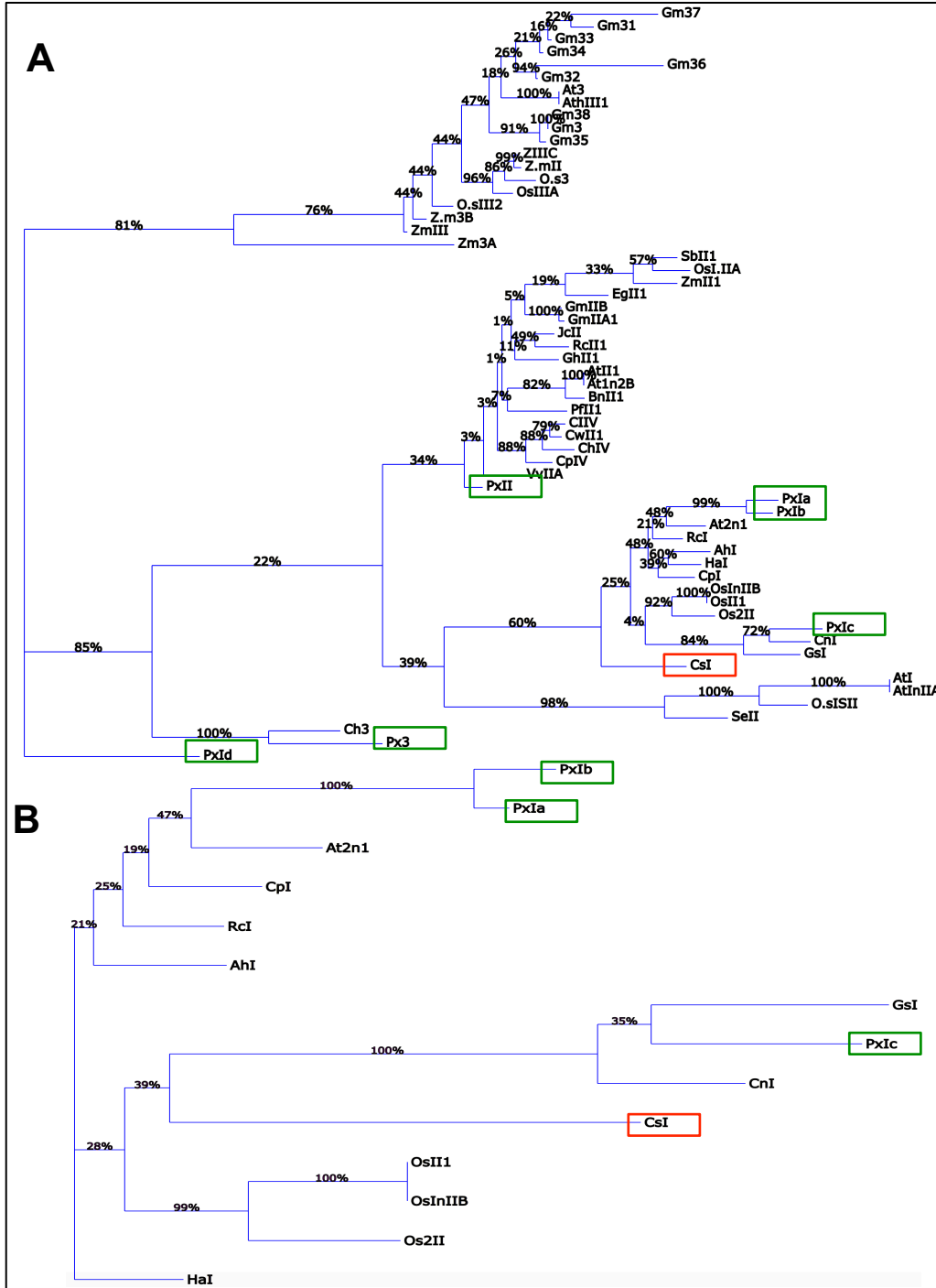
Figure 2.7. Phylogenetic tree of β-ketoacyl-ACP synthases. A) Overall Raxml output. B) Raxml output for genes in same clade as *C.sativum* gene. Green Box- Pxh (*Pelargonium × hortorum*), Red Box- Csa (*Coriandrum sativum*)*,* Ah (*Arachis hypogaea*),Ath (*Arabidopsis thailana),* Bna (*Brassica napus)*, Cw (*Cuphea wrightti*), Eg (*Elaeis guineenis)*, Gh (*Gossypium hirsutum*), Gm (*Glycine Max)*, O.sa (*Oryza sativa,*) Rco (*Ricinus communis*), Pf (*Perilla frutescens*), Sb- (Sorghum.*bicolor*), Han (*Helianthus annuus*), Vv (*Vitis vinifera*), Zma (*Zea mays*), Cn (*Cocos nucifera*), Jc (*Jatropha.curcas*), Cp (*Cuphea pulcherima)*, Cla (*Cuphea lanceolata*), Ch (*Cuphea hookeriana*), Gs (*Glycine suja*)*.* Suffix: Roman/Arabic Numerals correspond to KAS I,

KAS II and KAS III.
Analysis of target fatty acid biosynthesis genes transcript levels

Since unsaturated anacardic acids (22:1n5 and 24:1n5) and their precursors

UMFA's $16:1^{\Delta 11}$ and $18:1^{\Delta 13}$ are restricted to glandular trichomes, transcripts of

genes encoding fatty acid biosynthesis genes responsible for the production of

these UMFA's were hypothesized to be trichome specific. Based on

phylogenetic similarity of coriander gene transcripts involved in UMFA

biosynthesis, *Pelargonium × hortorum* ACP 2, FAT-A1, KAS I-a, KAS I-b and

KAS Ic were hypothesized to be the sequences that would be trichome specific.

A SYBR green qRT-PCR assay was used to compare the expression patterns of

KAS genes in trichome and bald pedicle tissue whereas a TaqMan qRT-PCR

assay was used for ACP and FAT since distinct sequences within each class had

high sequence similarity and thus required a probe for increased specificity. For

both assays ACTIN was used as reference gene (based on qPCR normalization

results that showed similar expression levels of ACTIN in both bald pedicle and

trichomes), Δ9 myristyl-ACP desaturase (MAD) and Δ9 stearoyl-ACP desaturase

(SAD) were used as positive control for higher ratio of expression in trichomes

and omega-3 desaturase (ω3) was used as positive control for higher ratio of

expression in bald pedicle. The ΔΔCT method was used to analyze qRT-PCR

data. In both assays, controls showed significant expression in their respective

tissues as expected, indicating the reliability of these results. The ACP 1, ACP 2,

FAT-A1, FAT-A2, KAS I-a/b and KAS Ic sequences were found to be significantly

higher in expression in trichome as compared to bald pedicle. KAS III shows a

non-significant trend of higher expression in trichome (Figure 2.8 and 2.9).

Combined expression of KAS Ia and Ib is shown since the assay couldn't

differentiate between the two paralogs.
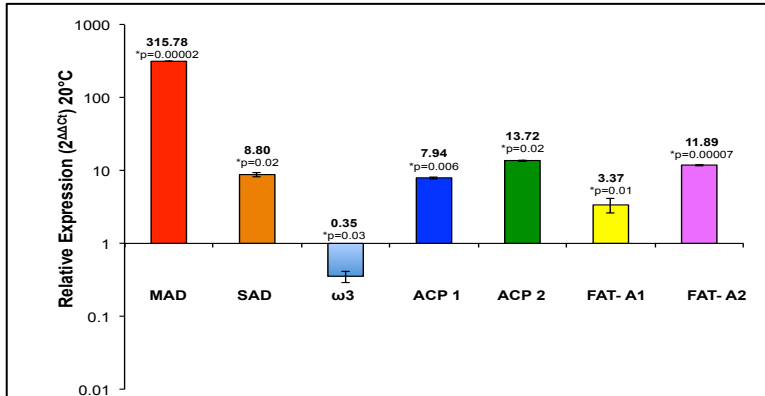
Figure 2.8. Relative expression of trichome compared to bald pedicle for ACPs and FAT-As. Δ9 Myristyl ACP desaturase (MAD),Δ9 stearoyl-ACP desaturase (SAD) and omega-3 desaturase (ω3) were used as controls. Y-axis shows average fold change valuesBased on 1:1 expression of trichome to bald pedicle, fold change values above 1 suggest higher expression in trichomes and below 1 suggest higher expression in bald pedicle. Bars represent the standard error of means and the means are represented by black numerals at the end of each bar with p-values of the t-test are shown for each gene. Asterisk before p-value indicates significantly different expression from 1:1 ratio.
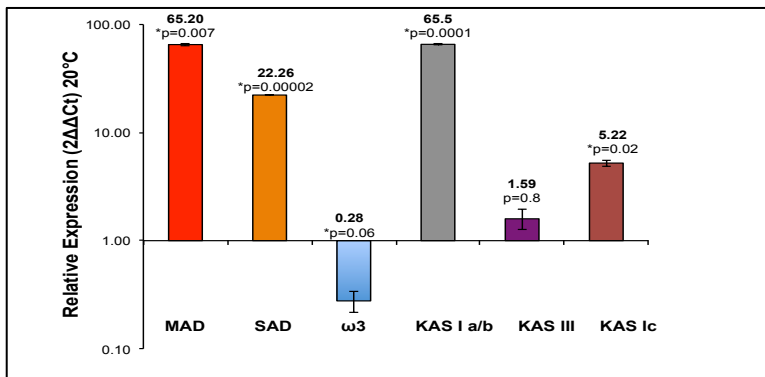
Figure 2.9 Relative expression of trichome compared to bald pedicle for KASs. Δ9 Myristyl ACP desaturase (MAD), Δ9 stearoyl-ACP desaturase (SAD) and omega-3 desaturase (ω3) were usedas controls. Y-axis shows average fold change values. Based on 1:1 expression of trichome to bald pedicle, fold change values above 1 suggest higher expression in trichomes and below 1 suggest higher expression in bald pedicle. Bars represent the standard error of means and the means are represented by black numerals at the end of each bar with p-values of the t-test are shown for each gene. Asterisk before p-value indicates significantly different expression from 1:1 ratio.

For direct expression comparison between selected specific genes, Tukeys multiple correction was used on one-way ANOVA. The delta Ct values of the TaqMan assay show that expression of MAD is significantly higher in trichome tissue compared to all the other genes. Both ACP's and FAT-A's show high expression in trichomes and expression of ACP 1 versus ACP 2 and FAT-A1 versus FAT-A2 are not different from each other in both trichome and bald pedicle (Figure 2.10 and 2.11). The delta Ct values of the SYBR green assay show expression of KAS 1-a/b is significantly higher in trichomes than all the genes. This may be due to the summed expression of two paralogs. KAS Ic expression is higher than KAS III in trichomes. Within bald pedicle expression of KAS 1-a/b, KAS III and KAS I-c is not significantly different from one another (Figure 2.12 and 2.13).

All the genes showing higher expression ratios of trichome to bald pedicle have delta Ct values that show lower expression within bald pedicle tissue compared to trichome tissue in both the assays (Figure 2.8 to 2.13). Based on these results it may not be possible to use tissue specific expression patterns alone to predict which ACP, TE or KAS is the best candidate that may be involved in UMFA synthesis. It is possible that all of them are involved in UMFA synthesis but this requires further biochemical evaluation.

Trichome Expression (ΔCt value)

Chart values (mean at end of each bar):
- MAD: -4.75
- SAD: 2.08
- ω3: 5.48
- ACP 1: -1.74
- ACP 2: -3.11
- FAT-A1: 3.26
- FAT-A2: 3.76
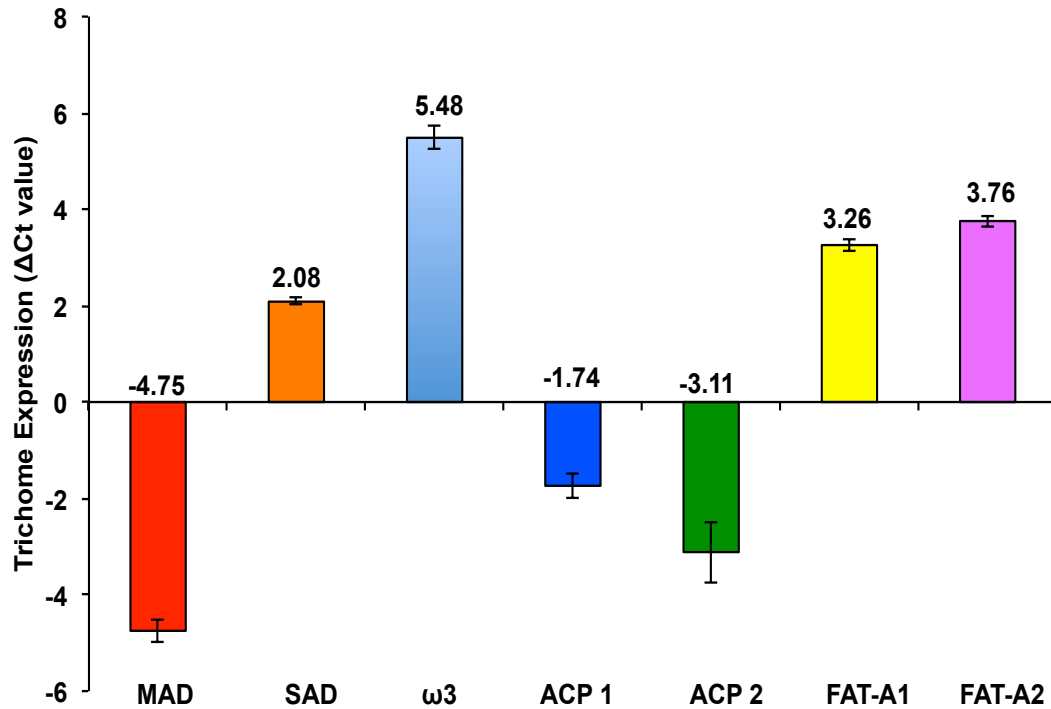
Post Hoc Analysis						
pvalues for gene to gene comparions in trichome tissue at 20°C						
Genes	ACP1	ACP2	MAD	ω3	SAD	FAT-A1	FAT-A2
---	---	---	---	---	---	---	---
ACP1	N/A	0.0607	<.0001	<.0001	<.0001	<.0001	<.0001
ACP2	0.061	N/A	0.02	<.0001	<.0001	<.0001	<.0001
MAD	<.0001	0.02	N/A	<.0001	<.0001	<.0001	<.0001
ω3	<.0001	<.0001	<.0001	N/A	<.0001	0.0015	0.013
SAD	<.0001	<.0001	<.0001	<.0001	N/A	0.139	0.016
FAT-A1	<.0001	<.0001	<.0001	0.001	0.139	N/A	0.884
FAT-A2	<.0001	<.0001	<.0001	0.013	0.016	0.884	N/A

Figure 2.10. Expression patterns of ACPs and FAT-As in the trichome tissue. MAD (Δ9 myristyl ACP desaturase), SAD (Δ9 stearoyl-ACP desaturase) and ω3 (omega-3 desaturase) were used as controls. The y-axis shows delta Ct values with the expression level inversely proportional to delta Ct values. The bars represent the standard error of means and the means are represented by black numerals at end of each bar. The p-values are shown for gene comparisons (table) using Tukeys multiple correction on one-way ANOVA. Values <0.05 show significant difference in expression between two genes.

			Post Hoc Analysis				
		pvalues for gene to gene comparions in bald pedicle tissue at 20°C					
Genes	ACP1	ACP2	MAD	ω3	SAD	FAT-A1	FAT-A2
ACP1	N/A	0.988	0.133	0.053	0.003	0.005	<.0001
ACP2	0.988	N/A	0.037	0.014	<.0001	0.002	<.0001
MAD	0.134	0.037	N/A	0.997	0.424	0.567	0.005
ω3	0.053	0.014	0.997	N/A	0.728	0.857	0.014
SAD	0.003	<.0001	0.424	0.728	N/A	1	0.197
FAT-A1	0.005	0.001	0.567	0.856	1	N/A	0.131
FAT-A2	<.0001	<.0001	0.005	0.013	0.196	0.131	N/A

Figure 2.11. Expression patterns of ACPs and FAT-As in the bald pedicle tissue. MAD (Δ9 myristyl ACP desaturase), SAD (Δ9 stearoyl-ACP desaturase) and ω3 (omega-3 desaturase) were used as controls. The y-axis shows delta Ct values with the expression level nversely proportional to delta Ct values. The bars represent the standard error of means and the means are represented by black numerals at end of each bar. The p-values are shown for gene comparisons (table) using Tukeys multiple correction on one-way ANOVA. Values <0.05 show significant difference in expression between two genes.

Figure 2.12. Expression patterns of KASs in the trichome tissue. MAD (Δ9 myristyl ACP desaturase), SAD (Δ9 stearoyl-ACP desaturase) and ω3 (omega-3 desaturase) were used as controls. The y-axis shows delta Ct values with the expression level inversely proportional to delta Ct values. The bars represent the standard error of means and the means are represented by black numerals at end of each bar. The p-values are shown for gene comparisons (table) using Tukeys multiple correction on one-way ANOVA. Values <0.05 show significant difference in expression between two genes.
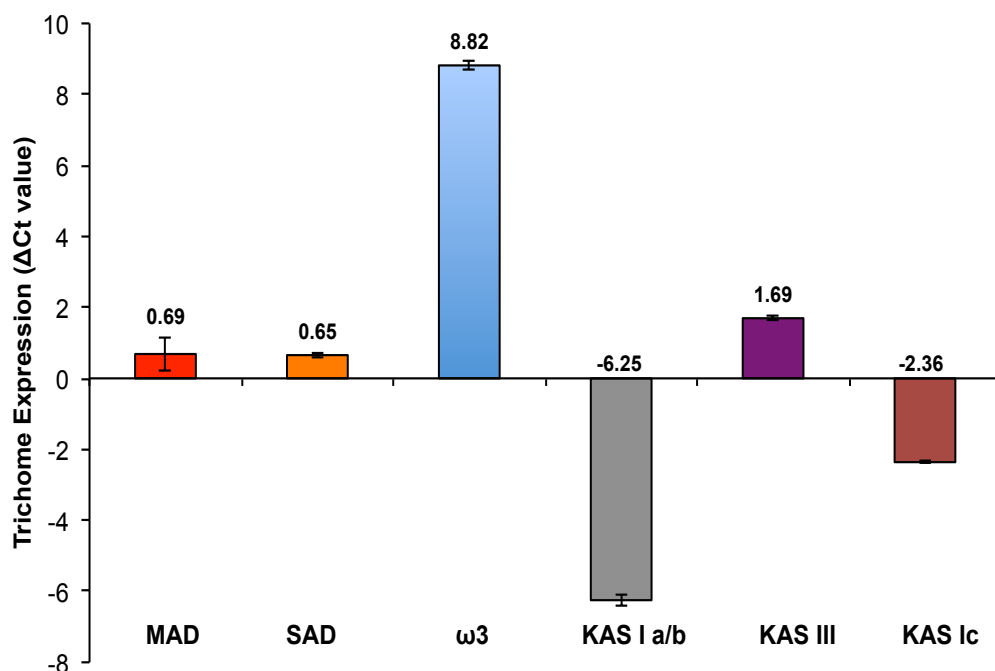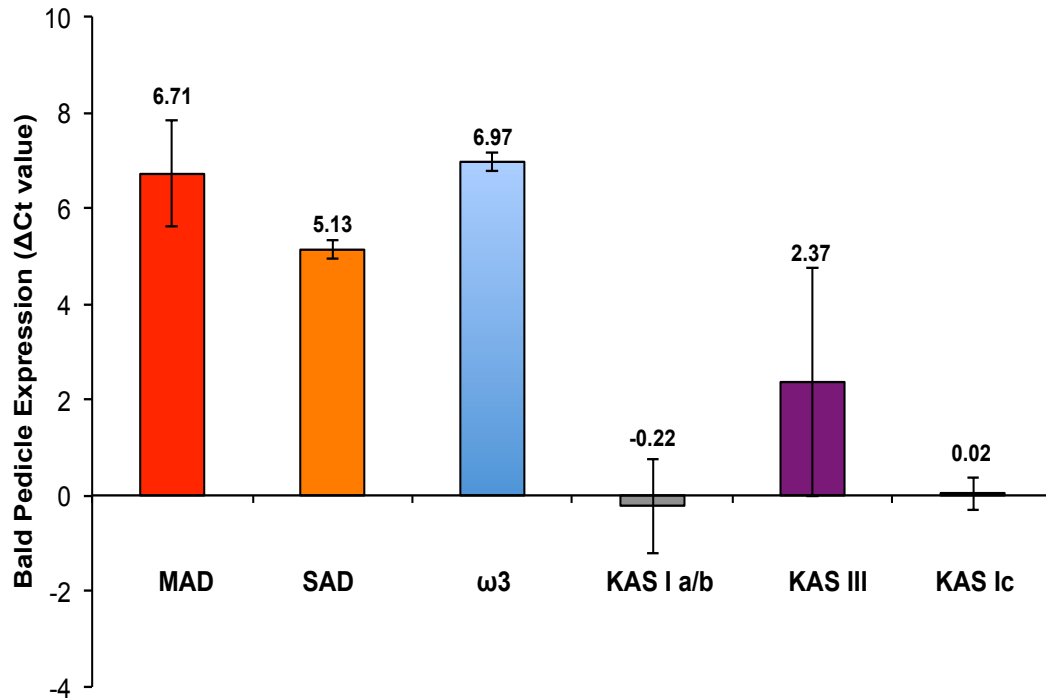
Post Hoc Analysis					
pValues for gene to gene comparisons in bald pedicle tissue at 20°C					
Genes	KASI a/b	KASIII	KASI-c	MAD	ω3	SAD
KASI a/b	N/A	0.633	1	0.012	0.009	0.061
KASIII	0.633	N/A	0.715	0.162	0.128	0.571
KASI-c	1	0.715	N/A	0.016	0.012	0.078
MAD	0.012	0.162	0.015	N/A	1	0.922
ω3	0.009	0.127	0.012	1	N/A	0.867
SAD	0.061	0.57	0.077	0.922	0.867	N/A

Figure 2.13. Expression patterns of KASs in the bald pedicle tissue. MAD (Δ9 myristyl ACP desaturase), SAD (Δ9 stearoyl-ACP desaturase) and ω3 (omega-3 desaturase) were used as controls. The y-axis shows delta Ct values with the expression level inversely proportional to delta Ct values. The bars represent the standard error of means and the means are represented by black numerals at end of each bar. The p-values are shown for gene comparisons (table) using Tukeys multiple correction on one-way ANOVA. Values <0.05 show significant difference in expression between two genes.

Conclusion

Complete sequences of 2 ACP's, 6 KAS's and 6 FATs were identified and used to analyze tissue specific expression in glandular trichomes of *Pelargonium* × *hortorum*. Combined results of expression analysis and phylogenetic analysis showed that ACP 2, KAS I-c and FAT-A1 were highly expressed in trichomes and aligned closely with homologs of genes from coriander that are known to be involved in UMFA biosynthesis. ACP 1, KAS Ia, KAS Ib and FAT-A2 did not align with known genes involved in UMFA biosynthesis but also showed high expression in trichomes. Both phylogenetic analysis (based on sequence homology) and tissue specificity analysis were used as methods to characterize the EST genes but that may not always to be an absolute method to define or validate function. It is possible that the low expression genes and the genes that did not align with genes of plants producing unusual monoenes may still be involved in production of UMFAs in geranium. Furthermore, there are 2 ACPS, 3 KASs and 2 FATs that were highly expressed in trichomes, so it is possible that they are either paralogs (gene duplicates) that have the exact same function or they are isoforms (splice variants) that are found at the specific stage of development and involved in UMFA metabolism. To further define their potential role UMFA productions these genes need to be expressed in heterologous systems.

CHAPTER 3

EFFECT OF TEMPERATURE ON PRODUCTION OF UNUSUAL MONOENES,
ANACARDIC ACID AND FAS GENE EXPRESSION IN THE TRICHOMES OF
PELARGONIUM × HORTORUM.

Summary

Anacardic acid (AnAc) is a collective term for congeners of 2-hydroxy-6-alkyl
benzoic acid. AnAc is ubiquitously produced in trichomes of all *Pelargonium ×
hortorum* (garden geranium) plants but specific congeners of AnAc (AnAc 22:1n5
and AnAc 24:1n5) are known to impart a pest resistance phenotype to geranium.
These specific congeners are derived from the unusual monoenoic fatty acids
(UMFAs) $16:1^{\Delta 11}$ and $18:1^{\Delta 13}$ that are synthesized only in glandular trichomes of
pest-resistant plants. Thus, UMFAs and their AnAc metabolites provide useful
biological markers that differentiate the biosynthetic pathways for unusual
mononenes and for common fatty acids. To elucidate the effect of temperature
and identify the genetic factors associated with UMFA and AnAc biosynthesis, a
de novo RNA transcriptome was generated from trichomes and bald pedicle of
Pelargonium × hortorum obtained from plants at 18°C and 23°C. A total of
486398 transcripts were generated and bioinformatics analysis was used to
identify sequences associated with temperature dependent alteration in UMFA
and AnAc synthesis that were differentially expressed in a specific tissue or at

the specific temperature. Expression data for target sequences obtained from the RNA transcriptome were further tested and validated using quantitative real time qRT-PCR (including tissues at 28°C) and biochemical analysis (HPLC-DAD and GC-FID) were used to evaluate production of AnAc and UMFAs. Correlation analyses of qRT-PCR data and levels of UMFAs and derived AnAc indicated that not only is 23°C is the optimal temperature for UMFA and AnAc synthesis compared to 18°C and 28°C but also indicated higher temperature negatively affects the production of metabolites. Production of UMFAs ($16:1^{\Delta 11}$ and $18:1^{\Delta 13}$) is positively correlated with production of AnAc (22:1n5 and 24:1n5) at all temperature's indicating a direct relationship between the amount of substrate and the metabolite at a given temperature. Finally, expression of ACP 1, ACP 2, KAS I-a/b, and KAS I-c were significantly correlated with production of target metabolites at specific temperatures (18°C, 23°C and 28°C) further validating their potential involvement in UMFA metabolism. This approach can be used to identify more genetic components and to study effects of environmental factors like light intensity and photoperiod affecting UMFA and AnAc metabolism.

Introduction

Unusual monoenoic fatty acids are important because their distinct chemical structures make them valuable feedstocks for the chemical industry where they have a wide variety of applications as polymers, fuels and renewable sources of energy[5, 95]. Additionally, within trichomes of *Pelargonium × hortorum* these UMFAs ($16:1^{\Delta 11}$ and $18:1^{\Delta 13}$) are the direct precursor for production of

specialized ω5 AnAc metabolites (AnAc 22:1n5 and AnAc 24:1n5) that not only confer pest resistance to plants but also have antibacterial, antifungal and anticancer properties[36]. Fatty acid biosynthesis enzymes are required for common fatty acids as well as the UMFA biosynthesis and research in this area of plant biology suggests that UMFAs are products of primary metabolism that may require distinct isoforms of fatty acid biosynthesis enzymes for their synthesis[4]. This is based on the identification of novel FAS genes like ACPs, β-keto acyl-ACP synthase (KASs), acyl-ACP thioesterase (TEs) and acyl-ACP desaturases (AADs) that have specific roles in the production of UMFA in their native plants [9, 22, 23, 32, 33, 96]. Apart from Δ9 myristyl-ACP desaturase (MAD) no other genetic components involved in UMFA synthesis has not be identified and the biosynthetic pathway leading to anacardic acid has not yet been elucidated completely[75]. Similarly, limited information is available on the influence of abiotic environmental factors (like light intensity, moisture, photoperiod, humidity, temperature and soil quality) on gene expression correlated with primary metabolism and derived specialized metabolite accumulation [97-102]. Changes in temperature affect the viscosity of trichome exudate containing ω5 AnAc metabolites which alters the lipid profile and bioactivity related to pest resistance[97]. Thus temperature is an important abiotic factor that can be used to correlate changes in expression of genes suspected to be involved with AnAc and UMFA production and it can also be used for screening and identifying genes involved in UMFA and AnAc metabolism.

To study the biochemical pathway and identify the genetic components involved in production of UMFAs that result in AnAc 22:1n5 and AnAc 24:1n5 accumulation a *Pelargonium × hortorum* (geranium) EST (expressed sequence tag) database was initially used to identify target fatty acid biosynthesis cDNA sequences. To expand the identification of genetic factors and to assess the effect of temperature, a high throughput next-generation sequencing (NGS) Illumina platform was employed as both a quantitative and a qualitative discovery approach[103-106]. NGS provides deeper sequencing information and better quality as compared to EST database along with gene expression analysis. This approach provided complete sequence information that was either missing or incomplete in the EST database and it provided the capacity to study the *Pelargonium × hortorum* transcriptome in the absence of a sequenced genome [103, 104, 107]. Fatty acid biosynthesis gene expression data was obtained for ACPs, KASs and FAT-As from the *de novo* RNA transcriptome and data were validated using qRT-PCR. HPLC-DAD and GC-FID were used to quantify AnAc (22:1n5 and 24:1n5) and UMFAs ($16:1^{\Delta 11}$ and $18:1^{\Delta 13}$) respectively in trichomes harvested from plants grown at defined temperatures. Both genetic and biochemical outputs were used to assess the effect of temperature on UMFA and AnAC biosynthesis and were correlated with changes in gene expression to identify the novel genes involved in UMFA metabolism.

Material and Methods

Plant Material

Pelargonium × hortorum accession 88-51-10 was a kind gift of Dr. Richard Craig, The Pennsylvania State University. The plants were vegetatively propagated and were grown in one-gallon pots containing MetroMix 360. Plants were maintained in environmental growth chambers at $18^{\circ}C$, $23^{\circ}C$ and $28^{\circ}C$ with a 16-hour photoperiod and photosynthetic photon flux = 250 µmol $m^{-2} s^{-1}$.

RNA isolation and metabolite extraction

Flower pedicles were harvested into 50 ml conical tubes on ice and then flash frozen in liquid nitrogen and stored at $-80^{\circ}C$ until used. Trichomes were sheared from the surface of frozen pedicles from multiple tubes as described [57, 76]. Trichomes from 2 x 50 ml conical tubes (trichomes were suspended in DEPC water) were combined in one tube for RNA isolation and metabolite extraction whereas three intact bald pedicles obtained after shearing trichomes were used for a single RNA and metabolite extraction. Both trichome and bald pedicle samples were treated with 800 µl of metabolite extraction buffer (methanol:chloroform:water (v/v/v) = 2.5:1:0.5) and centrifuged at 300 × g for 7 min, 4°C. After centrifugation, the supernatant was used for metabolite extraction in chloroform and the pellet was used for RNA extraction using Plant Spectrum total RNA extraction kit (Sigma-Aldrich) according to the manufacturers instructions (see, Appendix 3 for details). Three biological replicates were obtained from both trichomes and bald pedicles at each treatment temperature

($18°C$, $23°C$, and $28°C$). RNA quality and quantity was analyzed using both
nanodrop and bioanalyzer. RNA samples were DNase treated using Ambion
Turbo DNA-Free ^{TM} kit (ThermoFisher). One µg of RNA obtained from both
tissues at all three temperatures were used for cDNA synthesis using
SuperScript® III First-Strand Synthesis System (Invitrogen). cDNAs obtained
were used for qRT-PCR (TaqMan or SYBR green assays). Total RNA (5 µg)
from 3 biological replicates of trichomes and bald pedicle at $18°C$ and $23°C$ (12
samples total), was sent to the University of Louisville's Center for Genetics in
Molecular Medicine's (CGeMM) sequencing core facility for preparation of RNA
transcriptome. Metabolites obtained in the chloroform extract were used for
HPLC-DAD and GC-FID analysis.

qRT-PCR analysis of expression

SYBR green and TaqMan assays were conducted in VIIA7^{Tm} Real-time PCR
system from Applied Biosystems, software version 1.2.4. The default ABI
PCR conditions were 50°C (2 minutes), 95°C (10 minutes), followed 95°C (15
seconds) and 60°C (1 minute) for 40 cycles with temperature increment gradient
as 1.6°C/s. RNA extraction and cDNA synthesis was done as detailed above
using gene specific primers designed for either SYBR green or TaqMan assays
(see, Appendix 2, Table A2.4 and A2.5).

<u>SYBR Green Assay</u>

All-in-One™ qPCR Mix (GeneCopoeia), ROX reference dye at final concentration
of 150 nM, 2 µl of 1:10 dilution of cDNA and 0.2 µM primers was used for the

assay. Primers were designed manually for each sequence of interest based on the sequence alignments of the candidate genes, using Oligonucleotide Properties Calculator. Each primer length ranged from 18-30 bases, GC content between 50-60%. These primers were verified to have no self-annealing, complementarity or hairpin formation. Amplicon length ranged between 150-300 bp for various primer sets. Each primer set was tested for specificity using standard end point PCR and qRT-PCR with plasmid preparations of each gene as a template (template dilutions ranging from 0.0001 ng to 10 ng). Standard PCR results were viewed via Gel Electrophoresis and qRT-PCR results were analyzed based on cycle threshold values to verify primer specificity.

TaqMan assay

PCR reactions contained 2 µl of 1:10 dilution of cDNA added to 5 µl of TaqMan^R gene expression Master Mix (Life Technologies) along with 0.5 µl of gene expression assay (Life Technologies). For each gene expression assay, primers and probe, the location was selected manually based on sequence alignment of candidate genes. The selected regions were then incorporated into the program Primer Express™ (Perkin-Elmer, Applied Biosystems, USA) to verify the primer-probe design and efficiency. The 5' end of the probe had the fluorescent reporter dye, 6-carboxy-fluorescein (FAM) and the quencher 6-carboxy-tetramethyl-rhodamine (TAMRA) was located at the 3' end of the probe. Each assay was tested for specificity using qRT-PCR with plasmid preparations of each gene as a template (template dilutions ranging from 0.0001 ng to 10 ng). The qRT-PCR results were analyzed based on cycle threshold values to verify primer specificity.

HPLC analysis

AnAc content from the metabolite samples was estimated based on standard curves generated using pure AnAc 22:1n5 and AnAc 24:1n5 at 0.1, 0.25, 0.5, 1 and 10 mg/ml respectively (Appendix 3, Figure A3.1 and Figure A3.2). Metabolite samples obtained from trichomes at 18°C, 23°C, and 28°C (3 biological replicates from each) were dried under nitrogen gas then reconstituted in hexanes to obtain 0.5 mg/ml and 1 mg/ml sample extracts for each biological replicates. Each sample (20 μl) was injected in triplet technical replications for each biological replicate (1 biological replicate data = 3 technical replicates at each concentration). HPLC analysis was accomplished using HP Agilent HPLC 1100 series column at 2.5 mL/min flow rate, 20μl injection volume, a solvent system 50% A and 50% B (solvent A – acetonitrile/acetic acid; 99/1 and solvent B – methanol/acetic acid; 99/1).

GC analysis

Trichome metabolite extracts for three biological replicate samples from each temperature treatment (18°C, 23°C and 28°C) were dried under nitrogen gas. For quantitative analysis of fatty acid content, internal standard (triheptadecanoin) was added to the dried metabolite samples. Fatty acid methyl esters were generated using 1 ml boron triflouride (10% in methanol) and 300 μl of toluene was added to increase solubility of lipids prior to heating for 45 minutes at 90°C. Samples were cooled and quenched with 1 ml water prior to

hexane extractions (3 × 2 ml). The pooled hexane extractions of each sample

were concentrated under N_2 (g). Resulting fatty acid methyl esters were

analyzed by gas chromatography (GC) with an Agilent 5890 GC-FID equipped

with a 60 m, 0.25 mm, 0.2 μm CP-Sill88 column (Chrompack). The column was

programmed with spiltless inlet at 250°C, 21.76 psi, 10.3 ml/ml flow rate of

helium gas. The oven was set at an initial temperature of 150°C and ramped to

250°C at a rate of 2°C /min. The FID detectors was set at 250°C with hydrogen

flow rate of 3 ml/min, air flow rate of 400 ml/min. Three technical replicates (1 µl

/injection) for each biological replicate were used for the analysis. Authentic GC

standards C8-C22 (Sigma Aldrich, Appendix 3, Table A3.1) were used as a

external analytical standard to identify peaks in the experimental samples. Most

peaks within the C8-C22 were further verified by comparison of retention times to

other single of mixtures of standards. Identification of sample peaks was further

confirmed by spiking samples with authentic standards.

RNA transcriptome

The Truseq Stranded total RNA with Ribo Zero Plant (RS-122-2401) kit was used

to prepare total RNA libraries from 700 ng total RNA. Fragmentation of total

RNA was performed at 94°C for 30 seconds instead of 5 minutes in order to

optimize library size for larger fragments, which produced average fragment

lengths of 400 bp. Libraries were confirmed on the Agilent 2100 Bioanalyzer and

quantitated using the Illumina Library Quantification Kit, ABI Prism qPCR Mix

from Kapa Biosystems and the ABI7900HT real-time PCR instrument. Two times

151 bp paired-end sequencing was performed with the 500 high-output v2 (300cycle) sequencing kit on the Illumina NextSeq500 instrument.

Bioinformatics Analysis

Generation of *de novo* transcriptome and bioinformatics analysis was accomplished at the KBRIN Bioinformatics Core facility. The CGeMM DNA Core used Illumina TruSeqLT chemistry with 151 paired-end reads and sequenced twelve samples (trichome 18°C, n=3; bald pedicle 18°C, n=3; trichome 23°C, n=3; bald pedicle 23°C, n=3) using standard adapters.

Adapter   AGATCGGAAGAGCACACGTCTGAACTCCAGTCA

AdapterRead2   AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT

The resulting twelve samples were barcoded and identified. Sequences were run using the RNASeq protocol on the CGeMM Illumina NextSeq 500 platform[105]. Each replicate was sequenced with four lanes. A total 24 paired-end of RNA sequencing data (48 single files) was obtained and 96 paired-end raw sequencing files (.fastq) were downloaded from Illumina's Basespace website onto the KBRIN server (kbrin.hsb.louisville.edu). Each of the four lanes for each condition was concatenated together using the unix cat command. Quality control (QC) of the raw sequence data was performed using FastQC (version 0.10.1)[108] and FastQC results indicated that all the mean quality scores per bases are above 20. Therefore, it was determined that minor sequence trimming was not needed. *De novo* transcripts were assembled using the Trinity software package for each of the four experimental samples using the custom

splitSamples.pl script[109]. Given a large number of sequences for this project, the

first step was to perform *in silico* read normalization for each of the partitions

independently in order to reduce the number of reads used for *de novo* transcript

assembly. Following *in silico* read normalization for each partition, the resulting

*.normalized_K25_C30_pctSD200.fq files were combined using the unix cat

command into the files ALL_R1.fastq and ALL_R2.fastq. The resulting *in silico*

normalized reads results in 48,301,282 read pairs to be used for *de novo*

transcript assembly that is roughly 10% of the original data. *De novo* transcript

assembly on the normalized reads was performed using Trinity version 2.1.1[82]

and a total of 486,398 transcript contigs (contigs are a group of overlapping DNA

reads that have consensus sequences [77, 78]) resulted.

To assign a putative function to individual *de novo* transcripts, a database of non-

redundant plant proteins was constructed based on the NCBI nr database. For

this process, each of the files from the nr database (nr.*.tar.gz) was downloaded

from the NCBI ftp site ftp://ftp.ncbi.nlm.nih.gov/blast/db. These files were then

unpacked into a directory, resulting in approximately 150 Gb data. The genInfo

identifier number (GI) accessions for plant proteins were downloaded from the

NCBI protein database (http://www.ncbi.nlm.nih.gov/protein). First the query (all

[filter]) AND "green plants"[porgn:__txid33090] was used to return all green plant

proteins (5,821,183 such sequences). The GIs for each of these sequences

were then downloaded using the "Send To" link with the options Destination "File"

and Format "GI List". The resulting file was stored as sequence.gi.txt. The NCBI

blast executable was then downloaded from

ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/2.2.31/ncbi-blast-2.2.31+-
win64.exe and installed. Then the downloaded nr database was filtered for only

the GIs in the sequence.gi.txt using the custom script getPlantSequences.pl that

takes 10,000 sequences at a time. A total of 583 plant_<num>.faa files resulted.

These were then concatenated together into a single file ALLPLANTSEQ.faa.

Many of the headers contain redundant headers for the exact same sequence.

Therefore, redundant headers were removed using the custom script

removeDuplicateHeaders.pl. Using the non-redundant ALLPLANTSEQ.faa, a

blastable database was constructed. The contigs resulting from Trinity were then

compared against the database of known plant sequences into an XML output for

use by blast2go[110].   At the completion of the blastx step, a file of the

concatenated xml output resulted which was then parsed using a custom perl

script in order to reformat some of the xml lines so they could be appropriately

parsed by blast2go. Blast2go basic v3.1.3 was then downloaded from the

blast2go website https://www.blast2go.com. The raw fasta file

Trinity_UPDATED_CTGNAMES_part0-48.fasta was uploaded into blast2go in

order to add annotations under the start followed by load sequences and menu.

Annotations from the file tp0-48_PARSED.xml were added to the blast2go

project using the File followed by Load and Load Blast Results. Sequences and

annotations of sequences were then exported as fasta files within blast2go using

the File followed by Export and Export as FASTA option with the default

parameters set. Using the newly obtained reference contigs with added

annotations as a guide, the original fastq reads for each of the twelve samples

were aligned to the reference contigs for further differential gene expression analysis using the Trinity pipeline as described in http://trinityrnaseq.sourceforge.net/analysis/diff_expression_analysis.html. The first step in the process was the alignment process using bowtie2[111]. A custom perl script getGeneCounts.pl was created to run this process for each sample, which creates a directory for each sample.

To perform differential gene expression analysis, a gene matrix file was created where the first column represents each gene (contigs in this case) and each row represents the counts for each gene in each of the twelve conditions. A custom script createExpressionMatrices.pl was constructed to handle matrix construction using the Trinity script. Differential gene expression was run using a custom perl script getDEGs.pl which uses edgeR as the method for differential gene expression[112]. Differentially expressed genes for each of five comparisons were parsed according to an FDR cutoff of 0.05 and a |logFC| ≥ 1. The top 10 differentially expressed genes from each comparison, as ranked by lowest FDR value, were extracted. Further analysis of the contigs was performed to determine their properties, similar to the analysis performed and described in Y. Li *et. al*., 2015 and M. Salem *et. al*., 2015 [113, 114].

To assign Gene Ontology terms to each contig, the blast results were parsed according to the matching GI[115]. The GIs were converted to UniProt release 2015_12 identifiers using the file idmapping.dat.example obtained from ftp://ftp.uniprot.org/pub/databases/uniprot/current_release/knowledgebase/idmapping/idmapping.dat.gz[116]. After downloading the UniProt mappings, the resulting

file was parsed to only contain lines with a matching GenInfo Identifier (GI) using the parseUniprotFile.pl custom perl script resulting in the file uniprot2GI.dat. The resulting xml file from the blast results was parsed, and each GI for the top 20 hits to each contig was converted to a UniProt identifier (if applicable) using the file created and a custom perl script parseUniprotFile.pl. The resulting UniProt identifiers were then associated with each of the respective annotated Gene Ontologies using the association file download

ftp://ftp.ebi.ac.uk/pub/databases/GO/goa/UNIPROT/gene_association.goa_uniprot.gz[117]. A number of identifier association files were created using the UNIPROT and GO information, including Contigs2GO.dat which associates each of the reference contigs with its GO ids and Contigs2UniProt.dat which associates each of the reference contigs with its UniProt IDs. Using the list of differentially expressed genes, a list of associated GO identifiers for each differentially expressed gene was created using the perl script createDEG_GO_Tables.pl. For each of the five comparisons, three files result, one for all differentially expressed genes; one for up-regulated DEGs, and one for down-regulated DEGs. A custom R script FindGOEnrichments.R was created which uses the Bioconductor package GOstats to find enriched GO categories for all comparisons[118, 119]. There are three files for each comparison, one for biological process (BP), one for molecular function (MF) and one for cellular component (CC).

A blast database of the contigs was created and known nucleotide sequences for *Pelargonium × hortorum* were downloaded from the NCBI nucleotide database

resulting in 194 sequences. The "Send To/ File/Format FASTA" link was

selected with a size range of 0 to 100,000 bp in order to remove the complete

chloroplast genome. The known sequences (stored in the file

NCBIsequences.fasta) were blasted against the database and 187 of the 194

known sequences from NCBI had matches to the assembled contigs, in most

cases, with matching inferred annotations. Of the seven that did not match, four

are microsatellite DNA sequences, which should not be matched since they are

not transcribed, while three are transcribed mRNAs. This indicates that the

assembly and annotation are rather robust, matching 187/190 known mRNAs

(98.4%).

Identification of Target sequences from RNA seq Database

For identification of target sequences, ACPs, KASs, TEs, PKSs (polyketide

synthase), and KCSs (keto-acylCoA synthase) from the *de novo* RNA seq

database, ncbi_blast_2.3.0+(1).dmg program was downloaded from the National

Center for Biotechnology Information (NCBI) website. All of the ESTs sequences

were saved individually as .fa files. Each .fa file was ran against 486,398 contigs

(obtained from Bioinformatics Core) using R search script (python operator

2.7.11), Linux commands (Terminal Mac OS X program) and downloaded blast

program. The output generated a list of contigs that match the sequence of

interest; these contigs were then incorporated into the Vector NTI express

(version 1.1.5) contig assembly program. Based on contig alignment and 99-

100% sequence similarity, the sequence of interest was identified from the RNA

seq database. This method helped in identification and verification of the EST

sequences and the identification of additional isoforms/paralogs of the fatty acid

biosynthesis genes missing from the EST database.

Venn Diagrams and Heatmaps

Pratek Genomics suite 6.6 was used to generate Venn Diagrams and Heatmaps.

Program Settings for Venn Diagram, read count was set to > 0 in at least 1

condition (trichome 18°C, bald pedicle 18°C, trichome 23°C and bald pedicle

23°C). Heatmaps were generated using hierarchical clustering based on

Euclidean distance as the similarity measure for genes of interest and gene's

associated with GOID: 0006631.

Statistical Analysis

Data for qRT-PCR were analyzed using paired t-tests to compare the null

hypothesis of a 1:1 expression of trichome to bald pedicle ratio. One-way

ANOVA with Tukey's multiple comparisons and correction test was used to

compare gene-to-gene expression levels. All of the remaining analyses were

done with R version 3.1.1(2014-07010). Mixed model analyses were done with

the lmer function of the lme4 package[120, 121]. Technical replicate effects were

modeled as random and nested within each Treatment*(biological replicate)

combination. The biological replicate value was modeled as a random effect and

crossed with the fixed treatment effects. Model residuals were found in all cases

to satisfy the assumptions of normality and homogeneity of variance. Tukeys

multiple comparison contrasts were estimated and tested with the glht function in the multcomp package of R with the single step method used for adjusting p-values[122]. False discovery rate p-values (FDRp) were computed with the p.adjust function of R. Correlation analysis was performed with the cor.test function of R. Q-values were computed with the qvalue function of the bioconductor package using the default settings[123-126].

Results and Discussion

Differential expression analysis from the *de novo* transcriptome assembly

The *de novo* RNA transcriptome generated for *Pelargonium × hortorum* resulted in a total of 486,398 transcript contigs. Normalized read counts for each contig (greater than 0) were not equally distributed in four physiological conditions used to generate the database (Figure 3.1). Common sets of genes representing 49% of the total identified were found in all samples (tissue and temperature treatments) (Figure 3.1). Overall analysis of the differentially expressed genes (DEGs) showed 58-59% of genes are up-regulated at 23°C compared to 18°C in both trichomes and bald pedicles, indicating an overall temperature effect. Trichome genes (51-52%) were up-regulated compared to bald pedicle at both temperatures, indicating tissue effect (Table 3.1).

The top 10 DEGs identified in the RNA-SEQ database revealed an up-regulation of fatty acid desaturase (FAD) in trichome tissue at 23°C as compared to trichome tissue at 18°C and up-regulation of stearoyl-ACP desaturase (SAD), acyl-ACP thioesterase (TE) and 2 acyl-ACP reductases in trichomes as

compared to bald pedicle at 18°C (Table 3.2). Interestingly, none of the up-regulated FAD, SAD, TE or acyl-ACP reductases aligned with EST database sequences, indicating that these fatty acid genes are new isoforms or paralogs of fatty acid biosynthesis genes that were not identified in the EST database. These genes are strong candidates for studying temperature and tissue effect because they rank in top the 10 categories of up-regulated genes as compared to a total of 348 up-regulated genes in trichome 23°C versus trichome 18°C and 9399 up-regulated genes at trichome 18°C versus bald pedicle 18°C (Table 3.1). No fatty acid biosynthesis genes were identified in top 10 of the down-regulated genes (3 different comparisons) or up-regulated at trichome 23°C versus bald pedicle 23°C (Table 3.2). In order to evaluate differential expression for specific fatty acid biosynthesis genes (genes of interest), they need to be extracted from the RNA database for further analysis.
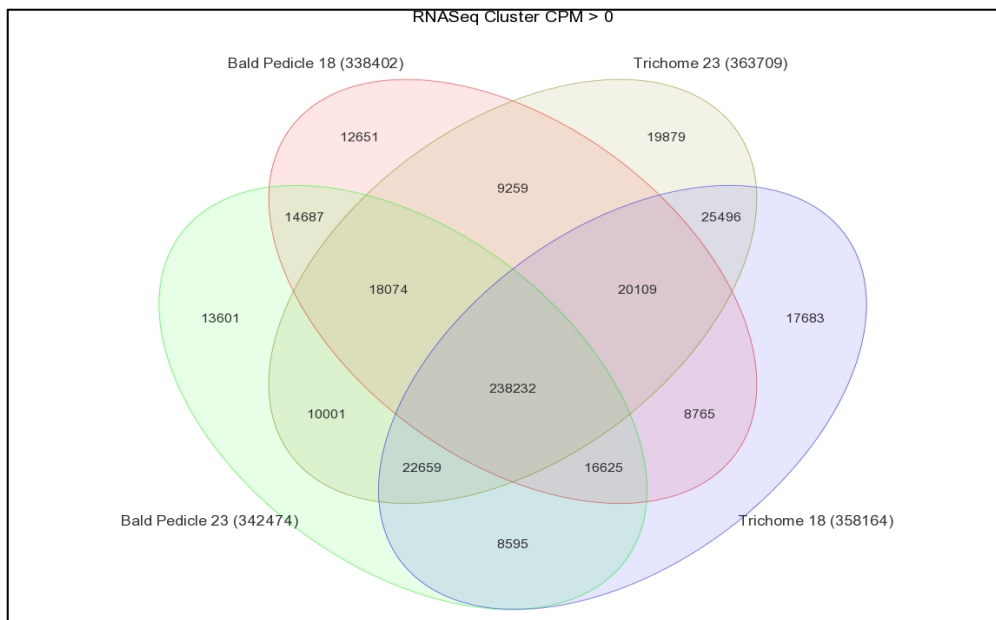
Figure 3.1. Venn diagram for the *de novo* RNA transcriptome distributed in each tissue and temperature treatment. Normalized read count was set to > 0 for each condition. The number of contigs in

a given condition is in parentheses.

Table 3.1. Differentially expressed genes (DEGs) of *Pelargonium × hortorum de novo* RNA transcriptome.

Comparsions	Number of DEGs	Number up regulated	Number down regulated
Trichome 23°C vs. Trichome 18°C	605	348	257
Bald pedicle 23°C vs. Bald Pedicle 18°C	445	260	185
Trichome 18°C vs. Bald Pedicle 18°C	18257	9399	8858
Trichome 23°C vs. Bald Pedicle 23°C	28274	14662	13612

Table 3.2. Top 10 Differentially expressed genes of *Pelargonium × hortorum de novo* RNA transcriptome.

Comparison	Gene Description (top 10 up regulated genes)	Gene Description (top 10 down regulated genes)
Trichome (23°C vs. 18°C)	Ribosomal protein s14	vacuolar-sorting receptor 6-like
	Kda class i heat shock	Uncharacterized protein isoform 1
	Orf114a gene product	26s protease regulatory subunit 8 homolog a
	Cucumisin-like	hypothetical chloroplast rf2
	cyclopropane-fatty-acyl-phospholipid synthase	phosphatidylinositol phosphatidylcholine transfer protein sfh9
	bag family molecular chaperone regulator 6	family transposase isoform 2
	fatty acid desaturase chloroplastic	exocyst complex component sec10
	probable anion transporter 5	serine carboxypeptidase-like
	non-functional nadph-dependent codeinone reductase 2-like	19-like isoform 1
	hypothetical protein CICLE_v10032920mg	NA
18°C (Trichome vs. Bald Pedicle)	cytochrome p450 82c4-like	atp synthase cf1 alpha subunit
	acyl-acp thioesterase	ribosomal protein s14
	probable 1-deoxy-d-xylulose-5-phosphate synthase chloroplastic	photosystem ii cp47 chlorophyll apoprotein
	retrotransposon ty1-copia sub-class	glycosyl hydrolase superfamily protein isoform 1
	7-deoxyloganetin glucosyltransferase-like	cytochrome p450 82c4-like
	fatty acyl- reductase 2	hypothetical protein
	hypothetical protein POPTR_0001s35870g	remorin-like isoform x1
	ribosomal protein s12	Unknown
	fatty acyl- reductase 3	calcium-binding ef-hand family
	stearoyl acp desaturase 2	NA
23°C (Trichome vs. Bald Pedicle)	cytochrome p450 82c4-like	atp synthase cf1 alpha subunit
	probable 1-deoxy-d-xylulose-5-phosphate synthase chloroplastic	photosystem ii cp47 chlorophyll apoprotein
	enoyl reductase	calcium-binding ef-hand family
	hypothetical protein POPTR_0001s35870g	cytochrome p450 82c4-like
	retrotransposon ty1-copia sub-class	membrane family protein
	cytochrome p450	remorin-like isoform x1
	hypothetical chloroplast rf19	photosystem ii cp43 partial
	zinc finger ccch domain-containing protein 20-like	probable 1-deoxy-d-xylulose-5-phosphate synthase chloroplastic
	non-ltr retroelement reverse transcriptase	dna glycosylase superfamily protein isoform 1
	(-)-germacrene d synthase-like	hypothetical protein

Sequences encoding isforms of ACP, KAS, FAT, MAD, PKS and KCS identified in the EST database were identified in the RNA-seq database using the bioinformatics steps outlined in material and methods. The distribution of normalized read counts showed that most of the genes of interest were highly expressed in trichomes compared to bald pedicle with exception of FAT-B2, FAT-B3, and KAS II (Figure 3.2 and Table 3.3). Complete sequences of FAT-B2, FAT-B3, KAS II and KAS Id that were incomplete in the EST database were obtained from RNA-Seq database (Appendix 6, Table A6.1). Due to the lack of reference genome, a gene with multiple variants or incomplete duplicates with few base pair errors are included in the heat map (Figure 3.2). Only complete contigs representing the genes of interest are considered for differential expression analysis (Table 3.3). There was an insignificant trend for an increase in fold change at 23°C as compared to 18°C for all the genes except KAS II and PKS I that showed an insignificant trend for decrease in fold change at 23°C as compared to 18°C (Table 3.3). The expression levels of these genes require validation via quantitative real-time PCR (qRT-PCR), which is more sensitive technique and targets the expression of specific gene[127, 128]. Additionally, *de novo* transcriptome assembly can have low statistical power for the less abundant genes (genes that are not highly expressed) and their changes in expression thus need to be validated by qRT-PCR [106, 127, 128]. KAS III and FAT-A3 genes identified in the EST database were not detected in the RNA transcriptome. A possible explanation for this may be lower abundance of these

genes and *de novo* transcriptome has detection constraints in identifying low

abundant genes[129].
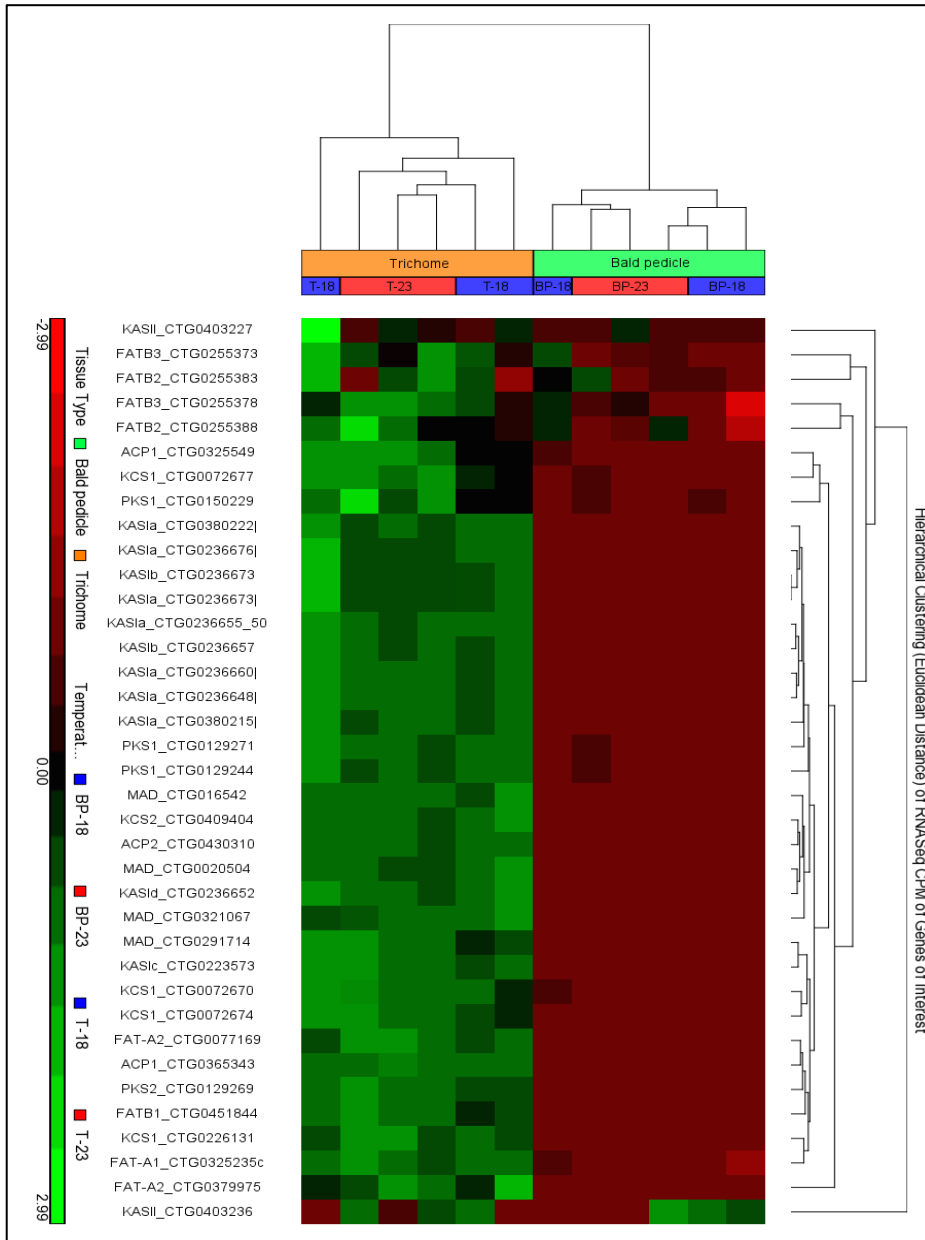
Figure 3.2. Heat map for genes of interests based on normalized read counts (RNA-Seq). Columns correspond to trichome 18°C, trichome 23°C, bald pedicle 18°C and bald pedicle 23°C. Each lane in the column corresponds to a biological replicate (3 replicates per sample). Each row corresponds to the gene of interest. Green to red indicates a continuum of high to low expression. The dendrograms on top and to the right were obtained via hierarchical clustering of a pair-wise Euclidean distance matrix. Genes, acyl carrier protein (ACP), β ketoacyl-ACP synthase (KAS), fatty acid acyl-ACP thioesterase (FAT), Δ^9 Myristoyl ACP desaturase (MAD), polyketide synthase (PKS) and 3-Ketoacyl-CoA Synthase (KCS).

Table 3.3. Fold change and FDR p values for genes of interest (RNA-seq)

Gene Identified	Length (bp)	Complete	Trichome vs. Bald Pedicle 18°C		Trichome vs. Bald Pedicle 23°C		Trichome 23C° vs. Trichome 18°C	
			Fold Change	FDR	Fold Change	FDR	Fold Change	FDR
MAD_CTG0165429	1271	YES	5.90	3.43E-18	6.17	1.61E-18	0.15	1
ACP 1_CTG0365343	640	YES	2.88	2.40E-07	3.05	1.34E-12	0.28	1
ACP 2_CTG0430310	745	YES	3.93	1.23E-11	4.31	1.10E-18	0.02	1
FAT-A1_CTG0325235c	1682	YES	2.17	3.00E-03	2.56	1.05E-07	0.19	1
FAT-A2_CTG0077169	1599	810/1599	5.66	1.02E-18	6.05	2.90E-19	0.47	1
FAT-A3	1596	Not found	NA	NA	NA	NA	NA	NA
FAT-B1_CTG0451844	1376	YES	6.37	4.08E-16	6.60	4.17E-26	0.61	1
FAT-B2_CTG0255388	**1327**	**YES**	**0.44**	**1.00**	**0.89**	**0.33**	**0.58**	**1**
FAT-B3_CTG0255378	**1977**	**YES**	**0.58**	**0.89**	**0.96**	**0.13**	**0.56**	**1**
KAS Ia_CTG0236648	1905	312/1905	5.47	9.89E-19	6.18	2.91E-28	-0.01	1
KAS Ib_CTG0236657	1871	YES	5.45	1.38E-13	6.21	1.34E-26	0.04	1
KAS Ic_CTG0223573	1645	YES	4.10	4.01E-10	4.57	3.37E-19	0.29	1
KAS Id_CTG0236655_5	**2038**	**YES**	**5.71**	**2.01E-18**	**6.27**	**4.01E-23**	**0.05**	**1**
KAS II_CTG0403227	**2348**	**YES**	**6.75**	**0.1**	**0.56**	**1**	**-1.34**	**1**
KAS III	1753	Not Found	NA	NA	NA	NA	NA	NA
PKS 1_CTG0129244	1406	302/1406	2.25	1.77E-04	1.72	1.10E-04	-0.17	1
PKS 2_CTG0129269	1399	YES	2.43	3.00E-05	2.76	3.23E-10	0.32	1
KCS 1_CTG0072674	1843	YES	2.84	1.40E-05	3.45	1.12E-16	0.42	1
KCS 2_CTG0409404	1546	YES	5.91	1.56E-18	6.05	2.59E-17	0.01	1
KCS 3_CTG0150219	1775	YES	3.47	1.37E-06	4.36	7.70E-20	0.44	1
KCS 4_CTG0222066	1419	YES	2.99	7.72E-06	3.47	4.41E-13	0.52	1
KCS 5_CTG0237072	1401	YES	5.02	6.01E-14	5.44	1.06E-25	0.34	1

Genes of Interest (from the EST database) that were identified in RNA-seq database, base pair (bp) length, fold changes and false discovery rate (FDR) p-values for each comparisons, trichome versus bald pedicle at 18°C, trichome versus bald pedicle at 23°C, trichome 23°C versus trichome 18°C. Bold Highlight - Complete sequence information obtained for sequences that were incomplete in the EST database.

Identification of EST database genes from the RNA-SEQ transcriptome led to identification of additional variants of fatty acid biosynthesis genes, 3 FAT-As , 4 FAT-Bs, 2 KAS Is, 1PKS and 3 KCSs sequences (Table 3.4). With *de novo* transcriptome assembly it is difficult to predict if these distinct sequences are isoforms or paralogs due to absence of reference genome for geranium. Thus,

these sequences are indicated only as gene variants in this study[82, 106, 109, 130-132].

Most of these genes were highly expressed in trichomes at both temperatures.

FAT-A4 shows a significant increase in expression at 23°C (FDR=0.01) whereas

other genes show insignificant trends for trichome 23°C versus trichome 18°C

comparison (Table 3.4).

Table 3.4. Fold change and FDR p-values for fatty acid biosynthesis genes
variants (RNA-Seq).

Gene Identified	Length(bp)	Trichome vs. Bald Pedicle 18°C		Trichome vs. Bald Pedicle 23°C		Trichome 23C° vs. Trichome 18°C	
		Fold Change	FDR	Fold Change	FDR	Fold Change	FDR
FAT-A4_CTG0380018	3596	3.75	4.06E-07	5.66	9.65E-25	1.77	**0.01**
FAT-A5_CTG0380020	3632	14.69	1.05E-38	13.63	4.96E-38	-1.27	0.52
FAT-A6_CTG0077168	1376	5.08	3.74E-07	6.46	4.88E-27	0.15	1
FAT-A7_CTG0380021	2288	6.03	1.69E-19	6.21	3.31E-20	0.43	1
FAT-B4_CTG0255373	2605	1.57	0.812	2.33	0.017	0.13	1
FAT-B5_CTG0451842	3658	7.39	2.67E-12	8.38	3.97-13	0.38	1
FAT-B6_CTG0388939	1824	7.07	1.40E-20	6.70	3.21E-10	-1.51	0.18
FAT-B7_CTG0255372	1308	-1.79	0.88	1.84	0.63	1.41	1
KAS Ie_CTG0236652	1255	5.89	3.72E-19	6.11	1.18E-25	-0.15	1
KAS If_CTG0209438_3	2632	0.05	1	0.20	0.95	0.02	1
KAS II(Variant1)_CTG0178043_3	1277	4.68	0.42	0.68	0.98	0.55	1
KAS II(Variant2)_CTG0178069	1842	0.99	0.71	1.36	0.21	0.04	1
KAS II(variant3)_CTG0403242	2348	7.27	0.21	-4.40	0.71	-7.52	0.92
PKS 3_CTG0129267	2813	2.22	1.00E-03	2.21	1.20E-05	0.03	1
KSC 6_CTG0072671	2392	3.40	5.36E-03	2.74	7.50E-05	0.23	1
KCS 7_CTG0072674	1956	2.84	1.30E-05	3.45	1.11E-16	0.42	1
KCS 8_CTG0129477	1484	4.96	8.53E-14	5.43	2.10E-24	0.27	1

New gene variants identified, sequence length, fold changes and FDR p-values for
comparisons of trichome versus bald pedicle at 18°C, trichome versus bald pedicle at
23°C and trichome 23°C versus 18°C. Significant FDR p-value for trichome 23°C versus
trichome 18°C comparison indicated in bold.

A total of 471 genes were identified from the RNA-seq database using Gene

ontology ID: 0006631 associated with metabolic process (Figure 3.3). Unlike the

heatmap for selected EST genes, this heat map reports genes that are highly

expressed not only in trichomes but also in bald pedicle and it also shows

temperature effect for few genes (Figure 3.3). Normalized read counts for each

contig (greater than 0) were not equally distributed in the four samples, (two

tissues, two temperature treatments) used to generate the database (Figure 3.4).

In all four samples 60% (281/471) of the genes were present whereas the

remaining genes were found only in a specific tissue type or at specific

temperatures combinations. Genes that were only present in trichomes at both

temperatures (35/471) were selected for further evaluation and their identity

(annotation) was determined based on BLASTX search engine version 2.5.1+

(Table 3.5).

The differential analysis of these genes did not show an effect of temperature

(FDR=1). Nevertheless, 6 different variants of fatty acid hydroxylase (used in

synthesis of unusual hydroxy fatty acids in some plants)[133] and 2 variants of KAS

I were identified (Table 3.5).
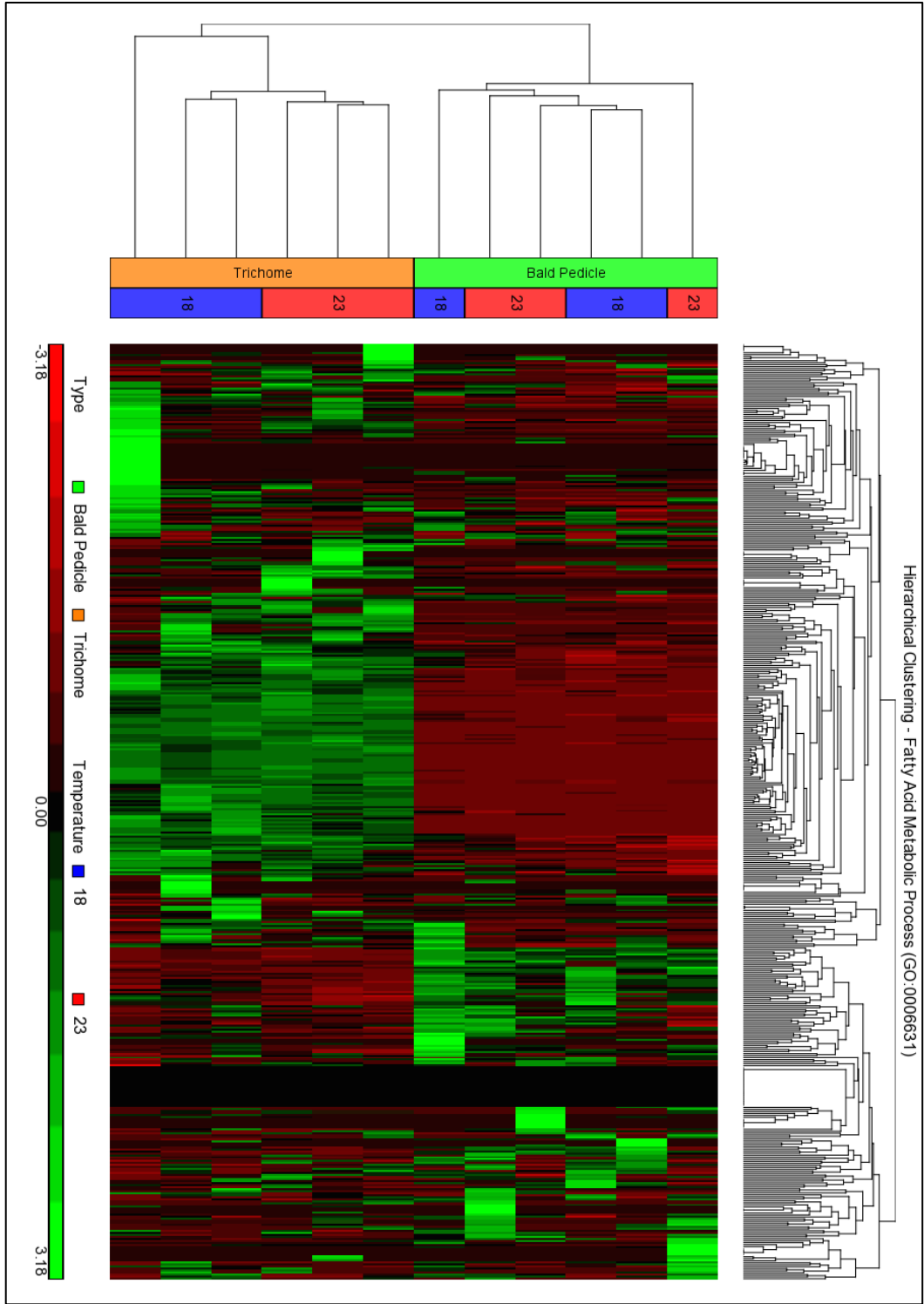
Figure3.3. Heatmap for genes with gene ontology ID: GO:0006631 associated with fatty acid metabolic processes. The heatmap is based on normalized read counts. Columns correspond to trichome 18°C, trichome 23°C, bald pedicle18°C, bald pedicle 23°C. Each lane in the column corresponds to a biological replicate (3 replicates per sample). Each row corresponds to a gene. Green to Red indicates a continuum of high to low expression. The dendrograms on top and to the right were obtained via hierarchical

clustering of a pair-wise Euclidean distance matrix.
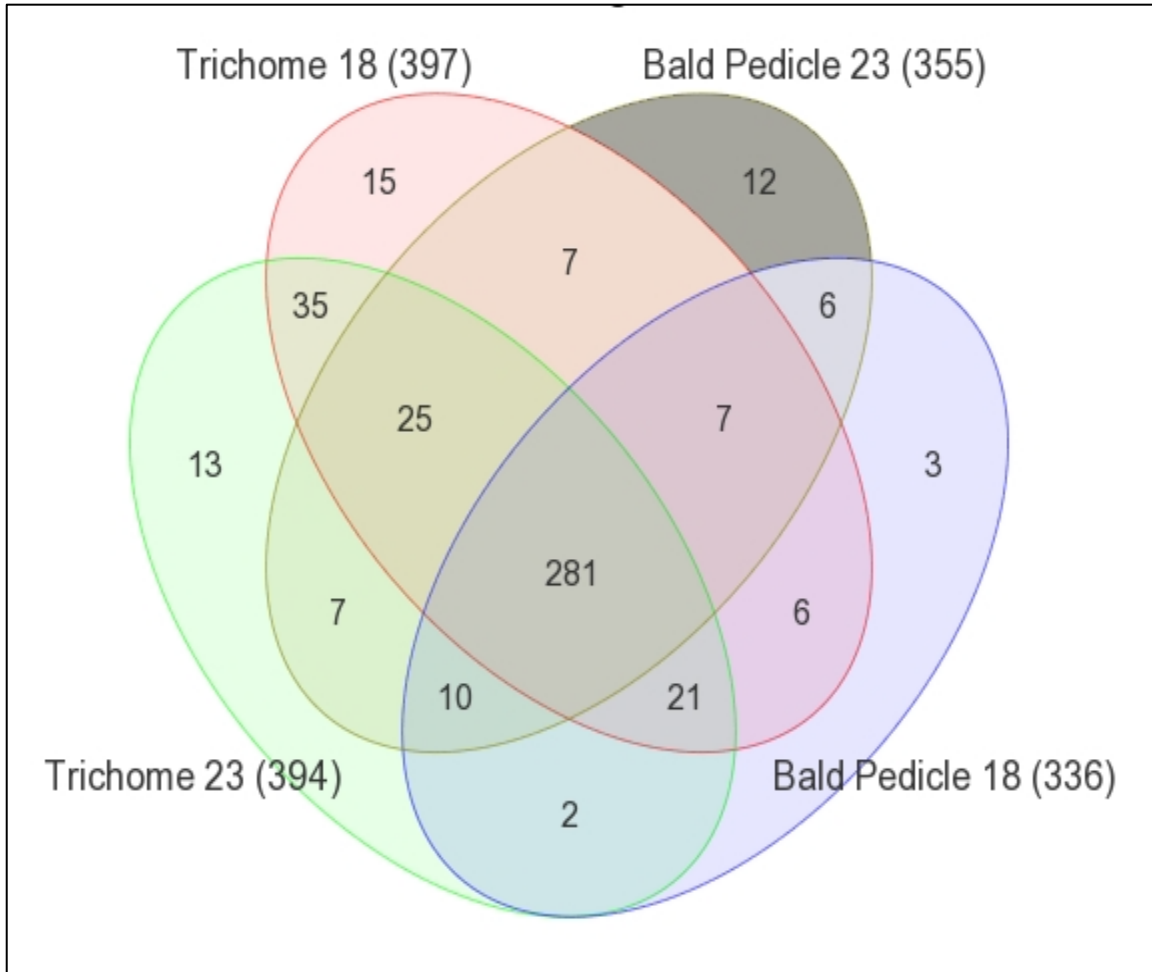
Figure 3.4. Venn diagram for RNA contigs within gene ontology ID: GO:0006631 associated with fatty acid metabolic processes. The numbers indicates contigs that are common between all samples (trichome at 18°C, trichome at 23°C, bald pedicle at 18°C and bald pedicle at 23°C), numbers indicating overlapping areas between the samples and numbers unique to a sample. Normalized Read count is set to > 0 for each condition. The number of contigs in a given sample is in parentheses.

Table 3.5. Identity of 35 genes associated with fatty acid metabolic processes (ID:GO:0006631) found exclusively in trichomes at both temperatures.

Sequence #	Blast X results for gene identiy	% Match	Plant	Trichome (23C° vs. 18°C) Fold Change	FDR
CTG0013871	stearoyl-[ACP] 9-desaturase, chloroplastic	88%	Solanum tuberosum	0.07	1
CTG0022266	stearoyl-ACP-desaturase	87%	Theobroma cacao	0.50	1
CTG0038053	GNS1/SUR4 membrane protein	76%	Cynara cardunculus var. scolymus	1.60	1
CTG0129631	Acyl-[ACP] desaturase	92%	Gossypium arboreum	0.03	1
CTG0133638	3-oxoacyl-[ACP] synthase I	83%	Triticum urartu	0.09	1
CTG0157307	stearoyl-ACP 9-desaturase 6	86%	Ricinus communis	0.76	1
CTG0157309	Fatty acid desaturase, type 2	87%	Cynara cardunculus var. scolymus	0.18	1
CTG0158156	Chloroplast J-like domain 1	68%	Theobroma cacao	-0.01	1
CTG0171331	Fatty acid hydroxylase 1 isoform 2	81%	Theobroma cacao	0.92	1
CTG0171333	Fatty acid hydroxylase 1 isoform 2	62%	Theobroma cacao	-0.01	1
CTG0171335	AMP-binding, conserved site-containing protein	72%	Cynara cardunculus var. scolymus	1.86	1
CTG0184753	AMP-binding, conserved site-containing protein	72%	Cynara cardunculus var. scolymus	-1.79	1
CTG0231144	stearoyl ACP desaturase 02	96%	Pistacia chinensis	0.04	1
CTG0236653	beta-ketoacyl-[ACP] synthase I	93%	Arabidopsis thaliana	0.60	1
CTG0236668	beta-ketoacyl-[ACP] synthase I	91%	Arabidopsis thaliana	-1.41	1
CTG0254781	stearoyl acyl carrier protein desaturase	62%	Cocos nucifera	0.10	1
CTG0254795	myristyl-ACP desaturase	70%	Pelargonium × hortorum	0.20	1
CTG0262387	acyl-CoA oxidase family protein	72%	Populus trichocarpa	-0.06	1
CTG0262392	acyl-CoA oxidase ACX3	66%	Arabidopsis thaliana	-0.06	1
CTG0262395	acyl-CoA oxidase 3	74%	Prunus persica	-0.47	1
CTG0262402	acyl-CoA oxidase 3	77%	Prunus persica	-0.21	1
CTG0262409	acyl-CoA oxidase family protein	79%	Populus trichocarpa	-0.10	1
CTG0262414	acyl-CoA oxidase, putative	82%	Ricinus communis	-0.49	1
CTG0262415	acyl-CoA oxidase 3	83%	Prunus persica	-0.09	1
CTG0273427	Fatty acid 2-hydroxylase	46%	Zostera marina	-0.18	1
CTG0285223	Fatty acid 2-hydroxylase	70%	Glycine soja	0.80	1
CTG0285224	Fatty acid 2-hydroxylase	70%	Glycine soja	-0.31	1
CTG0285225	fatty acid 2-hydroxylase 2-like	76%	Dorcoceras hygrometricum	1.28	1
CTG0295821	3-oxoacyl-[acyl-carrier-protein] synthase, KASI	100%	Escherichia coli IS5	-1.82	1
CTG0376554	Stearoyl-ACP Desaturase	72%	Salvia miltiorrhiza	-0.50	1
CTG0397211	myristyl-ACP desaturase	91%	Pelargonium × hortorum	1.55	1
CTG0405943	Acyl-ACP thioesterase	72%	Helianthus annuus	1.05	1
CTG0430616	myristyl-ACP desaturase	73%	Pelargonium × hortorum	-0.08	1
CTG0442248	acyl-CoA oxidase family protein	79%	Populus trichocarpa	0.24	1
CTG0465399	beta-ketoacyl-[acyl carrier protein] synthase I	92%	Arabidopsis thaliana	-0.64	1

BlastX 2.5.1+ output (% Identity with other plant species).

Effect of temperature on gene expression, anacardic acid production, and UMFA biosynthesis.

SYBR green and TaqMan assays were used to study gene expression of ACPs, KASs and FAT-As as described in Chapter 2. The assays were repeated for three different temperatures 18°C, 23°C and 28°C. Temperature did not affect the higher expression for any of the tested genes (MAD, ACP 1, ACP 2, FAT-A1, FAT-A2, KAS I-a/b, KAS I-d and KAS III) in trichomes compared to bald pedicle (Figure 3.5). This observation indicates that trichome specificity is unaffecetd by temperature. Within the trichome samples, expression of ACP 2 was significantly higher than ACP 1 at all temperatures. This further suggests the potential of ACP 2 as a novel ACP paralog involved in UMFA synthesis (Figure 3.6). Both ACP 1 and ACP 2 showed a significantly lower expression at 28°C as compared to 18°C and 23°C while their expression levels was not significantly different at 18°C and 23°C (Figure 3.6, Table 3.5).

Trichome expression of FAT-A2 was significantly higher than FAT-A1 only at 28°C, whereas expression level remained the same at 18°C and 23°C (Figure 3.6). Overall expression of FAT-A1 was not affected by the temperature whereas FAT-A2 showed significant increase in expression at 28°C compared to 18°C and 23°C (Figure 3.6, Table 3.5).

Trichome expression of KAS I-a/b was significantly higher than KAS I-c and KAS III at all temperatures further suggesting the potential of KAS I-a/b as a novel KAS paralog potentially involved in UMFA synthesis (Figure 3.6). Both KAS I-a/b and KAS Ic showed a significant decrease in expression at 28°C as compared to 18°C and 23°C whereas KAS III expression was not affected by temperature (Figure 3.6, Table 3.5). Expression levels of all the KASs in this study were not significantly different at 18°C compared to 23°C (Figure 3.6, Table 3.5). Overall results suggested that temperature affects the expression of ACPs, KASs and TEs. To identify specific fatty acid genes involved in UMFA synthesis, respective expressions patterns at distinct temperatures should be correlated with changes in production of UMFAs and AnAc.
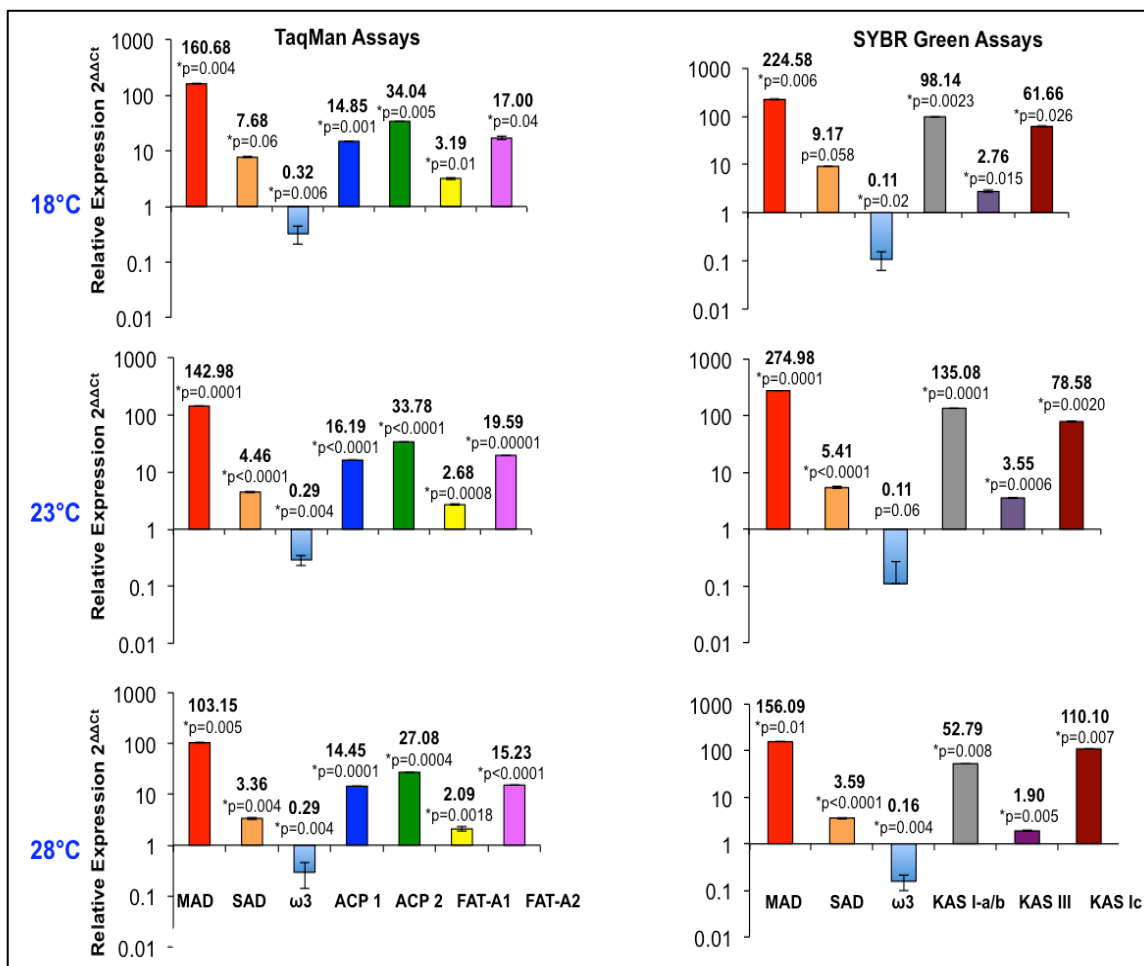
Figure 3.5. Relative expression of selected genes in trichome compared to bald pedicle at 18°C, 23°C and 28°C. MAD (Δ9 myristyl ACP desaturase), SAD (Δ9 stearoyl-ACP desaturase) and ω3 (omega-3 desaturase) were used as controls in both assays. TaqMan assay was used for ACPs (acyl carrier proteins) and FAT-As (fatty acid thioesterase). SYBR green assay was used for KAS (β-ketoacyl-ACPsynthase). Y-axis shows average fold change values. Based on 1:1 expression of trichome to bald pedicle, fold change values above 1 suggests higher expression in trichomes and below 1 suggest higher expression in bald pedicle. Bars represent the standard error of means and the means are provided at end of each bar. p-values of the t-test are shown for each gene, asterisk before p-value indicates significantly different expression from 1:1 ratio.
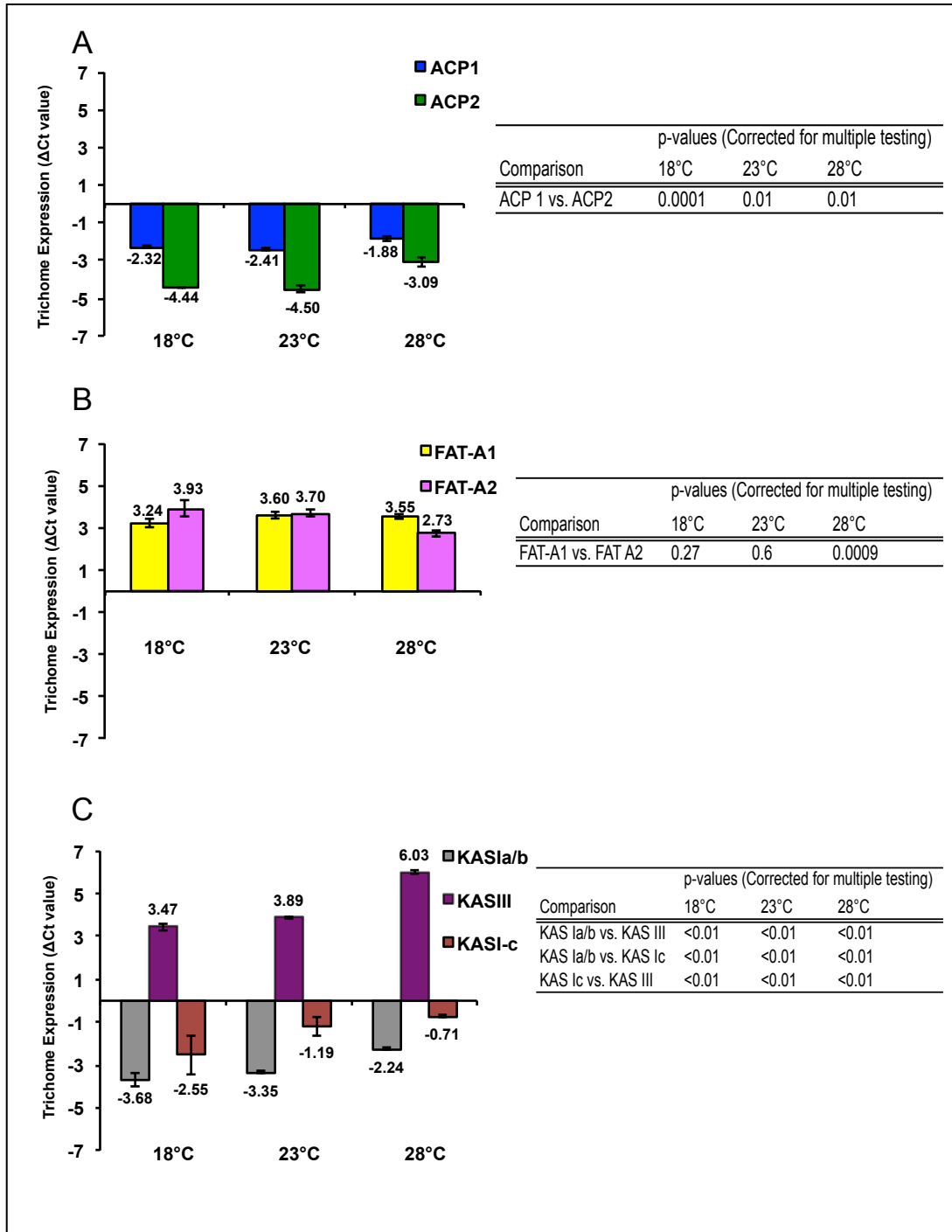
Figure 3.6. Comparison of expression between selected fatty acid biosynthesis genes in trichome tissues at 18°C, 23°C and 28°C. ACPs (Acyl carrier proteins), FAT-As(fatty acid acyl-ACP thioesterase) and KAS (β-ketoacyl-ACP synthase). Y-axis shows the average delta cycle threshold (Ct) values. Bars represent the standard error of means and the means arerepresented by black numerals at end of each bar. The p-values (corrected for multiple comparisons) in the tables next to graph show comparison between two genes. Values <0.05 are highly significant. Lower delta Ct values indicate higher expression.

Table 3.6. FDR p-values for testing effect of temperature on target genes.

Genes	Assay	Temperature Effect	18°C vs. 23°C	18C° vs. 28°C	23°C vs. 28°C
		FDR-pvalues			
MAD	Taq-Man	0.37	0.60	1.00	0.55
SAD	Taq-Man	6.00E-03	0.84	0.04	0.05
ω3	Taq-Man	4.00E-04	0.83	0.05	0.02
ACP 1	Taq-Man	6.00E-04	0.86	0.08	0.03
ACP 2	Taq-Man	3.00E-04	0.98	0.03	0.02
FAT-A1	Taq-Man	0.16	0.28	0.37	0.97
FAT-A2	Taq-Man	4.00E-03	0.76	0.03	0.04
MAD	SYBR Green	6.00E-04	0.31	0.03	0.12
SAD	SYBR Green	3.00E-04	0.92	0.02	0.02
ω3	SYBR Green	2.00E-04	0.89	0.03	0.01
KAS Ia/b	SYBR Green	2.-0E05	0.32	3.00E-05	6.00E-04
KAS Ic	SYBR Green	1.05E-06	0.06	1.79E-06	5.26E-06
KAS III	SYBR Green	0.04	0.21	0.10	0.74

p-values <0.05 are highly significant

Within the trichomes the amount of AnAc 24:1n5 was significantly higher than AnAc 22:1n5 at all temperatures (Figure 3.7) and amount of $16:1^{\Delta 11}$ was significantly higher than $18:1^{\Delta 13}$ at all temperatures (Figure 3.8). Even though content of $16:1^{\Delta 11}$ is greater than $18:1^{\Delta 13}$ more AnAc 24:1n5 was accumulated indicating a possibility of substrate preference for $18:1^{\Delta 13}$ by the downstream enzymes for production of AnAc 24:1n5. Furthermore, ratios of UMFAs ($18:1^{\Delta 13}$ to $16:1^{\Delta 11}$) showed insignificant trend for an increase at 23°C and a decrease at 28°C whereas ratios of AnAc (24:1n5 to 22:1n5) showed insignificant trend for an increase at 23°C and significant trend for a increase at 28° (Figure 3.9).

AnAc congeners and their respective substrate UMFA's showed significant higher production at 23°C compared to 18°C and 28°C and a significant

decrease in production at 28°C compared to 18°C and 23°C (Figure 3.7 and

Figure 3.8). This suggests that 23°C is the optimal temperature for production of

UMFAs and AnAc as compared to 18°C and 28°C and higher temperature

negatively affects the production of both UMFAs and AnAc.

Anacardic Acid	Temperature Effect	18°C vs. 23°C	18C° vs. 28°C	23°C vs. 28°C
		FDR p-values		
22:1n5	4.70E-05	0.02	0.01	2.00E-04
24:1n5	1.32E-06	3.00E-05	2.00E-03	4.75E-07
		p-values		
		18°C	23°C	28°C
22:1n5 versus 24:1n5		4.22E-28	3.88E-25	6.84E-27
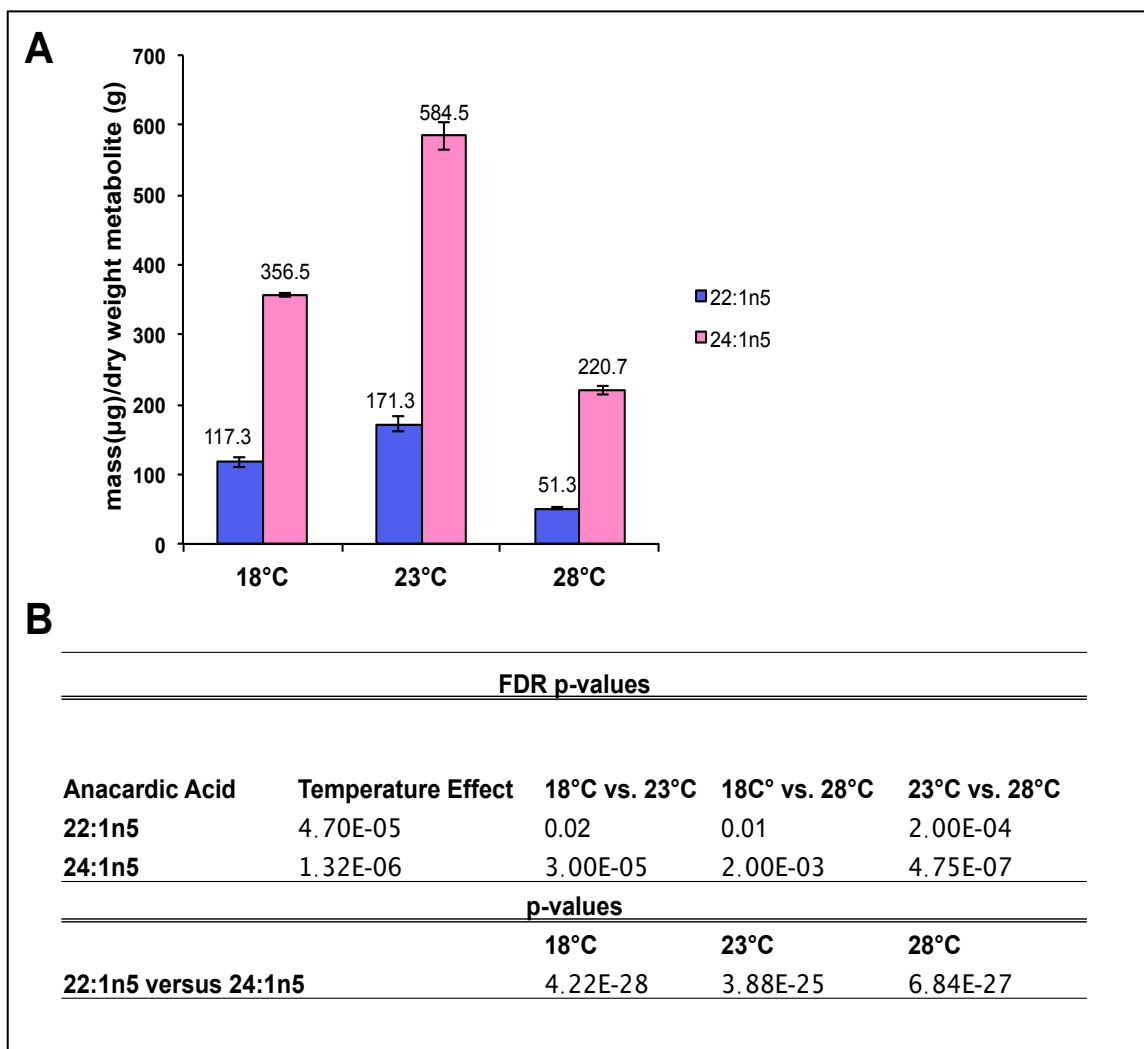
Figure 3.7. Production of n5 anacardic acids at 18°C, 23°C and 28°C. (A) HPLC
analysis of anacardic acid (22:1n5 and 24:1n5) production at distinct
temperatures.
(B) FDR p-values for temperature effects and individual comparisons between
all three temperatures and p-value comparison for 22:1n5 versus 24:1n5 . p-
values <0.05 are highly significant**.**

A

B

		FDR p-values		
UMFAs	Temperature Effect	18°C vs. 23°C	18C° vs. 28°C	23°C vs. 28°C
$16:1^{\Delta 11}$	7.19E-05	0.01	0.02	3.00E-04
$18:1^{\Delta 13}$	7.43E-06	3.00E-03	7.00E-03	8.52-05
		p-values		
		18°C	23°C	28°C
$16:1^{\Delta 11}$ versus $18:1^{\Delta 13}$		4.50E-04	2.00E-02	3.20E-05
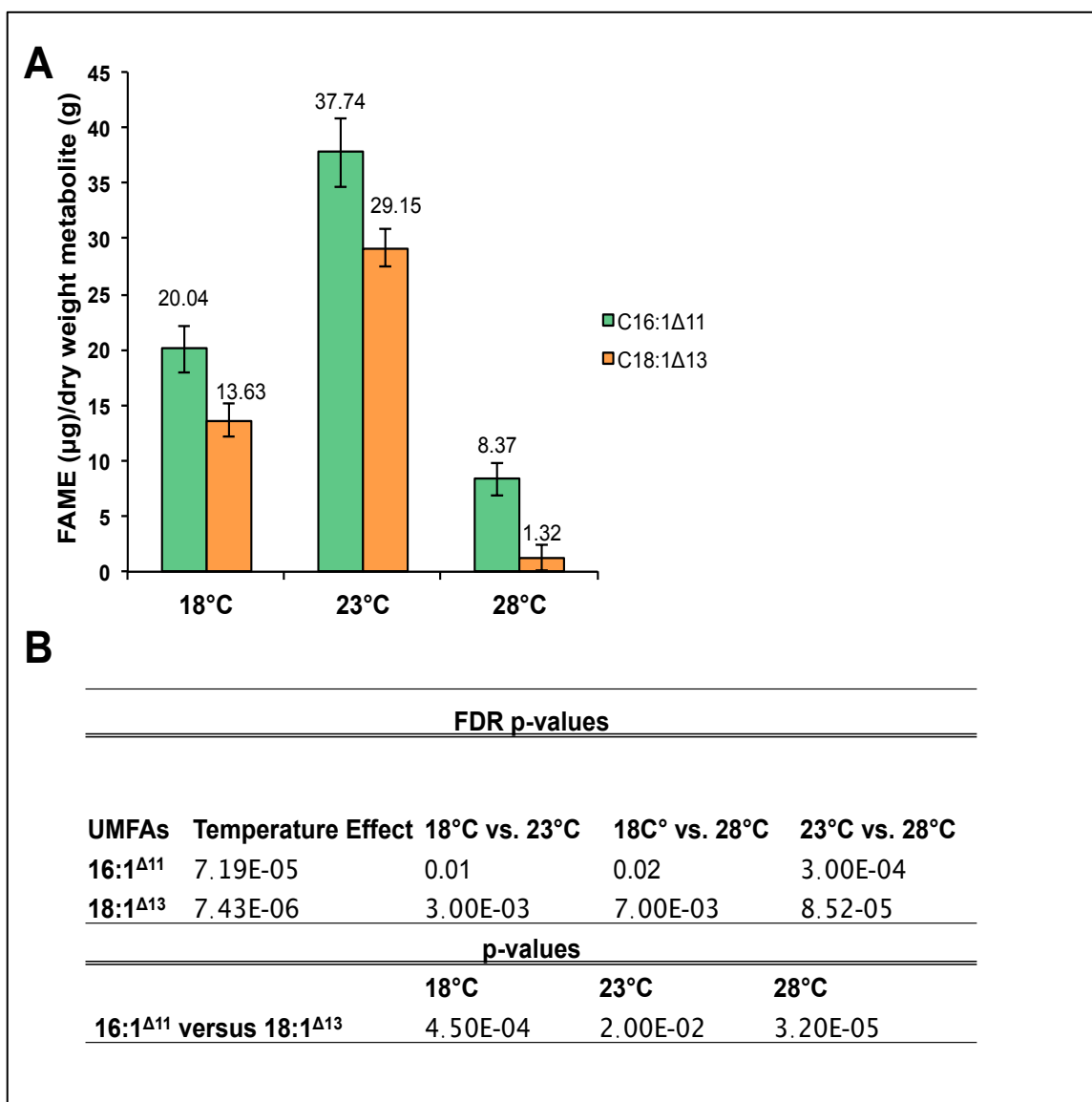
Figure 3.7. Production of UMFAs at 18°C, 23°C and 28°C. (A) GC analysis of UMFAs ($16:1^{\Delta 11}$ and $18:1^{\Delta 13}$) production at distinct temperatures. (B) FDR p-values for temperature effects and individual comparisons between all three temperatures and p-value comparison for UMFAs $16:1^{\Delta 11}$ and $18:1^{\Delta 13}$. p-values <0.05 are highly significant.
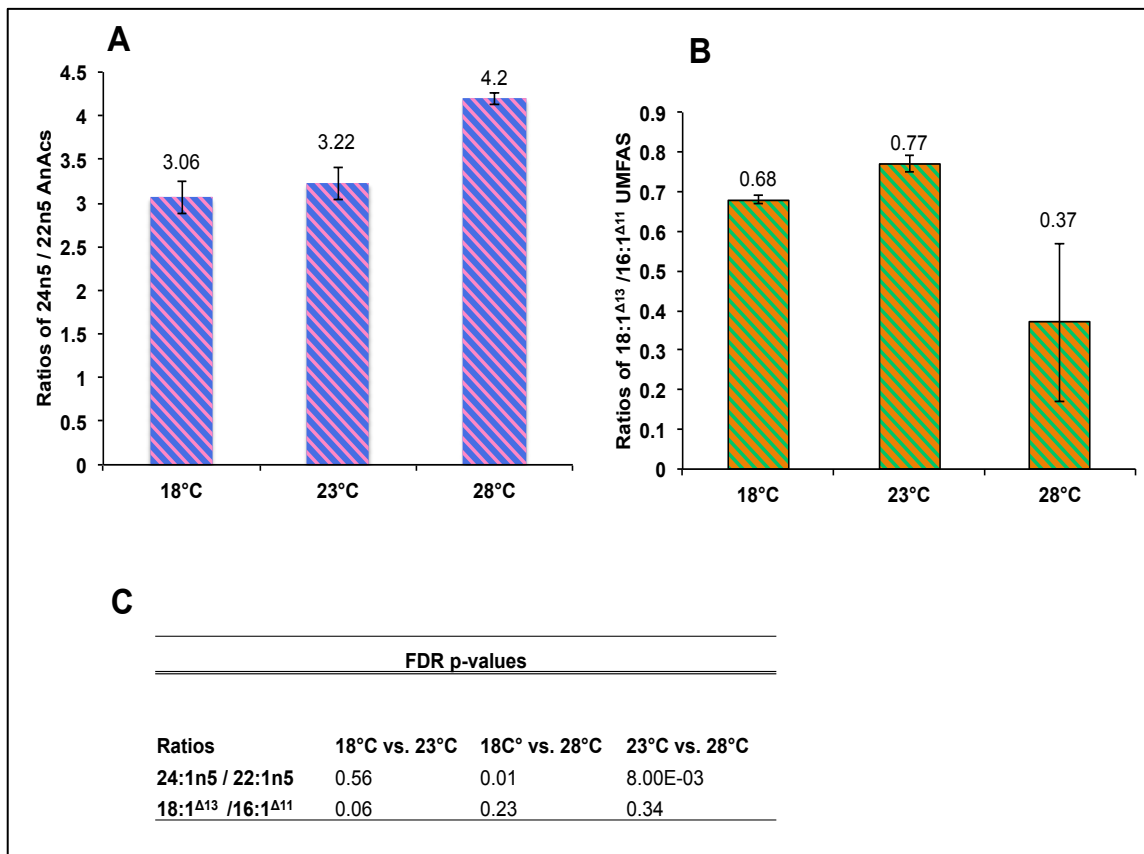
Figure 3.9. Ratios of n5 AnAc and UMFAs at 18°C, 23°C and 28°C. (A) Ratios of AnAc 24:1n5 to 22:1n5, (B) Ratios of UMFAs $18:1^{\Delta 13}$ to $16:1^{\Delta 11}$ and (C) p-values for individual ratio of AnAc and UMFA comparisons between all three temperatures. p-values <0.05 are highly significant.

The correlation analysis measures the quantities of AnAc 22:1n5, AnAc 24:1n5, UMFA $16:1^{\Delta 11}$ and UMFA $18:1^{\Delta 13}$ indicated all four are positively correlated at all three temperatures. Thus there is a a direct relationship between the amount of substrate and the derived metabolite at a given temperature (Figure 3.10). Expression levels of ACP 1 and ACP 2 were positively correlated to both AnAc (22:1n5, 24:1n5) and UMFAs ($16:1^{\Delta 11}$,$18:1^{\Delta 13}$) showing a similar trend of increase in relative expression at 23°C and a significant decrease at 28°C

(Figures 3.11 and 3.12). This suggests that both ACPs may be potentially involved in UMFA synthesis.

Deceread expression levels from 23°C to 28°C of KAS I-a/b shows a marginally significant correlation and KAS Ic shows a significant correlation with both AnAc (22:1n5, 24:1n5) and UMFAs ($16:1^{\Delta11}$,$18:1^{\Delta13}$), (Figures 3.13 and 3.14). The decreased expression level from 18°C to 23°C for both KAS I-a/b and KAS I-c was not significant and this change cannot be correlated with levels of both AnAc (22:1n5, 24:1n5) and UMFAs ($16:1^{\Delta11}$,$18:1^{\Delta13}$) at those temperatures (Figures 3.6, 3.13 and 3.14). KASs are condensing enzymes that add 2 carbon units to the growing acyl chain, so it was expected that if KASs expression decreases at a particular temperature then ratios of 18 to 16 carbon should also decrease at that temperature; however, the ratios of UMFAs (18 to 16) did not decrease at 23°C and only showed a significant decrease at 28°C (Figure 3.9). Based on this observation and for inference of better conclusions, expression of KASs at different temperature would require additional biological replicates.

Expression level for FAT-As was not correlated with levels of both AnAc (22:1n5, 24:1n5) and UMFAs ($16:1^{\Delta11}$,$18:1^{\Delta13}$) at any of the temperatures indicating that either temperature does not affect the expression levels of FAT-As or that FAT-A1 and FAT-A2 are not the target candidates involved in UMFA synthesis (Appendix 3, Table A3.2)
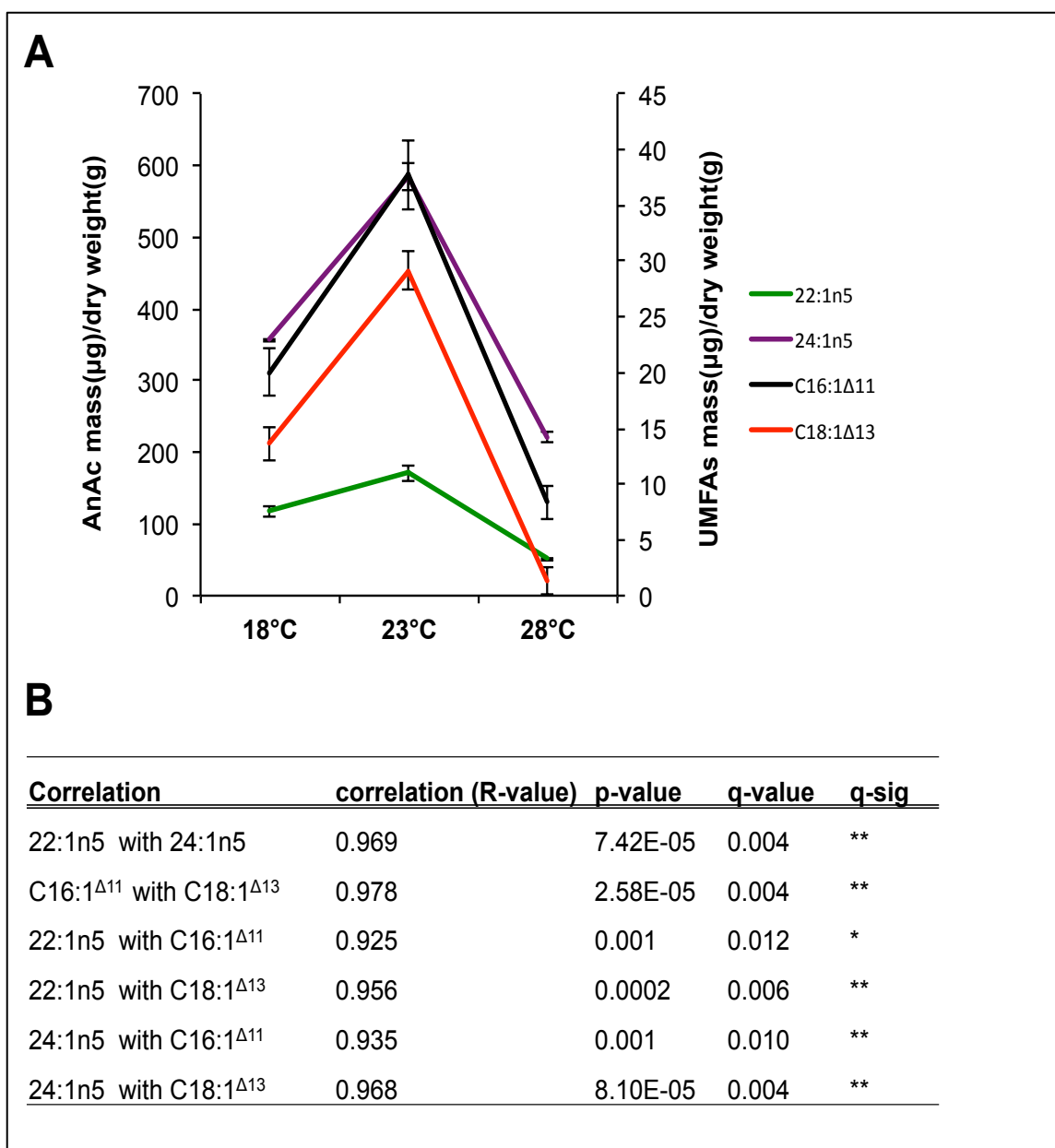
Figure 3.10. Correlations of UMFAs and anacardic acids production at distinct temperatures. (A) Correlations of AnAc 22:1n5, 24:1n5) and UMFAs ($16:1^{\Delta 11}, 18:1^{\Delta 13}$). (B) Corelation R-value for each comparison along with p-values and q-values. Mixed model correlation analysis was performed with the cor.test function of R to generate q-values for each comparison. q-value significance - Not significant (ns), Marginally significant (.), significant (*) and highly significant (**).
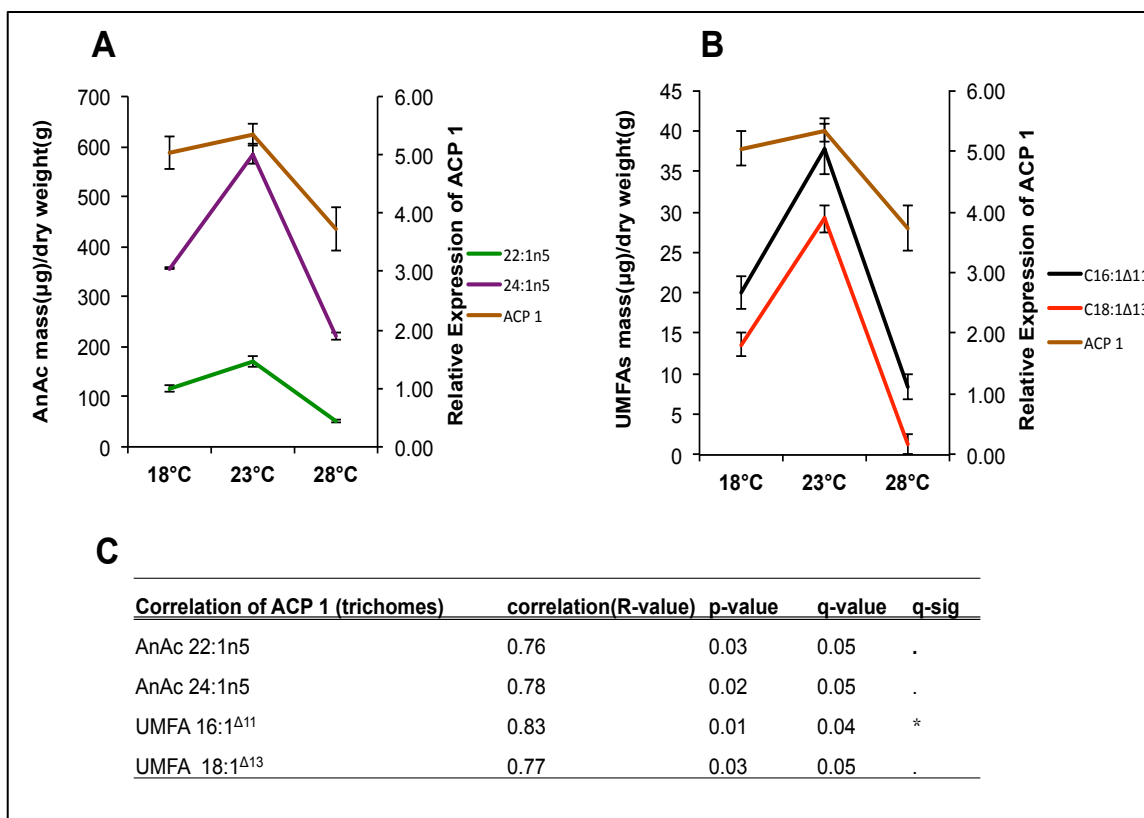
Figure 3.11. Correlations of UMFAs and anacardic acids production with ACP 1 expression at distinct temperatures. (A) Correlations of AnAc 22:1n5 and 24:1n5 with ACP 1. (B) Correlations of UMFAs 16:1$^{\Delta 11}$ and 18:1$^{\Delta 13}$ with ACP 1. (C) Corelation R-value for each comparison along with p-values and q-values. Mixed model correlation analysis was performed with the cor.test function of R to generate q-values for each comparison. q-value significance - Not significant (ns), Marginally significant (.), significant (*) and highly significant (**).

Table C:

Correlation of ACP 1 (trichomes)	correlation(R-value)	p-value	q-value	q-sig
AnAc 22:1n5	0.76	0.03	0.05	.
AnAc 24:1n5	0.78	0.02	0.05	.
UMFA 16:1$^{\Delta 11}$	0.83	0.01	0.04	*
UMFA 18:1$^{\Delta 13}$	0.77	0.03	0.05	.
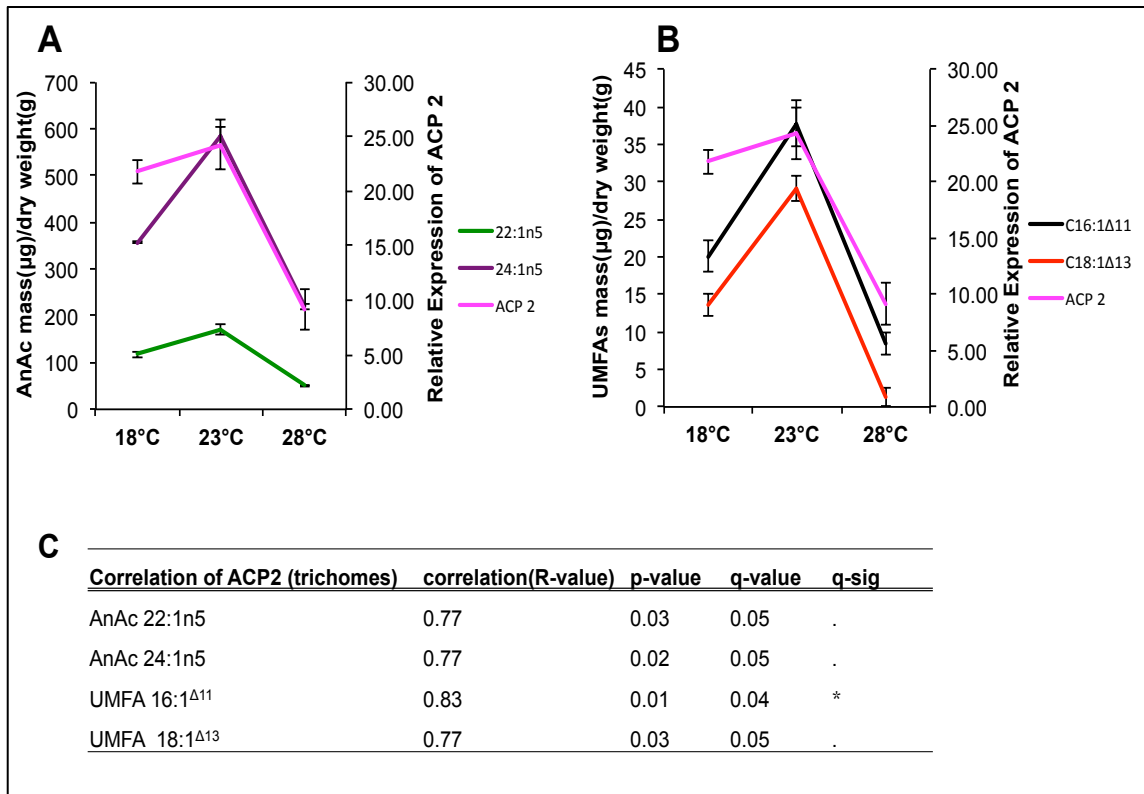
Figure 3.12. Correlations of UMFAs and anacardic acids production with ACP 2 expression at distinct temperatures. (A) Correlations of AnAc 22:1n5 and 24:1n5 with ACP 2. (B) Correlations of UMFAs $16:1^{\Delta 11}$ and $18:1^{\Delta 13}$ with ACP 2. (C) Corelation R-value for each comparison along with p-values and q-values. Mixed model correlation analysis was performed with the cor.test function of R to generate q-values for each comparison. q-value significance - Not significant (ns), Marginally significant (.), significant (*) and highly significant (**).

A

B

C

Correlation of KAS Ia/b (trichomes)	correlation(R-value)	p-value	q-value	q-sig
AnAc 22:1n5	0.76	0.03	0.059	.
AnAc 24:1n5	0.67	0.07	0.093	.
UMFA 16:1$^{\Delta 11}$	0.69	0.06	0.088	.
UMFA 18:1$^{\Delta 13}$	0.67	0.07	0.092	.
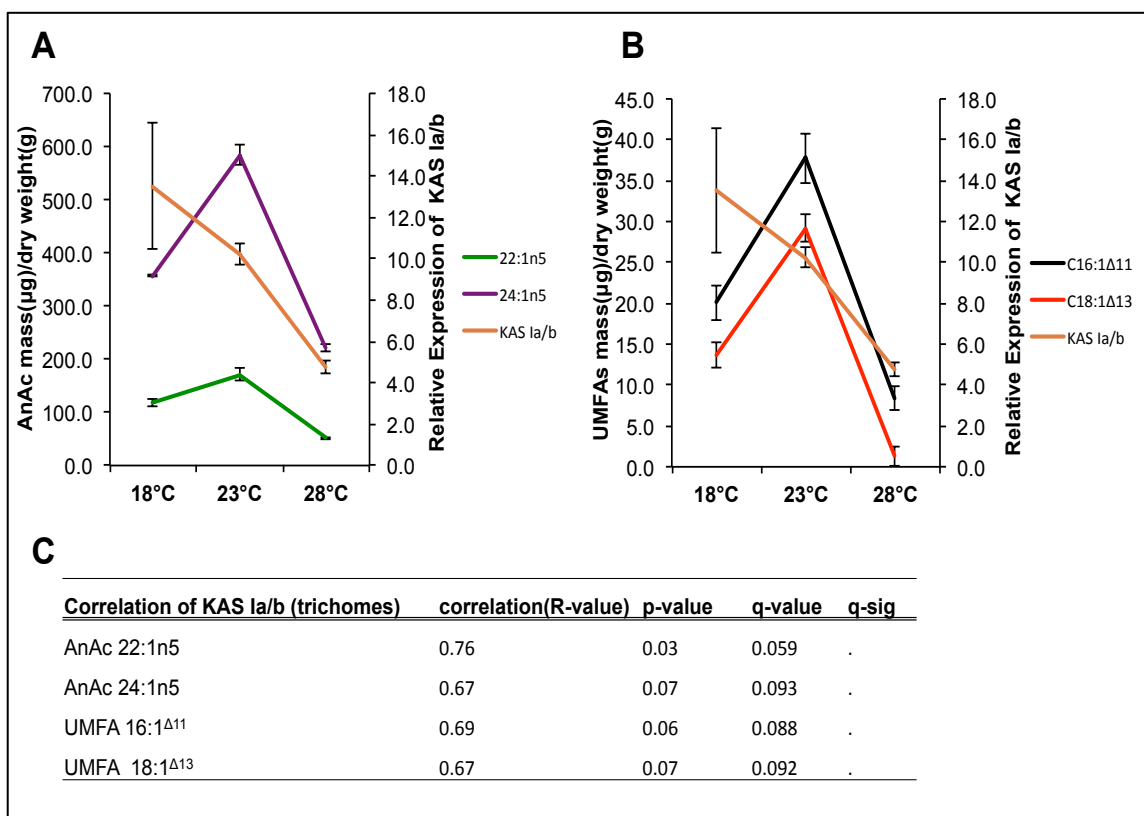
Figure 3.13. Correlations of UMFAs and anacardic acids production with KAS Ia/b expression at distinct temperatures. (A) Correlations of AnAc 22:1n5 and 24:1n5 with KAS Ia/b. (B) Correlations of UMFAs 16:1$^{\Delta 11}$ and 18:1$^{\Delta 13}$ with KAS Ia/b. (C) Corelation R-value for each comparison along with p-values and q-values. Mixed model correlation analysis was performed with the cor.test function of R to generate q-values for each comparison. q-value significance - Not significant (ns), Marginally significant (.), significant (*) and highly significant (**).

A

AnAc mass(µg)/dry weight(g): 700.0, 600.0, 500.0, 400.0, 300.0, 200.0, 100.0, 0.0

Relative Expression of KAS Ic: 0.120, 0.100, 0.080, 0.060, 0.040, 0.020, 0.000

18°C   23°C   28°C

22:1n5
24:1n5
KAS Ic

B

UMFAs mass(µg)/dry weight(g): 45.0, 40.0, 35.0, 30.0, 25.0, 20.0, 15.0, 10.0, 5.0, 0.0

Relative Expression of KAS Ic: 0.120, 0.100, 0.080, 0.060, 0.040, 0.020, 0.000

18°C   23°C   28°C

C16:1Δ11
C18:1Δ13
KAS Ic

C

Correlation of KAS Ic (trichomes)	correlation(R-value)	p-value	q-value	q-sig
AnAc 22:1n5	0.83	0.01	0.04	*
AnAc 24:1n5	0.76	0.03	0.06	.
UMFA $16:1^{\Delta 11}$	0.76	0.03	0.06	.
UMFA $18:1^{\Delta 13}$	0.76	0.03	0.06	.
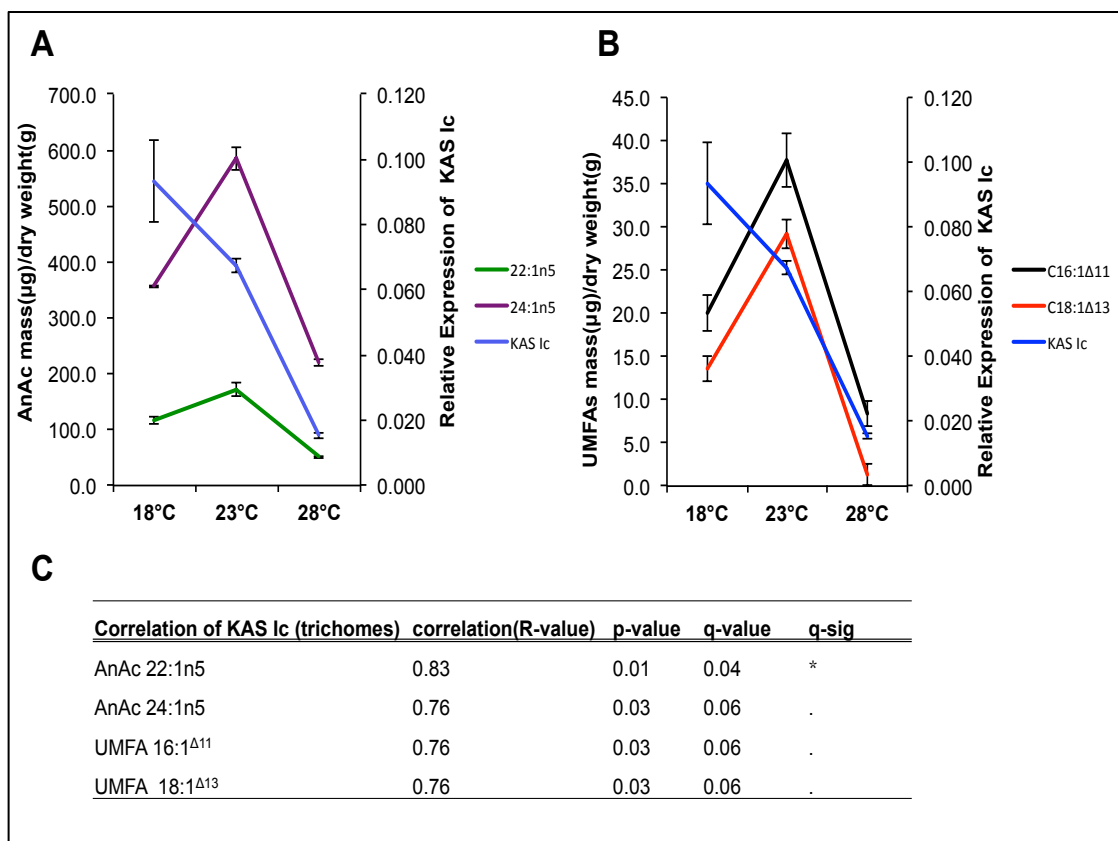
Figure 3.14. Correlations of UMFAs and anacardic acids production with KAS Ic expression at distinct temperatures. (A) Correlations of AnAc 22:1n5 and 24:1n5 with KAS Ic. (B) Correlations of UMFAs $16:1^{\Delta 11}$ and $18:1^{\Delta 13}$ with KAS Ic. (C) Corelation R-value for each comparison along with p-values and q-values. Mixed model correlation analysis was performed with the cor.test function of R to generate q-values for each comparison. q-value significance - Not significant (ns), Marginally significant (.), significant (*) and highly significant (**).

Conclusion

A Pelargonium × hortorum (geranium) *de novo* RNA transcriptome representing distinct tissue (trichomes and bald pedicle) subjected to temperature treatments was generated to study the potential changes and correlations between production of UMFAs and their derived AnAc along with expression of potential genes involved in UMFA biosynthesis. This approach was used to identify

specific paralogs of genes involved in metabolism of UMFAs and AnAc. The RNA-Seq database validated existing EST database genes and also provided additional identification of complete sequences of EST genes and identification of genes that were missing from EST database. Examples of these genes includes FAT-A that is highly expressed in trichomes and showed increased in expression at 23°C, novel reductases, hydrolases, fatty acid acyl-ACP thioesterases, β ketoacyl-ACP synthases, polyketide synthases, and 3-ketoacyl-CoA synthases. A limitation of this *de novo* transcriptome assembly without a reference genome was the lack of distinction between genes that were isoforms (that result from alternative splicing) or paralogs (that result from gene duplications)[107, 134-136]. This distinction is important because splice variants provide important insights into metabolic regulation since they can be tissue, stage or time specific during the plant development[135]. Potential future direction would be to analyze transcriptome data using bioinformatics tools like IsoSVM to identify isoforms or paralogs[135]. Moreover, the *de novo* RNA transcriptome generated for geranium is a novel platform for identification of various genetic elements like enzymes and transcription factors involved in trichome metabolism. The correlation of effect of temperature on gene expression, UMFA and AnAc production lead to the conclusion that 23°C was the optimal temperature for UMFA and AnAc synthesis compared to 18°C and 28°C. Increased temperature (28°C) caused significant reduction in gene expression and production of metabolites as expected based on previous work[97]. Production of UMFAs ($16:1^{\Delta 11}$ and $18:1^{\Delta 13}$) was positively correlated with production of AnAc (22:1n5 and 24:1n5) at all temperature's

indicating a relationship between the substrate and the metabolite at a given temperature. Interesting the ratios of AnAc (24:1n5/ 22:1n5) and UMFAs ($18:1^{\Delta 13}/16:1^{\Delta 11}$) did not show corelation at 28°C and the amount AnAc 24:1n5 was more than AnAc 22:1 at all three temperatures whereas the amount AnAc 24:1n5 substrate $18:1^{\Delta 13}$ was less compared to $16:1^{\Delta 11}$ at all three temperatures. This indicates a possibility of substrate preference for $18:1^{\Delta 13}$ by the downstream enzymes for production of AnAc 24:1n5. Since ACP 1, ACP 2, KAS I-a/b, and KAS I-c were significantly correlated with production of target metabolites, they are potentially involved in UMFA synthesis. Expressing them in heterologous plants or bacterial systems can help in further validation of their function in UMFA metabolism.

CHAPTER 4

MICRO-RNA DATABASE FOR *PELARGONIUM × HORTORUM.*

Summary

Anacardic acids (AnAc) are produced in *Pelargonium × horturum* (garden

geranium) but specific congeners of AnAc (AnAc 22:1n5 and AnAc 24:1n5) are

known to condition a pest-resistance phenotype in specific geranium lines.

These congeners are derived from unusual monoenoic fatty acids (UMFAs)

$16:1^{\Delta 11}$ and $18:1^{\Delta 13}$ that are synthesized only in glandular trichomes of pest-

resistant geranium[36, 45]. Thus, these UMFAs and their AnAc metabolites provide

useful biochemical marker that differentiates the UMFA biosynthetic pathway

from the common fatty acid pathway found in all plants. This is important as the

UMFAs and AnAc have many industrial, medicinal and agricultural applications

and thus elucidating genetic factors that affect UMFA and AnAc biosynthesis

warrant further study. In this effort, an EST database and a *de novo* RNA

transcriptome (Chapter 2 and Chapter 3) were used to provide information on

genetic factors that influence primary metabolites and their derived specialized

metabolites. This study further extends these genetic resources by creating a

micro-RNA (miRNA) transcriptome for geranium. The miRNA transcriptomes

have only recently drawn research attention and are thought to be important for

regulation of gene expression. To generate a geranium miRNA database that represents sequences with broad physiological roles, two distinct tissues (trichomes and bald pedicles) and a temperature treatment (plants grown at 18°C and 23°C) were utilized. Furthermore, gene targets for geranium miRNA were identified using a Plant Small RNA Target Analysis Server, which facilitated selection of target miRNAs for future investigation.

Introduction

Synthesis of Plant MicroRNA

Micro-RNAs (miRNAs) are small non-protein coding single stranded RNA's that regulate posttranscriptional gene expression by cleaving mRNA or inhibiting the translation of target genes[137] (Figure 4.1). In plants miRNA sizes range from 21-24 nucleotides and are found to affect important biological processes like growth, development, metabolic pathways, and abiotic and biotic stress response[138, 139]. Plant miRNA research is fairly recent and the bulk of knowledge about their mechanism of action and biogenesis is focused on model plant species like *Arabidopsis* and rice (Figure 4.1). Other details like origin, function, and evolution are topics of on-going research since miRNAs are characterized in only a limited number of plant species[137].

Micro-RNA synthesis in plants begins with transcription and splicing of MIR gene by RNA polymerase II to form primary miRNA transcripts (Figure 4.1)[137, 140, 141]. The dicer- like (DLC) enzyme family identifies the primary miRNA transcripts that are folded into hairpin structures. Unlike animals, plants DLC enzymes vary in

size and differ in number between plant species. For example, in *Arabidopsis*

there are four types of DLCs - DLC1, DLC2, DLC3 and DLC4 that generate

miRNA of 21, 22, 23 and 24 nucleotides respectively (Figure 4.1)[137].
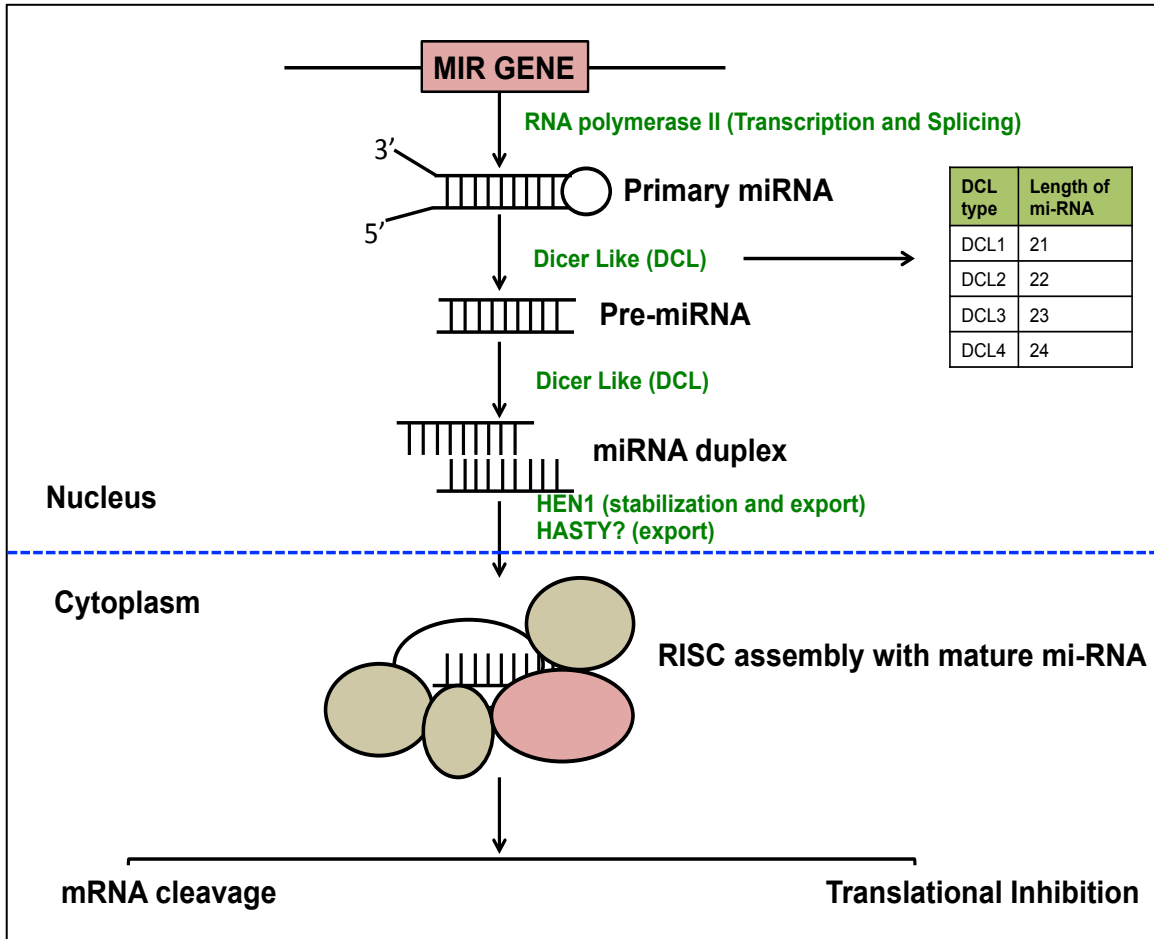
Figure 4.1. A simplified overview of miRNA biogenesis in plants [137, 140, 141].
HEN1: Hua Enhancer 1, RISC: RNA-induced silencing complex. This is a
modified figure adpoted from Budak et.al., 2015[137].

DLCs cleave primary miRNA to pre-miRNA and then pre-miRNA to miRNA

duplex within the nucleus. Hua Enhancer 1 (HEN1) stabilizes miRNA by

methylation of 3' terminus and then exports it into the cytoplasm. Within the

cytoplasm mature miRNA becomes a part of RNA-induced silencing complex

(RISC) and binds to its target based on complementarity and regulates the

expression of a gene either by RNA cleavage or translational inhibition (Figure 4.1)[137]. Unlike the case in animals, there is a high degree of sequence complementary between plant miRNA and their target transcript. Thus bioinformatics computational methods can be employed to identify targets of plant miRNAs[142]. Target identification can help in prediction of the physiological roles of the miRNAs based on the limited information available from the plant miRNA database. This method of identification has been validated by experiments in a number of plant species[140].

Plant miRNAs regulatory functions

Plant miRNA targets include transcripts that encode regulatory proteins, genes and transcription factors essential for signal transduction, plant development and metabolism[140 143, 144]. Plant miRNAs play a critical role in auxin signaling, organ boundary formation, defining organ polarity, flowering time, growth of leaves and reproductive development[145-150]. Interestingly miRNAs are also involved in trichome development[151-155]; examples of these include miRNA156 that suppresses expression of a plant specific transcription factor *SQUAMOSA PROMOTER BINDING PROTEIN LIKE* (SPL) and increases the density of trichomes on plant surface in *Arabidopsis*[156, 157]. In contrast, miRNA171 suppresses the activity of LOST MERISTEMS genes that promote trichome development, thereby decreasing the density of trichomes on stems and floral organs of *Arabidopsis*[154]. Additionally, plant miRNAs are involved in lipid metabolism[158-160], examples of these include miRNA2102 in *Panicum virgatum* (switchgrass) which targets fatty acid desaturase and lipid binding protein to

enhance biofuel production[159] and in *Solanum tuberosum* (potato), fatty acid biosynthesis is regulated by miRNA530 that targets biotin carboxylase carrier protein, and miRNA1442 that targets β-ketoacyl-ACP synthase I gene[161].

Therefore, miRNAs are important factors for studying UMFA and anacardic acid biosynthesis in the trichomes of geraniums. Next generation sequencing (Illumina platform) was used to create a *de novo* geranium mi-RNA database that includes two distinct tissues (trichomes and bald pedicles) as well as trichome tissue subjected to two temperatures (18°C and 23°C). In total, 336 putative miRNAs were identified and their distribution was evaluated using bioinformatics tools. Targets for miRNAs were predicted based on a Plant Small RNA Target Analysis Server.

Material and Methods

Plant Material

Pelargonium × hortorum accession 88-51-10 was a kind gift of Dr. Richard Craig, The Pennsylvania State University. Plants were vegetatively propagated and grown in one-gallon pots containing MetroMix 360. Plants were maintained in environmental growth chambers at 18°C or 23°C and a 16-hour photoperiod with 250 µmol m^{-2} s^{-1} photosynthetic photon flux.

RNA sample preparation

Pedicles were harvested into 50 ml-conical tubes on ice and then flash frozen in liquid nitrogen and stored at -80°C until used. Trichomes were sheared from the surface of frozen pedicles from multiple tubes as described [57, 76]. RNA was purified as described (Appendix 3). Plant Spectrum total RNA extraction kit (Sigma Aldrich, Cat # STRN10) was used to enrich miRNA samples from total RNA. RNA samples were Dnase treated using Ambion Turbo DNA-Free ™ kit (ThermoFisher Cat #AM1907) before miRNA Library preparation. RNA quality and quantity were analyzed using both nanodrop and bioanalyzer.

miRNA Library Preparation

Total RNA samples (5 µg) were submitted to the University of Louisville's Center for Genetics in Molecular Medicine's (CGeMM) sequencing core facility for preparation of the miRNA library. Briefly, the Truseq Small RNA kit v2 (RS-200-0012) was used to prepare miRNA libraries from 1 µg total RNA. Each Library was individually gel purified on a Novex TBE 6% gel and re-suspended in 10 µl 10 mM Tris-Cl, pH 8.5. Libraries were validated and quantitated by running 1 µl on the Agilent Technologies 2100 Bioanalyzer DNA High Sensitivity Chip. Thirty six cycle single sequencing reads were generated on the Illumina NextSeq500 instrument utilizing the 500 Mid-output v2 sequencing kit.

Bioinformatics Analysis

Bioinformatics analysis was conducted utilizing the University of Louisville Bioinformatics core. Three single-end raw sequencing files (.fastq) representing four physiological conditions (trichome 18°C, bald pedicle 18°C and trichome 23°C) were downloaded from Illumina's BaseSpace (https://basespace.illumina.com/) using the mirdeep2[104, 162, 163] (Figure 4.2)
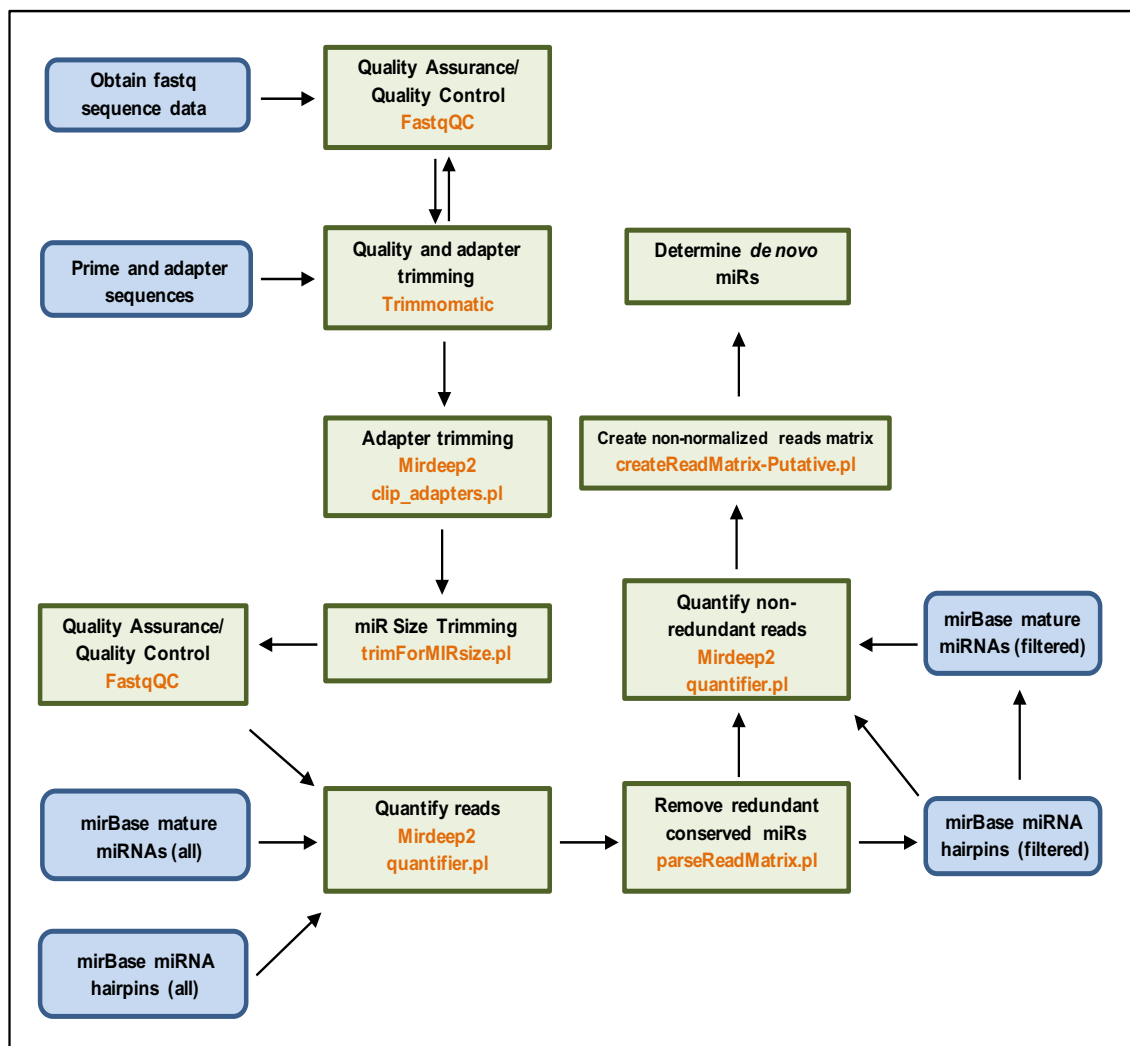
Figure 4.2. Data analysis of pipeline using mirdeep2 for *de novo* miRNA detection. The analysis layout was obtained from Dr. Eric Rouchka (Bioinformatics core).

Bioinformatics approach for *de novo* miRNAs sequences.

Each of the three single-end raw .fastq files for each replicate was concatenated into one single-end .fastq file using the unix cat command. Quality control (QC) of the raw sequence data was performed using FastQC (version 0.10.1). The FastQC results indicated that quality trimming was not necessary since the minimum quality value for all samples is well above Q30 (1 in 1000 error rate), (Appendix 4, Figure A4.1 [108]). Preliminary adapter trimming was performed on each of the samples using a custom file adapters ToTrim.fa which contains a subset of the Illumina TruSeq Small RNA adapter and primer sequences (https://support.illumina.com/content/dam/illuminasupport/documents/documentation/chemistry_documentation/experiment-design/illumina-adapter-sequences_1000000002694-00.pdf). Sequences were trimmed off the adapters with Trimmomatric v0.33 (Appendix 4, Table A4.1)[164]. Peaks at 24 bp were selected for representing the length of plant mature miRNAs (Appendix 4, Figure A4.2). Sequences were trimmed again for a size selection between 18 and 28 nucleotides to be in line with mature miRNA sequences using the custom perl script trimForMiRNAsize.pl. After size selection, roughly 20 to 25 percent of the original sequences remained (Appendix 4, Table A4.2). The final size-trimmed sequences were run through fastQC and the remaining trimmed sequences were collapsed to remove redundant sequences using fastx_collapser. Mirdeep2 alignment tool was used to compare the collapsed sequence files against mature and hairpin miRNAs from mirBase[162, 165].

A read matrix was constructed that included the reads identified for each of the

38,764 known miRNAs across all species using the custom perl script

createReadMatrix.pl. This file was parsed to contain only those miRNAs with a

non-zero count, reducing the overall set to 6,581 miRNAs. These were further

parsed to combine shared miRNAs across species. The miR with the highest

read mapping was used as the putative miR sequence. Complete *Pelargonium* x

hortorum miRNA sequence information is presented in Appendix 4, Table A4.2.

Venn Diagrams

Pratek Genomics suite 6.6 was used to generate a Venn Diagram with program

settings of read count set to > 0 in at least 1 condition (trichome 18°C, bald

pedicle 18°C and trichome 23°C).

Target sequence Analysis

A Plant Small RNA Target Analysis Server (psRNATarget) was used to generate

target prediction tables for pxh-miRNAs (Tables 4.1-4.5)[166]. This analysis utilized

settings for comparison selected (User-submitted small RNAs / preloaded

transcripts), user-submitted small RNAs (miRNAs of *Pelargonium* × *hortorum*),

preloaded transcript/selected library (Arabidopsis thaliana, unigene, DFCI Gene

Index (AGI), version 15,released on 2010_04_08),

(Link:ftp://occams.dfci.harvard.edu/pub/bio/tgi/data/Arabidopsis_thaliana/AGI.rele

ase_15.zip), maximum expectation/low false positive prediction (3.0), length for

complementarity scoring hspsize (20), number of top target genes for each small RNA (200), target accessibility - allowed maximum energy to unpair the target site UPE (25), flanking length around target site for target accessibility analysis (17 bp in upstream and 13 bp in downstream) and range of central mismatch leading to translational inhibition (9-11 nt). The output was generated based on the information available for miRNA from other plants database and shows geranium miRNA sequence, target description with recent discoveries in plant miRNA's target recognition, number of miRNA/target site pairs, type of RNA inhibition (cleavage of ribosome or translational inhibition) and multiplicity of target site.

Results and Discussion

Pelargonium x hortorum de novo miRNA database

A total of 336 putative miRNAs were identified from trichome and bald pedicle of *Pelargonium x hortorum* (Table 4.1). Sequences of all the miRNAs are presented in Appendix 4, Table A4.3. The top two putative miRNAs (miR-166 and miR-319) represent over half of all expressed miRNAs with normalized read counts 144453 and 143655 respectively. miR-166 is implicated in root growth in *Arabidopsis* by targeting the Class III Homeodomain-Leucine Zipper (HD-ZIP III) transcripts[167]. In addition, miR-166 is required for shoot meristem activity, leaf polarity, and vascular patterning of shoot and root in *Arabidopsis* and corn[108, 168-172]. miR-319 targets TCF transcription factors that help to regulate cell proliferation[173, 174]. In addition, miR319a targets TCP4, which is involved in petal growth and development in *Arabidopsis*[175]

Table 4.1. Putative miRNAs detected in *Pelargonium x hotorum.*

miR-166	miR-184-3p	miR-23a-3p	miR-342-3p	miR-210-5p	miR-169b	miR-317
miR-319	miR-5168-3p	miR-8109	miR-186	miR-330	miR-190	miR-31
miR-1128	miR-408-3p	miR-25	miR-454-3p	miR-339	miR-194	miR-31a-5p
miR-396	miR-235-3p	miR-24	miR-1421w-3p	miR-410	miR-20	miR-322-3p
miR-390-5p	miR-92-3p	miR-24a-3p	miR-395	miR-429	miR-20a-5p	miR-3607-3p
miR-393	miR-8175	miR-24	miR-395a-3p	miR-500-3p	miR-20	miR-3963
miR-21a-5p	miR-181a	miR-24a-3p	miR-395	miR-5298d	miR-20a-5p	miR-3968
miR-21	miR-535	miR-100	miR-395a-3p	miR-532	miR-218	miR-4286
miR-21a-5p	miR-2478	miR-164	miR-4995	miR-625-3p	miR-2276-3p	miR-4332
miR-21	miR-7122a	miR-425-5p	miR-7528	miR-744	miR-324	miR-4334-3p
miR-894	miR-27a-3p	miR-941	miR-872-5p	miR-8-3p	miR-335-3p	miR-4505
miR-168	miR-858	bantam	miR-34a-5p	miR-877	miR-340	miR-467f
miR-159	miR-320-3p	miR-146-5p	miR-4171-5p	miR-2355-3p	miR-3535	miR-4792
miR-165a-5p	miR-151-3p	miR-378-3p	miR-263	miR-276	miR-3661	miR-501-3p
miR-827	miR-103-3p	miR-28-3p	miR-275	miR-2778a-5p	miR-374-5p	miR-502
miR-5083	miR-107	miR-200a-3p	miR-473	miR-281-5p	miR-3934-5p	miR-5106
let-7f	miR-22	miR-130a	miR-106	miR-484	miR-5054	miR-5205b
miR-403-3p	miR-22a	miR-130	miR-149-5p	miR-489-3p	miR-516b	miR-548ay-5p
let-7	miR-22	miR-130a	miR-15a	miR-574	miR-530-5p	miR-548a
let-7a	miR-22a	miR-130	miR-23-3p	miR-9	miR-550-5p	miR-548ay-5p
let-7j-5p	miR-477a	miR-183-5p	miR-384-5p	miR-9-5p	miR-6222-5p	miR-548a
miR-3630-3p	miR-5368	miR-301a-3p	miR-399	miR-9a	miR-628-3p	miR-548d-5p
miR-1448	let-7e	miR-301	miR-769	miR-9	miR-654-5p	miR-5735-3p
miR-10a-5p	let-7d-5p	miR-301a-3p	miR-8590	miR-9-5p	miR-6813-5p	miR-589
miR-10	miR-167-5p	miR-301	miR-252-5p	miR-99	miR-7501	miR-6134
miR-398	miR-171	miR-101	miR-29a	miR-9	miR-828	miR-6167
miR-397-5p	miR-1910-5p	miR-1507-3p	miR-7475-5p	miR-9-5p	miR-840-5p	miR-629-5p
let-7	miR-221-3p	miR-205	miR-1180	miR-9a	miR-1127	miR-63-3p
let-7a	let-7b-5p	miR-584	miR-138	miR-1246	miR-124b-3p	miR-652
miR-182	miR-1436	miR-6240	miR-140-3p	miR-1863	miR-126-3p	miR-653-5p
miR-156b	miR-172	miR-192	miR-1582	miR-3120	miR-126a-3p	miR-660
let-7	miR-16a-5p	miR-1895	miR-1692	miR-4451	miR-126-3p	miR-6752-3p
let-7j-5p	miR-16	miR-5139	miR-1839	miR-4492	miR-135-5p	miR-7
miR-156b	miR-16a-5p	let-7k-5p	miR-279	miR-5119	miR-142-5p	miR-767
miR-156	miR-16	miR-203	miR-315	miR-5293	miR-142-5p	miR-8649
miR-92a-3p	miR-423-3p	miR-394	miR-34-5p	miR-541-5p	miR-142a-5p	miR-9226
miR-5049c	miR-141	let-7c	miR-3931-3p	miR-5532	miR-1468-5p	
miR-26	let-7g	miR-148-3p	miR-4504	miR-6087	miR-15	
miR-26a	miR-170-5p	miR-421	miR-750	miR-6788-5p	miR-153	
miR-26	miR-93	miR-482	miR-8155	miR-7158-3p	miR-193-5p	
miR-26a	miR-6300	miR-424-3p	miR-2111-5p	miR-738	miR-195	
miR-1133	miR-375	miR-486	miR-238	miR-850	miR-195a-5p	
miR-6478	miR-125	miR-81a	miR-3470b	miR-8523	miR-1957a	
miR-191-5p	miR-157a-5p	miR-128	miR-5538	miR-8716	miR-195	
miR-191-5p	miR-222	miR-1260	miR-702-3p	miR-8938	miR-195a-5p	
miR-191a	miR-5072	miR-1511	miR-1307	miR-1120b-3p	miR-199a-5p	
miR-27-3p	miR-5181-3p	miR-3954	miR-146-5p	miR-1303	miR-2170	
miR-2916	miR-160	miR-6483	miR-146a-5p	miR-133	miR-3018	
miR-30a-5p	miR-6476a	miR-7767-5p	miR-156aa	miR-169	miR-31	
let-7i-5p	miR-98-5p	miR-345	miR-196	miR-169aa	miR-31a-5p	

Red highlight, miR-166 and miR-319 (most abundant). The list of micro-RNA is based on total normalized read counts, highest to lowest from top to bottom.

Distribution of miRNAs

Normalized read counts (>0) for each miRNA was not equally distributed in the three physiological conditions used to generate this database (Figure 4.3). It is important to note that due to lower read counts and lack of biological replicates, these data cannot be used for quantifying expression of these miRNAs as this database screens for presence or absence of miRNA in a given tissue and/or temperature. In total, 110 miRNAs (33%) were observed to be present in all three conditions whereas the remaining miRNAs were found only in a specific tissue type or at a specific temperature. Only 27 miRNAs (8%) were present exclusively in trichome tissue and 35 miRNAs (10%) were found exclusively in bald pedicle tissue while a total of 66 miRNAs (19%) were detected only in trichome at 18°C and 19 miRNAs (5%) that were found only in trichome at 23°C. Tissue specificity or temperature effects need to be verified/tested with a larger dataset of additional replicates.
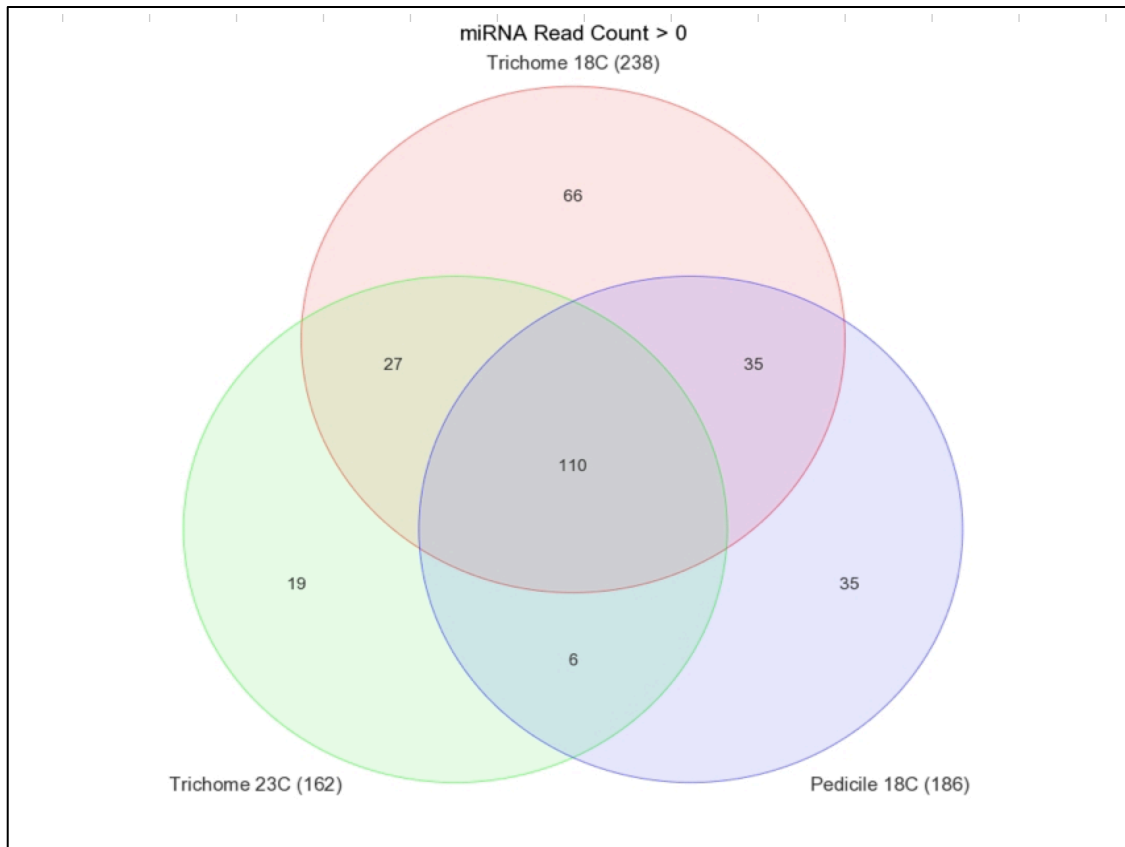
Figure 4.3. Venn diagram for micro-RNAs detected in *Pelargonium × hotorum*. The miRNA sequences were obtained from two tissues and two physiological conditions including trichomes at 18°C, trichomes at 23°C and bald pedicle at 18°C. Read count is set to > 0 for each sample.

Target prediction of miRNAs

The high degree of sequence complimentary between plant miRNA and their target transcript facilities using servers like psRNATarget that employ computational algorithms to screen user submitted miRNAs against available plant miRNA database [142, 166]. The psRNATarget server was used for target prediction of miRNA detected in trichomes at both temperatures (Table 4.2), bald pedicle at 18°C (Table 4.3), trichomes at 18°C (Table 4.4) and trichomes at 23°C (Table 4.5). The psRNATarget server generated details for target prediction, number of target sites, type of RNA inhibition (cleavage of ribosome or

translational inhibition) and target multiplicity for each pxh-miRNAs detected[166].

This approach did not provide results for 31 miRNA sequences (21%).   Target

identification helped in prediction of potential physiological roles of the geranium

miRNAs and highlighted a few miRNAs for future investigation.

Table 4.2. Target description for miRNAs detected in trichomes (18°C and 23°C)

pxh-miRNA	Target Site Pairs	Target Description	Inhibition
miR-184-3p	9	H60S ribosomal protein L26, *Botryotinia fuckeliana*	Cleavage
miR-1436	1	Homologue to UniRef100_Q9LT84 Cluster: UPF0496 protein, *Arabidopsis thaliana*	Translation
miR-6476a	14	Bidirectional sugar transporter SWEET1, *Arabidopsis thaliana*	Cleavage
miR-1895	6	Acidic ribosomal protein P3 , *Corchorus olitorius*	Cleavage
miR-128	3	F-box family protein*, Arabidopsis thaliana*	Cleavage
miR-1511	7	Alpha-glucan water dikinase 2 precursor, *Arabidopsis thaliana*	Cleavage
miR-3954	12	Cysteine/histidine-rich C1 domain-containing protein, *Arabidopsis thaliana*	Cleavage
miR-6483	24	Homeodomain protein 14, *Arabidopsis thaliana*	Cleavage
miR-7767-5p	3	Tochitinase-like protein 1, *Arabidopsis thaliana*	Cleavage
miR-4171-5p	7	Root hair defective 3 gene*, Arabidopsis thaliana*	Cleavage
miR-1180	3	Caffeoyl-CoA O-methyltransferase, *Arabidopsis thaliana*	Cleavage
miR-138	12	Auxin-induced in root cultures protein 12-like, *Camelina sativa*	Translation
miR-140-3p	3	Endo-beta-N-acetylglucosaminidase~gene_id:K18I23.27, *Arabidopsis thaliana*	Translation
miR-1582	27	BZIP transcription factor bZIP122, *Glycine max* (Soybean)	Translation
miR-1692	14	UniRef100_Q85WY0 Cluster: ORF45d; n=1; *Pinus koraiensi*	Cleavage
miR-1839	3	Phosphoglucosamine mutase family protein, *Arabidopsis thaliana*	Translation
miR-279	7	Enhanced downy mildew 2, *Arabidopsis thaliana*	Translation
miR-315	25	Protein serine/threonine kinase gene, *Arabidopsis thaliana*	Cleavage
miR-34-5p	6	Selenium-binding protein-like, *Arabidopsis thaliana*	Cleavage
miR-3931-3p	5	transposase, *Phaeodactylum tricornutum*	Cleavage
miR-4504	19	Alpha/beta-Hydrolases superfamily protein, *Arabidopsis thaliana*	Cleavage
miR-750	10	NAD(P)-binding domain-containing protein, *Arabidopsis thaliana*	Cleavage

Target multiplicity is 1 for all the miRNAs. miR-6300, miR-6240, miR-5139, miR-1R260 and miR-8155 were not detected by the psRNATarget server.

Table 4.3. Target description for miRNAs detected in bald pedicle tissue at 18°C

pxh-miRNA	Target site pairs	Target Description	Inhibition
miR-2111-5p	18	F-box family protein, *Arabidopsis lyrata*	Cleavage
miR-238	3	Phosphatidylinositol-4-phosphate 5-kinase , *Arabidopsis thaliana*	Translation
miR-3470b	2	Vacuolar-sorting receptor 1 precursor , *Arabidopsis thaliana*	Cleavage
miR-5538	13	Chloroplast 50S ribosomal protein L2, *Arabidopsis thaliana*	Cleavage
miR-702-3p	2	Chloroplast PSI type III chlorophyll a/b-binding protein , *Brassica juncea*	Cleavage
miR-1120b-3p	6	WD-40 repeat family protein / beige-related , *Arabidopsis thaliana*	Cleavage
miR-1303	19	Gamma-glutamyl hydrolase, *Arabidopsis thaliana*	Cleavage
miR-133	1	ADP, ATP carrier protein 1, *Arabidopsis thaliana*	Cleavage
miR-169	13	Nuclear transcription factor Y subunit A-3, *Arabidopsis thaliana*	Cleavage
miR-190	33	Pentatricopeptide repeat-containing protein-like protein, *Arabidopsis thaliana*	Cleavage
miR-194	13	Adenosyl-L-methionine-dependent methyltransferase, *Arabidopsis thaliana*	Translation
miR-20	1	Jacalin-like plant lectin domain-containing protein, *Arabidopsis thaliana*	Cleavage
miR-218	6	:Cytochrome b561/ferric reductase transmembrane, *Arabidopsis thaliana*	Cleavage
miR-2276-3p	1	Weakly similar todomain-containing protein MOP10, *Arabidopsis thaliana*	Translation
miR-324	2	Similar to protein MIDASIN1, *Arabidopsis thaliana*	Cleavage
miR-335-3p	25	Kinase interacting (KIP1-like) family protein, *Arabidopsis thaliana*	Cleavage
miR-340	20	Transcription factor 25 (Nuclear localized protein 1), *Xenopus tropicalis*	Cleavage
miR-3535	5	Ethylene-responsive transcription factor ERF016, *Arabidopsis thaliana*	Cleavage
miR-3661	2	Ubiquitin-conjugating enzyme E2 -like protein, *Arabidopsis thaliana*	Cleavage
miR-374-5p	4	Hypothetical protein ARALYDRAFT_477717, *Arabidopsis lyrata*	Translation
miR-3934-5p	3	Protein kinase PK1, *Zea mays*	Cleavage
miR-516b	7	nonsense-mediated mRNA decay trans-acting factors~gene, *Arabidopsis thaliana*	Cleavage
miR-530-5p	4	Laccase-2 precursor, *Arabidopsis thaliana*	Translation
miR-550-5p	3	Elongation factor G, chloroplast precursor, *Glycine max*	Cleavage
miR-6222-5p	7	Transducin/WD40 domain-containing protein, *Arabidopsis thaliana*	Cleavage
miR-628-3p	3	Cellulose synthase catalytic subunit (Ath-A), *Arabidopsis thaliana*	Cleavage
miR-654-5p	5	Hypothetical protein PID|e327464 (gb|Z97338), *Arabidopsis thaliana*	Cleavage
miR-6813-5p	6	Mitochondrial carrier protein, putative, *Arabidopsis thaliana*	Cleavage
miR-7501	3	Protein arginine N-methyltransferase 1, *Arabidopsis thaliana*	Cleavage
miR-828	12	Hypothetical protein, *Arabidopsis thaliana*	Cleavage
miR-840-5p	5	ATP binding / alanine-tRNA ligase/ nucleic acid binding protein, *Arabidopsis thaliana*	Translation

Target multiplicity is 1 for all the miRNAs. miR-5054 was not detected by the psRNATarget server.

Table 4.4. Target description for miRNAs detected in trichome tissue at 18°C.

pxh-miRNA	Target site pairs	Target Description	Inhibition
miR-34a-5p	6	LRR receptor-like serine/threonine-protein kinase ERL2 mRNA, *Arabidopsis thaliana*	Cleavage
miR-29a	2	Protein TRF-like 7 mRNA, *Arabidopsis thaliana*	Translation
miR-7475-5p	17	DNA-binding protein, *Arabidopsis thaliana*	Cleavage
miR-2355-3p	13	Inorganic phosphate transporter and chloroplast precursor, *Arabidopsis thaliana*	Cleavage
miR-276	4	Nucleoside diphosphate kinase, *Arabidopsis thaliana*	Cleavage
miR-2778a-5p	26	Pectinacetylesterase, *Arabidopsis thaliana*	Cleavage
miR-281-5p	2	Flavin-containing monooxygenase, *Arabidopsis thaliana*	Cleavage
miR-484	3	Heat shock transcription factor HSF30 homolog, *Arabidopsis thaliana*	Cleavage
miR-489-3p	2	Disease resistance protein homolog, *Arabidopsis thaliana*	Translation
miR-574	1	Chromosome 1 sequence, *Arabidopsis thaliana*	Translation
miR-9	18	Mitochondrial precursor; n=1; *Arabidopsis thaliana*	Translation
miR-9-5p	17	Disease resistance protein, *Arabidopsis thaliana*	Cleavage
miR-99	2	BHLH protein-like, *Arabidopsis thaliana*	Cleavage
miR-124b-3p	1	Lysosome-related organelles complex 1 subunit 1 mRNA, *Arabidopsis thaliana*	Translation
miR-126-3p	3	12S cruciferin seed storage protein; n=3; *Arabidopsis thaliana*	Translation
miR-135-5p	11	AMP deaminase like protein, *Arabidopsis thaliana*	Cleavage
miR-142-5p	10	Adenosylmethionine decarboxylase 1 beta chain, *Arabidopsis thaliana*	Cleavage
miR-1468-5p	1	Putative protein (T16L4.170) mRNA, *Arabidopsis thaliana*	Translation
miR-15	2	Uncharacterized protein mRNA, *Arabidopsis thaliana*	Tranlastion
miR-153	8	TMV resistance protein, *Arabidopsis thaliana*	Tranlastion
miR-195	5	AP2-like ethylene-responsive transcription factor, *Arabidopsis thaliana*	Translation
miR-2170	21	Salicylic acid responsive calcium/calmodulin binding protein, *Arabidopsis thaliana*	Cleavage
miR-3018	35	Ubiquitin carrier protein , *Arabidopsis thaliana*	Cleavage
miR-31	3	Chloroplast tRNA-Ala, tRNA-Ile, 16S rRNA, tRNA-Val, *Solanum nigrum*	Translation
miR-322-3p	14	RNA recognition motif (RRM)-containing protein, *Arabidopsis thaliana*	Cleavage
miR-3607-3p	14	Nucleobase-ascorbate transporter 11, *Arabidopsis thaliana*	Cleavage
miR-3968	2	AP2-like ethylene-responsive transcription factor, *Arabidopsis thaliana*	Translation
miR-4334-3p	7	Phosphatidylinositol/phosphatidylcholine transfer protein-like, *Arabidopsis thaliana*	Cleavage

Table 4.4 *continued*

miRNA		Target	
miR-467f	41	AP2-like ethylene-responsive transcription factor, *Arabidopsis thaliana*	Cleavage
miR-501-3p	3	Zinc finger protein-like, *Arabidopsis thaliana*	Cleavage
miR-502	3	Zinc finger protein-like, Arabidopsis thaliana	Cleavage
miR-5106	1	ABC transporter permease protein, *Azorhizobium caulinodans*	Cleavage
miR-5205b	2	Mitochondrial phosphate translocator, *Arabidopsis thaliana*	Cleavage
miR-548ay-5p	44	Mitochondrial import inner membrane translocase, *Arabidopsis thaliana*	Translation
miR-548a	44	Ethylene-responsive transcription factor, *Arabidopsis thaliana*	Cleavage
miR-548d-5p	41	Ethylene-responsive transcription factor, Arabidopsis thaliana	Cleavage
miR-5735-3p	2	UniRef100_Q66GL9 Cluster: At3g52490; n=1; *Arabidopsis thaliana*	Cleavage
miR-589	4	Eukaryotic Translation initiation factor 5A-3; *Arabidopsis thaliana*	Cleavage
miR-6167	27	Chlorophyll a-b binding protein, *Arabidopsis thaliana*	Cleavage
miR-629-5p		Protein kinase AtSIK, *Arabidopsis thaliana*	Cleavage
miR-653-5p	10	CEN (centroradialis)-like phosphatidylethanolamine protein, *Arabidopsis thaliana*	Cleavage
miR-660	1	Pentatricopeptide repeat-containing protein mRNA, *Arabidopsis thaliana*	Translation
miR-6752-3p	2	Dna-directed rna polymerase II, *Asparagus officinalis*	Cleavage
miR-7	7	DNA binding / protein dimerization , *Arabidopsis thaliana*	Translation
miR-767	2	Penicillin-binding protein; n=1; *Mesorhizobium*	Translation
miR-8649	1	mutT like protein, *Arabidopsis thaliana*	Cleavage

Target multiplicity is 1 for all the miRNAs. miR-252-5p, miR-1127, miR-193-5p, miR-195a-5p, miR-1957a, miR-199a-5p, miR-31a-5p, miR-317, miR-31a-5p, miR-31a-5p, miR-322-2p, miR-3607-3p, miR-3963, miR-3968, miR-4286, miR-4332, miR-4505,miR-467f, miR-4792, miR-6134, miR-63-3p, miR-652 and miR-9226 were not detected by the psRNATarget server.

Table 4.5. Target description for miRNAs detected in trichome tissue at 23°C.

miRNA	Target site pairs	Target Description	Inhibition
miR-263	6	Chloroplast thylakoid lumen protein, *Arabidopsis thaliana*	Cleavage
miR-275	4	Transmembrane nine protein 6, *Arabidopsis thaliana* and *Vitis vinifera*	Cleavage
miR-1863	15	Putative retroelement polyprotein, *Arabidopsis thaliana*	Cleavage
miR-3120	2	Citrate synthase and peroxisomal precursor, *Arabidopsis thaliana*	Cleavage
miR-5293	46	Zinc knuckle (CCHC-type) family protein, *Arabidopsis thaliana*	Translation
miR-541-5p	3	Myosin-like protein, *Arabidopsis thaliana*	Cleavage
miR-5532	13	Adenylate kinase family-like protein, *Solanum tuberosum*	Cleavage
miR-6788-5p	25	Leucoanthocyanidin dioxygenase-like protein, *Arabidopsis thaliana*	Cleavage
miR-7158-3p	5	Transcription factor TCP23 [Arabidopsis thaliana]	Cleavage
miR-850	6	Bidirectional sugar transporter SWEET4 and nodulin MtN3 protein , *Arabidopsis thaliana*	Cleavage
miR-8716	12	ABC1 protein kinase 6, *Arabidopsis thaliana*	Cleavage
miR-8938	10	MATE efflux family protein, *Deinococcus geothermalis*	Cleavage

Target multiplicity is 1 for all the miRNAs. miR-1246, miR-4451, miR-4492, miR-5119,miR-6087, miR-738 and miR-8523 were not detected by the psRNATarget server.

Identification of potential physiological roles of geranium miRNAs

Ubiquitous miRNAs (trichomes and bald pedicle)

Unlike animal miRNAs, plant miRNAs are less conserved and tend to be species specific[141]. Yet, since the emergence of gene silencing is attributed to evolution, it may be possible to identify conserved RNA sequences across plant species. From geranium miRNAs that were identified in all three libraries (Appendix 4, Table A4.4), we found that pxh-miRNAs 319, 390-5p,159 ,403p, 398, 408-3p,103-3p, 171 are conserved and found in *Arabidopsis*, bryophytes and other flowering plants[137, 140, 176]. Some miRNAs in this category have experimental support for specific biological processes. For example, miRNAs 166, 172, 164

are involved in flower development and organ polarity[137, 140, 146, 155, 157, 177-181],

miR-319 is involved in leaf growth and development[137, 152, 182], mir-393 and 160

affects auxin signaling[140, 183], miR-168,403-3p, 319 and159 are involved in

miRNA biogenesis and metabolism[137, 176, 184,185], miR 403-3p and 159 has a role

in reproductive development [140, 176, 184] and miR 319 and 395 are specific to

stress response[137, 139]. This suggests the potential for similar functions of these

miRNAs in geranium.

Additionally, four potential miRNA targets were identified for studying

UMFAs and anacardic biosynthesis. These include, miR156 which is involved in

trichome distribution, trichome development and stress response[137, 154], miR103

and miR107 that are involved in regulation of polyphenolic compounds[186, 187] and

miR125,378-3p that is involved in lipid metabolism[160].

miRNAs identified in specific tissue (trichomes or bald pedicle)

The output generated from the psRNATarget server for miRNAs that were

present in both types of trichome tissue, only in bald pedicle or only in trichome

at a particular temperature facilitated selection of miRNAs for future evaluation

(Table 4.2, 4.3, 4.4 and 4.5).

For miRNAs detected only in bald pedicle tissue, potential physiological roles

were identified for miR-335-3p - involved in lipid metabolism[160], miR-828 -

conserved in few plant species and its overexpression reduces the anthocyanin

production in *Arabidopsis* [137, 188], miR-2111-5p - conserved (specially eudicots)

[137] and miR169 - acts on nuclear transcription factor Y and is involved in viral stress response[138, 166, 189].   All of these miRNAs might have similar function in geranium. Interestingly, miR1120b has only been identified in barley making gernanium only the second plant reported to have this miRNA [166, 190].

Within miRNA detected only in trichomes, pxh-miR-6300, 6420, 5193, 1260 and 8155 are particularly interesting because the plant server could not identify targets for these miRNAs, indicating these miRNAs may be specific to geranium trichomes or may not yet identified in any of the plant species. Other noteworthy miRNA candidates identified only in trichomes were miR-3954 which targets triose phosphate transporter gene required for metabolism and initiates siRNA biosynthesis [191-193] and miR-6483 which targets transmembrane proteins and regulates transport of cellular materials across the membrane [191, 192, 194]. To facilitate further studies of UMFA and AnAc synthesis in trichomes potential targets was uncovered. For example, miR-138 and miR 4171 are involved in the development of root hairs (trichomes)[166, 195] and miR-4504 and miR1180 regulate fatty acid biosynthesis enzymes[166, 195]. Additionally, 66 miRNAs were identified only in trichome at 18°C whereas 19 miRNAs were identified only in trichome at 23°C, all of these are important candidates to evaluate temperature effects (Tables 4.4, 4.5). An example of this is miR-5532 which is only present in trichome at 23°C and it is well studied in medicinal herb, *Picrorhiza kurroa,* for its regulatory effect on growth, development and secondary metabolite production[196].

Conclusions and future prospects

Although thousands of miRNAs have been identified in model plant species, they still remain unknown in most non-model plant species. To date, miRNA from only 71 plant species has been reported[141]. The observations made via the *Pelargonium × hotorum* miRNA catalogue and the sequences obtained are an important and novel addition to the limited plant miRNA data. The limitation of this study was that only three conditions without replicates were used for sequencing and thus further analysis on differential expression or target estimation could not be performed. However, with help of bioinformatics tools potential target sequences and miRNAs were suggested for further evaluation.

A future direction for geranium miRNA research would be using bioinformatics tools to identify specific miRNA targets using RNA-SEQ transcriptome database (Chapter 3). Furthermore, the larger sample size of transcripts from trichomes and bald pedicle at different temperature can be obtained and used to correlate differential expression of miRNA with target transcript abundance and validate presence of miRNA by using experimental techniques like quantitative real time PCR. Functional verification of miRNAs can be conducted via overexpression or knockdown/knockout expression of miRNAs. Evaluation of functions for miRNA identified in *Pelargonium × hotorum* will be not only a valuable addition to miRNA plant research but also will help in expanding the understanding of miRNA related pathways and gene targets for non-model plants like geraniums.

CHAPTER 5

CO-EXPRESSION OF *PELARGONIUM × HORTORUM* ACYL CARRIER
PROTEIN WITH Δ^9 14:0-ACP-DESATURASE.

Summary

Unusual monoenoic fatty acids (UMFA's) and derived specialized metabolites

called anacardic acids (AnAc) are produced in glandular trichomes of

Pelargonium × hortorum (geranium). The $16:1^{\Delta 11}$ and $18:1^{\Delta 13}$ UMFAs are

precursors for the synthesis of AnAc 22:1n5 ad 24:1n5 that confer pest

resistance in geranium. Since UMFA's are potentially produced through a

distinct "metabolic channel" from the common fatty acids biosynthetic pathway,

multiple sequences with homology to fatty acid biosynthesis (FAS) enzymes that

are highly expressed in trichomes were identified (Chapter 2). These enzymes

included acyl carrier proteins (ACPs) that are central to the process of fatty acid

synthesis as they are the carriers of acyl intermediate during FAS and can be

constitutively expressed or regulated based on development and tissue

specificity. To identify the potential role of geranium ACPs in production of

UMFAs, each ACP was co-expressed with myristoyl-ACP desaturase (MAD) in

two heterologous systems. No affects of either ACP 1 or ACP 2 co-expressed

with MAD on UMFAs were detected in tobacco plants. Interestingly, co-

expression of ACP 1 with MAD in *E. coli* lead to reduction in total fatty acids and ACP 2 lead to an overall increase in accumulation of fatty acids in *E. coli* . Neither ACPs was found to affect the specific relative proportion of $16:1^{\Delta11}$ or $18:1^{\Delta13}$ UMFAs.

Introduction

UMFAs have various applications in the chemical and pharmacological industry and it is important to identify all the genetic components involved in their synthesis and enable their large scale cost effective production [5]. To date all the genetic factors (enzymes, transcription factors, micro-RNAs) essential for UMFAs synthesis within a given plant have not been identified and this limits their production and utility in transgenic plants[197]. Research in this area of plant biology suggests that UMFAs are products of primary metabolism that may require isoforms of fatty acid biosynthesis enzymes for their synthesis[4]. This is based on the identification of various novel FAS genes like ACPs, β-ketoacyl-ACP synthase (KASs), acyl-ACP thioesterase (TEs) and acyl-ACP desaturases (AADs) that have specific roles in the production of UMFA in their native plants [9, 22, 23, 32, 33, 96].

AADs and ACPs play a vital role in UMFA synthesis. AADs incorporate the double bond at specific positions needed for production of UMFAs and ACPs are indispensable cofactors that are carriers of acyl intermediate during FAS and their enzyme specificity determines the fatty acid end product. Novel ACP

isoforms have been identified that are not only involved in production of UMFAs but also affect the activity of AADs and are involved production of UMFA[9, 33, 198]. Examples of these include, a novel ACP and Δ4 16:0 ACP desaturase that leads to production of UMFA $18:1^{\Delta6}$ (petroselinic acid) in *Coriandrum sativum* and a novel ACP and Δ6 16:0 ACP desatuarse that produces UMFA $16:1^{\Delta6}$ in *Thunbergia alata*[94].

Within trichomes of *Pelargonium × hortorum* a novel Δ9 14:0-acyl carrier protein (ACP) desaturase (MAD) has been identified [75]. The AAD gene is responsible for producing myristoleic acid ($14:1^{\Delta9}$) that is elongated to $16:1^{\Delta11}$ and $18:1^{\Delta13}$ [199]. Two acyl carrier proteins (ACP 1 and ACP 2) that are highly expressed in the trichomes of geranium were identified (Chapter 2) and expression of each was correlated with production of UMFAs and AnAc at 18°C, 23°C and 28°C (Chapter 3). Due to this, both ACPs were ideal candidates for further evaluation of their metabolic role with geranium MAD in UMFA biosynthesis. Both ACPs were individually co-expressed with MAD in *E. coli* and tobacco to study how expression affects the production of UMFAs [200, 201]. Production of $16:1^{\Delta11}$ and $18:1^{\Delta13}$ or higher ratios (unsaturated to saturated) was predicted to be altered by one or both of the ACPs when expressed with MAD thus indicating the potential involvement in UMFAs synthesis and increases the production of target UMFAs.

Material and Methods

General Lab Procedures

All enzymes, antibiotics and kits used for cloning were obtained from Roche diagnostics or New England BioLabs. General lab reagents, standards and solvents for gas chromatography were obtained from Fisher or Sigma Aldrich. Bacterial Expression cells lines BL21(DE3)pLysS and Rosetta™(DE3)pLysS were obtained from Novagen. *Agrobacterium tumefaciens* LBA4404 cell line obtained from Takara. All the primers designed for cloning experiments and RT-PCR (Appendix 5, Table A5.1) were obtained from Eurofins.

Cloning and Expression in *E. coli*

ChloroP (http://www.cbs.dtu.dk/services/ChloroP/) , SignalIP (http://www.cbs.dtu.dk/services/SignalP/) and Psort (http://psort.hgc.jp/form.html)

 servers were used to predict signal peptide sequence of ACPs[9]. Primers were designed to exclude the signal peptide sequence for expression of ACPs in *E. coli* and the ATG site was incorporated at the 5' end (Figure 5.1)[9, 202-204]. ACP 1 and ACP 2 cDNAs cloned in pBluescript SK- were used as templates. The 5' primer with NdeI site and 3' primer with XhoI site were used for amplification of these ACPs via standard end point PCR. The corresponding PCR product for each ACP was subsequently cloned between the NdeI (5') and Xho1 (3') sites of the pet22b vector (Novagen) to generate constructs pet22b-ACP 1 and pet22b-

ACP 2, respectively. Directional cloning was verified by restriction analysis and

sequencing of the constructs.

Each ACP was also cloned into the pet3d-MAD construct [75]. For these dual

constructs, XbaI/SpeI sites, ribosome binding site and ATG were incorporated in

5' primers and ribosome binding site, stop site and NcoI site were incorporated in

3' primers (Appendix, Table A5.1). A ribosome binding site was incorporated for

both primer sets since the same T7 promoter controlled both the genes. These

primers were used to amplify each ACP via touchdown PCR and the PCR

products were subsequently cloned between XbaI (5') and NcoI (3') sites of

pet3d-MAD construct to generate dual constructs pet3d-MAD-ACP 1 and pet3d-

MAD-ACP 2, respectively. Directional cloning was verified via restriction analysis

and sequencing. The changes in fatty acid profile as compared to wild type and

production of UMFA was a proof of expression by the dual constructs. The

expression of genes is also validated by the fact that MAD, known to produce

UMFA, was cloned after ACP, thus ensuring transcription of ACP in cell lines in

which the UMFA are detected.

For expression of all the constructs in *E. coli*, two cells lines were tested

BL21(DE3)pLysS and Rosetta™(DE3)pLysS. Rosetta™(DE3)pLysS was

chosen for expression since BL21(DE3)pLysS cells containing the ACP

constructs did not grow. MAD-pet3d construct was induced with 0.4 mM

isopropyl β-D-thiogalactopyranoside (IPTG) at 20°C for 4 hours, 8 hours and 18

hours for optimizing the induction time. Leaky expression of target transgenes was detected in un-induced samples for all genes in the study.

Three biological replicates of all the samples (Rosetta™ (DE3) pLysS, pet22b, pet22b-ACP 1, pet22b-ACP 2, pet3d, pet3d-MAD, pet3d-MAD-ACP 1 and pet3d-MAD-ACP 2) were induced with 0.4 mM IPTG for 18 hours at 20°C with moderate shaking (200 rpm). Cells were pelleted, frozen at -80°C then freeze dried. Dried cell pellets were used to prepare methyl esters for GC analysis.

ACP1
MASFTANSLSLTSISCSFTPIKAPARTSSLKSVSFSINGNGFSSLRL RQGPSRFQISCSAKPETVDKVCEIVKKQLALPEGTEVSGDSKFAAL GADSLDTVEIVMGLEEEFGINVEEESAQNIATVQDAADLIEKLVEKK APA

ACP2
MASFTPNSVSMTSISCSLRPNMAPTMISGMKSASFSINRNGFPSL RLQQGSSRLQVLCSAKAETVDKVCEIVRKQLAIPTDTEVSGESKFA ALGADSLDTVEIVMGLEEEFGISVEEESAQSIATVQDAADLIEKLVE KKDA

Figure 5.1.*Pelargonium × hortorum* acyl carrier protein (ACP) amino acid sequences. Amino acids highlighted in blue are signal peptide sequences determined using ChloroP, SignalIP and Psort server[9, 202-204] for ACP 1 and ACP 2.

Cloning and Expression in Tobacco

The geranium ∆9 14:0-ACP desaturase was previously cloned into pMON25661 (Schultz Lab, unpublished). This construct utilized the CaMV 35S Promoter and included the heterologous transit peptide (Arab-SSU1A-transit) and included the

E9 terminator sequence. The MAD expression cassette including promoter and

termination sequences was digested from MAD-pMON25661 and cloned into the

NotI site of the pRI201-AN (Takara) dual plant expression vector to create pRI-

MAD,   directional cloning was verified via restriction analysis and sequencing.

Each ACP was cloned into the NdeI/SalI sites of pRI-MAD using 5' primers with

NdeI site and 3' primers with SalI site (Appendix 5, Table A5.1) via touch down

PCR. PCR products for each ACP were subsequently cloned between NdeI (5')

and SalI (3') sites of pRI-MAD construct to generate dual constructs pRI-MAD-

ACP 1 and pRI-MAD-ACP 1 respectively. Directional cloning was verified via

restriction analysis and sequencing.

Plant expression constructs (pRI201-AN vector, pRI-MAD, pRI-MAD-ACP 1 and

pRI-MAD-ACP 1) were transformed into *Agrobacterium tumefaciens* LBA4404

via electroporation (Settings: Bacteria- Agr, Volts- 2.2 kV, pulse - 1ms).

Restriction analysis was used to confirm presence of the constructs in the cell

line (this was done by transforming the Agrobacterium constructs in *E. coli*).

Wild type tobacco was transformed using the standard leaf disc transformation

method (Figure 5.2)[205] with modification as outlined in Appendix 5. Kanamycin

resistant shoots were selected individually and transferred into the rooting media

with resulting plants transferred to soil (Figure 5.2).

Multiple independent primary transformants were obtained for each construct and

included pRI-MAD (n=4), pRI201-AN (n=2), pRI-MAD-ACP 1 (n=3) and pRI-

MAD-ACP 2 (n=3). Leaves of each plant were harvested, frozen at -80°C then

freeze dried prior to fatty acid analysis via GC. Fatty acid content of transgenic

seeds was also analyzed via GC . Transgenic seeds were also sterilized (10%

ethanol and 10% bleach) and placed on Murashige and Skoog medium

containing 100 mg/ml Kanamycin to obtain next generation from primary

transformants. The plants obtained from this procedure were transferred to pots.

The leaves and seeds of the next generation plants were harvested and freeze

dried for further analysis.

Reverse Transcription PCR

RNA was extracted from leaves of each transgenic plant using Qiagen RNeasy

Plant Mini Extraction Kit and RNA quality and quantity was analyzed using

Nanodrop 2000. RNA samples were DNase treated using Ambion Turbo DNA-

Free ^{TM} kit (ThermoFisher) and cDNA was synthesized using SuperScript® III

First-Strand Synthesis System (Invitrogen). Primers were designed (Appendix 5,

Table A5.1) to amplify specific genes from cDNA of each transgenic plant via

reverse transcription PCR to confirm the presence of transgene in the tobacco

plants. Wild Type cDNA was used as control to confirm the absence of

transgenes in wild type plants.
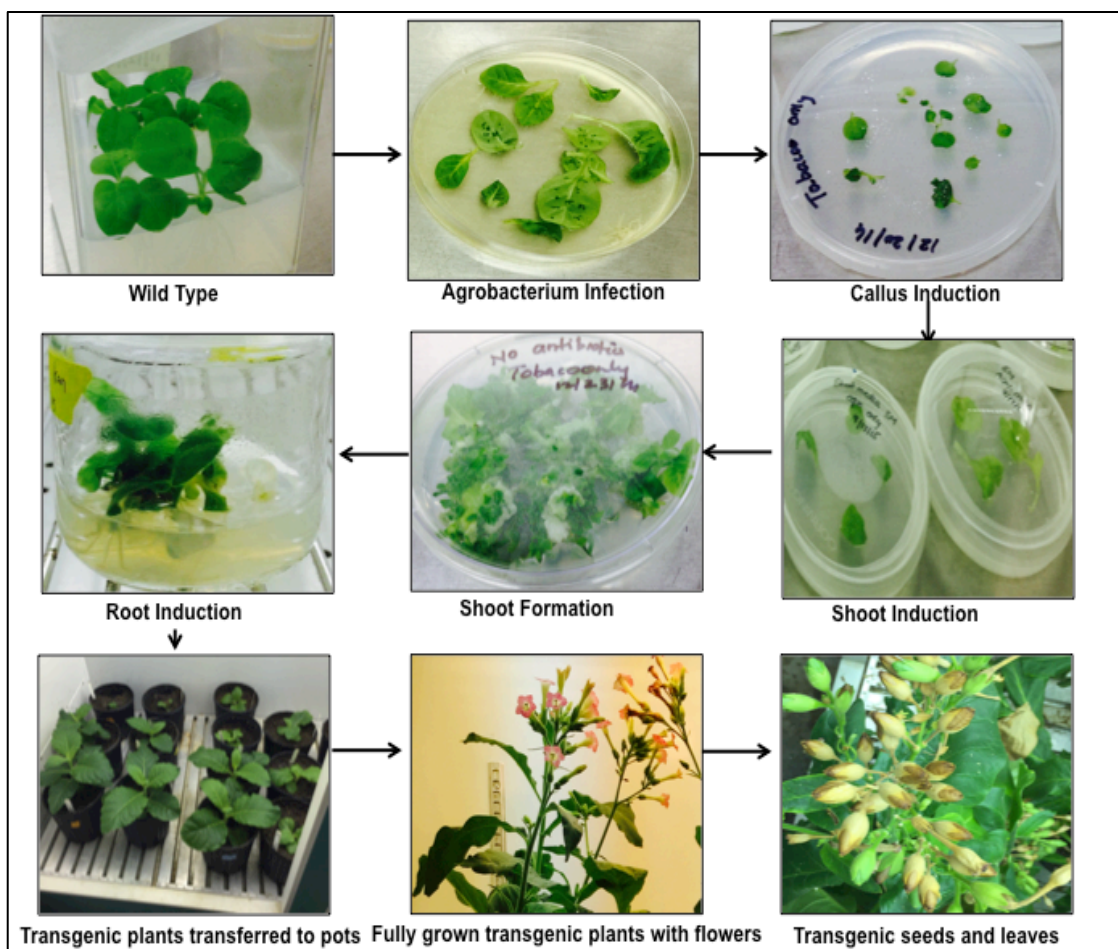
Figure 5.2. Tobacco leaf disk transformation. Leaf disk transformation was essentially as described [205] with modifications.

GC analysis

For quantitative analysis, triheptadecanoin was added as internal standard to 50 mg of dried transgenic leaves, 50 mg of transgenic seeds or to dried *E. coli* pellets. For all samples, methyl esters were generated using 1 ml boron triflouride (10% in methanol) and 300 µl of toluene was added to increase solubility of lipids prior to heating for 45 minutes at 90°C. Samples were cooled then quenched with 1 ml water prior to hexane extractions. Resulting fatty acid methyl esters were analyzed by gas chromatography (GC) with an Agilent 5890

GC-FID equipped with a 60 m, 0.25 mm, 0.2 μm CP-Sill88 column (Chrompack). The column was programmed with spiltless inlet at 250°C, 21.76 psi, 10.3 ml/ml flow rate and helium gas. Oven was set at initial temperature 150°C with ramp to 250°C at the rate of 2°C /min. Detectors FID were set at 250°C with hydrogen flow rate 3 ml/min, air flow rate 400 ml/min and helium flow rate 15 ml/min. Three technical replicates (1 µl /injection) for each biological replicate were used for the analysis. Authentic GC standards C8-C22 (Sigma Aldrich, Appendix 3, Table A3.1) were used as a external analytical standard to identify peaks in the experimental samples. Most peaks within the C8-C22 were further verified by comparison of retention times to other single of mixtures of standards. Identification of sample peaks was further confirmed by spiking samples with authentic standards.

Statistical Analysis

All analyses were done with R version 3.1.1 (2014-07010). Mixed model analyses were done with the lmer function of the lme4 package[120, 121]. Technical replicate effects were modeled as random and nested within each treatment multiplied by each biological replicate combination. Biological replicate values were modeled as a random effect and crossed with the fixed Treatment effects. Model residuals were found in all cases to satisfy the assumptions of normality and homogeneity of variance. Tukeys multiple comparison contrasts were estimated and tested with the multcomp package of R with the single step method used for adjusting p-values[122]. False discovery rate p-values (FDRp)

were computed with the p.adjust function of R. Correlation analysis was

performed with the cor.test function of R. The q-values were computed with the

qvalue function of the bioconductor package using the default settings[123-126].

Results and Discussion

Expression of Geranium ACP 1 and 2 in *E. coli*

 The effect of expressing geranium ACP 1 and 2 on the fatty acid composition of

E. coli was evaluated. Since previous work indicated cell lines carrying empty

foreign vector may alter the fatty acid profile of the bacteria [206] untransformed as

well as empty vector controls were utilized.   The results of this experiment

indicated changes in fatty acid profile of *E. coli* carrying empty vector (Figure 5.3,

Table 5.1) and thus, all comparisons were made to untransformed as well as to

cells transformed with empty vector. In lines expressing ACP 1 no changes in

the amount of saturated and unsaturated fatty acids was detected while cell lines

expressing ACP 2 showed highly significant increases in C12:0, C14:0, C16:0,

$16:1^{\Delta 9}$, $18:1^{\Delta 11}$ and total fatty acid content (Figure 5.3, Table 5.1).

Both ACP 1 and 2 were found to be highly expressed in trichomes with ACP 2

showing higher expression than ACP 1 (chapter 2). Furthermore, ACP 2 is

phylogenetically more similar to the *Coriandrum sativum* ACP (chapter 2) that is

involved in the biosynthesis of petroselinic acid ($18:1^{\Delta 6}$, a specific UMFA)[34].

Based on these results, the effect of simultaneous expression of ACP 1 or 2 in

conjunction with MAD in *E. coli* was assessed.

Co-expression of geranium MAD with ACP 1 or ACP 2 in *E. coli*

Expression of a $\Delta 9$ 14:0-ACP desaturase (MAD) from *Pelargonium × hortorum* in *E. coli* results in the production of UMFAs ($16:1^{\Delta 11}$ and $18:1^{\Delta 13}$) [75]. To examine the effect of geranium ACP 1 and 2 on production of $16:1^{\Delta 11}$ and $18:1^{\Delta 13}$, each was co-expressed with MAD and compared with controls (Figure 5.4, 5.5 and 5.6). Cell lines transformed with dual construct MAD-ACP 1 showed a significant decrease in production of all the saturated and unsaturated fatty acids (Figure 5.4, 5.5, Table 5.2). Expression of MAD-ACP 1 not only reduced the overall mass of both UMFAs but also decreased the overall percentage of UMFAs when compared to MAD alone or to MAD-ACP 2 expression lines (Figure 5.6, Table 5.3). This is not unexpected considering ACP 1 alone (in the absense of MAD) also resulted in reduced FAME production in these cell lines (Figure 5.3).

Cell lines transfomed with dual contruct MAD-ACP 2 showed significant increases in C12:0, C18:0, $16:1^{\Delta 9}$, $16:1^{\Delta 11}$ (UMFA),$18:1^{\Delta 11}$ and the total amount of unusual monoenes produced as compared to MAD alone or MAD-ACP 1 (Figure 5.4, Table 5.2).
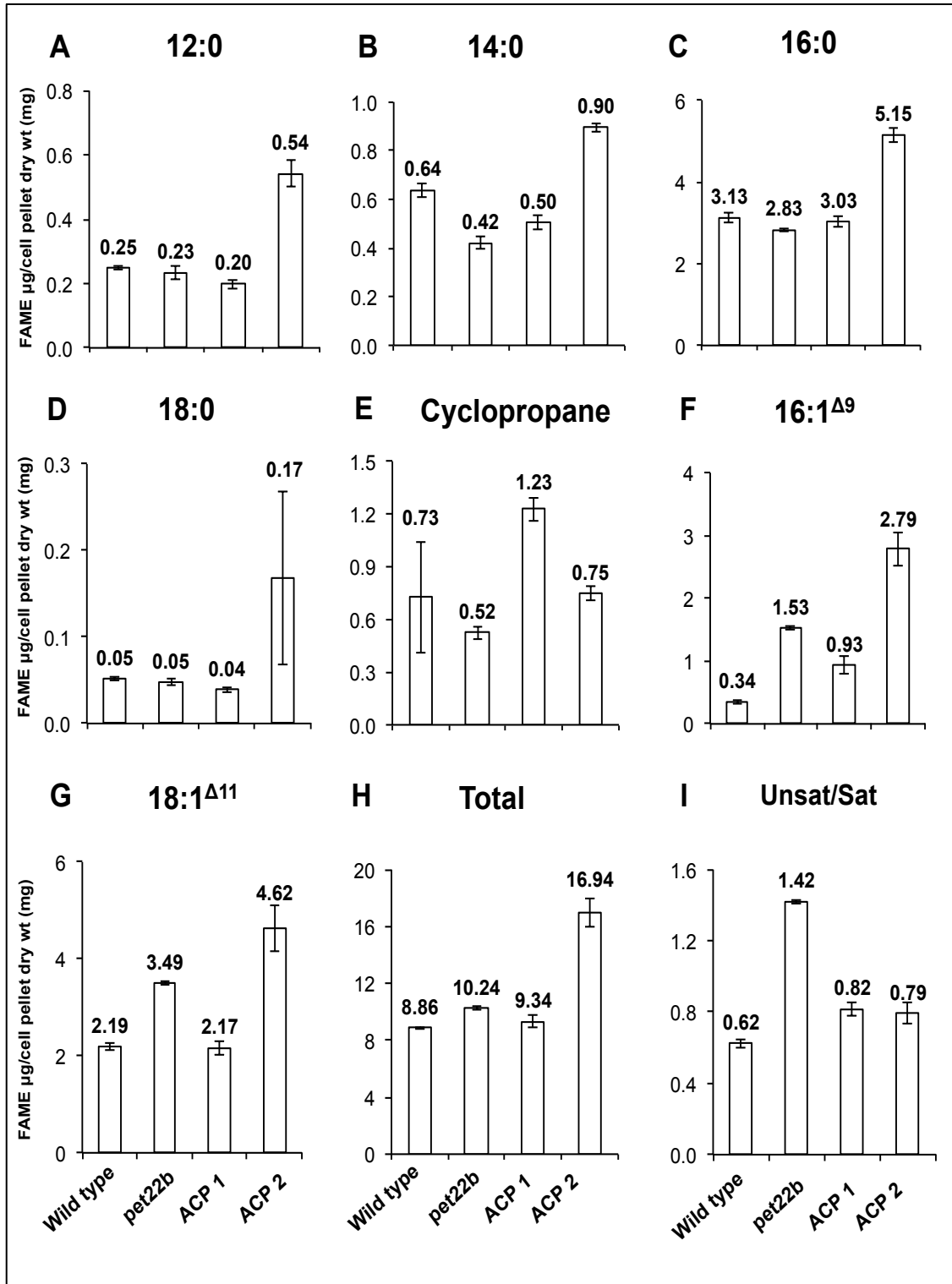
Figure 5.3. Fatty acid methyl ester content of Rosetta DE3PlysS expressing geranium ACP cDNAs. Average values (n=3) are indicated with error bars representing standard deviation of wild type (Rosetta DE3PLysS) and pet22b (vector alone), and lines expressing ACP 1 or ACP 2 in pet22b.

Table 5.1. FDR (False discovery rate) p-values for fatty acid content of *E. coli* expressing geranium ACP 1 and 2 cDNAs.

Fatty acids	FDR pvalues for comparison of each treatment					
	ACP 1 vs. ACP 2	Wild Type vs. ACP 1	pet22b vs. ACP 1	Wild Type vs. ACP 2	pet22b vs. ACP 2	Wild Type vs. pet22b
C12:0	0	0.01	0.25	0	0	1
C14:0	0	0	2.09E-07	0	0	0
C16:0	0	1	0.24	0	0	0.01
C16:1$^{\Delta 9}$	0	1.32E-08	4.19E-09	0	0	0
Cyclopropane	2.20E-05	2.20E-04	6.69E-09	0.99	0.38	0.83
C18:0	0.28	1	0.99	0.74	0.38	1
C18:1$^{\Delta 11}$	0	1	1.95E-14	0	1.11E-10	3.53E-14
Total	0	1	0.11	0	0	5.21E-03
Unsaturated/Saturated	0.02	0.0940207	1.36E-04	7.54E-04	7.70E-04	1.69E-04

FDR p-values are for one to one comparison of each treatment using two factor mixed model analysis. Values <0.05 show significant difference in FAMEs between cell lines. Rows that show FDR p-value as 0 are highly significant.

MAD-ACP 2 also showed an increase in total UMFA but also an increase in total saturated fatty acids such that the ratio of unsaturated/ saturates is not different than MAD alone or MAD-ACP 1 and (Figure 5.5, Table 5.2). Furthermore the ratio of UMFAs to unsaturated fatty acids was slightly higher in MAD-ACP 2 lines but not significantly different from MAD alone or MAD-ACP 1 (Figure 5.5,Table 5.2).

Additionally, the overall percentage of UMFAs for MAD-ACP 2 compared to MAD-ACP 1 is significantly higher but not different than MAD alone (Figure 5.6, Table 5.1). This indicates that MAD-ACP 2 is not preferentially shifting the FA profile to increased UMFA in comparison to MAD alone, but is increasing the overall mass of fatty acids.

Figure 5.4. Fatty acid methyl ester content of Rosetta DE3PlysS co-expressing geranium MAD with ACPs cDNAs set 1. Average values (n=3) are indicated with error bars representing standard deviation of wild type (Rosetta DE3PLysS) and pet3d (vector alone), and lines expressing MAD, MAD- ACP 1 and MAD-ACP 2.

Figure 5.5. Fatty acid methyl ester content of Rosetta DE3PlysS co-expressing geranium MAD with ACPs cDNAs set 2. Average values (n=3) are indicated with error bars representing standard deviation of wild type (Rosetta DE3PLysS) and pet3d (vector alone), and lines expressing MAD, MAD- ACP 1 and MAD-ACP 2.

Table 5.2. FDR p-values for fatty acid content of *E.coli* co-expressing geranium
MAD with geranium ACP 1 and 2 cDNAs.

FDR p-values for comparison of each treatment

Fatty Acids	MAD vs. MAD-ACP 1	MAD vs. MAD-ACP 2	Wild type vs. MAD	pet3d vs. MAD	MAD-ACP 2 vs. MAD-ACP 1	Wild type vs. MAD-ACP 1	pet3d vs. MAD-ACP 1	Wild type vs. MAD-ACP 2	pet3d vs. MAD-ACP 2	Wild type vs. pet3d
C12:0	0	0	0.99	3.85E-09	0	0	1.60E-04	6.22E-14	0	9.63E-10
C14:0	0	0	0	1.55E-11	0	0	0	1.71E-12	0	0
C16:0	0	0	0	0	0	0	0	0.39	0	0
C16:1$^{\Delta9}$	0	0	1.46E-11	8.88E-16	0	0.99	0	0	0	0
C16:1$^{\Delta11}$	1.52E-13	0	NA	NA	0	NA	NA	NA	NA	NA
Cyclopropane	1.52E-13	0	0.99	0.91	1.28E-07	4.34E-05	2.00E-03	0.84	0.28	0.87
C18:0	1.52E-13	0	0	0.12	0	0	0	0	0	5.62E-10
C18:1$^{\Delta11}$	1.52E-13	0	0	0.91	0	0	0	0	0	0
C18:1$^{\Delta13}$	0	7.68E-10	NA	NA	0	NA	NA	NA	NA	NA
Total	0	7.68E-10	0	2.00E-03	0	6.54E-11	0	0	0	4.19E-07
Unsat / Sat	1.40E-06	3.50E-05	0	0	0.78	2.10E-06	1.00E-03	1.00E-07	4.00E-03	1.60E-06
Sum of UMFAs	6.28E-03	0.04	NA	NA	4.70E-04	NA	NA	NA	NA	NA
UMFAs / Unsat	3.17E-01	0.84	NA	NA	0.15	NA	NA	NA	NA	NA
C16:1$^{\Delta11}$ / C16:1$^{\Delta9}$	0.25	0.06	NA	NA	0.53	NA	NA	NA	NA	NA
C18:1$^{\Delta13}$/ C18:1$^{\Delta11}$	9.60E-07	3.00E-03	NA	NA	0.38	NA	NA	NA	NA	NA

FDR p-values are for one to one comparison of each treatment using two factor mixed
model analysis. Values <0.05 show significant difference in fatty acid amount between
two genes. Rows that show FDR p-value as 0 are highly significant.
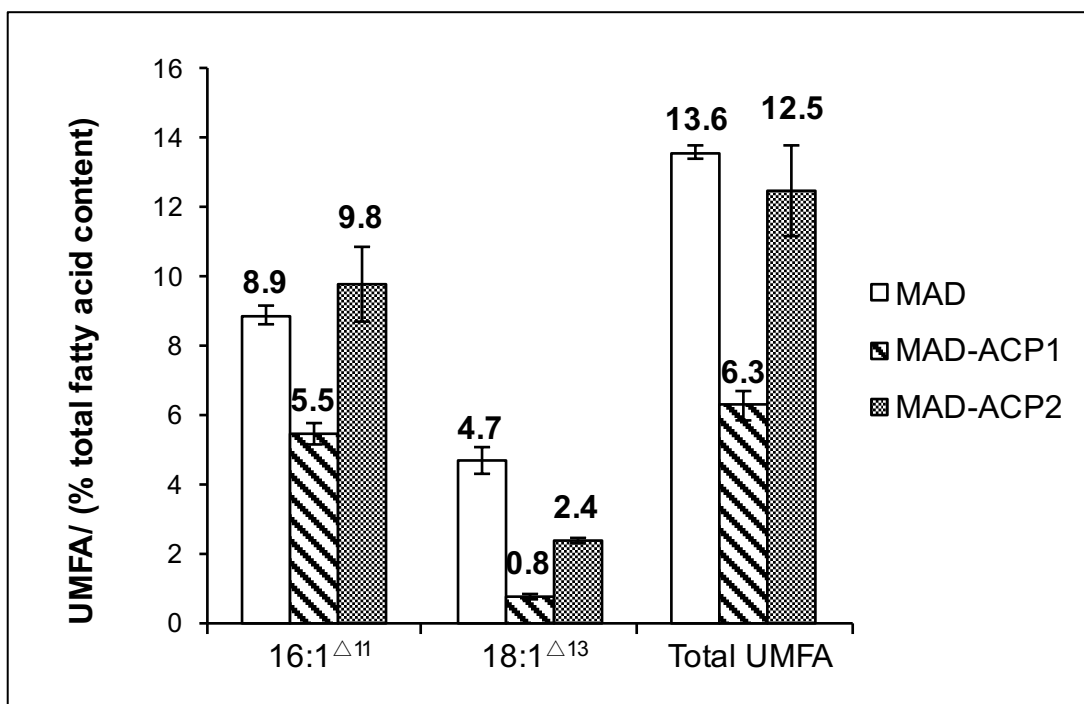
Figure 5.6. UMFA methyl ester content of Rosetta DE3PlysS co-expressing geranium MAD and ACP 1 and 2. Y-axis represents the percentage of UMFA produced as compared to total fatty acid content (mean $\pm$ SEM). Numerals at end of each bar represent the average percentage of UMFA in different treatments(*E. coli* containing MAD, MAD-ACP 1 and MAD-ACP 2). X-axis represents UMFAs, $16:1^{\Delta 11}, 18:1^{\Delta 13}$ and sum of both UMFAs.

Table 5.3. FDR p-values for and UMFA content of *E. coli* co-expressing geranium MAD with geranium ACP cDNAs.

	FDR pvalues for comparison of each treatment		
Fatty Acids	MAD vs MAD-ACP1	MAD vs MAD-ACP2	MAD-ACP2 vs MAD-ACP1
$C16:1^{\Delta 11}$	5.80E-10	0.13	2.10E-06
$C18:1^{\Delta 13}$	4.80E-10	1.30E-08	1.50E-12
Sum of UMFAs	2.30E-15	0.06	1.90E-07

FDR p-values are for one to one comparison of each treatment using two factor mixed model analysis. Values <0.05 show significant difference in fatty acid amount between two genes.

Co-expression of Geranium MAD with ACP 1 and 2 in Tobacco

To study the biochemical effects of geranium ACP 1 and ACP 2 in a model plant system, transgenic tobacco lines were generated with using expression vector pRI 201 AN empty vector, pRI 201-AN-MAD, pRI 201-AN-MAD-ACP 1 and pRI 201-AN-MAD-ACP 2 . Fatty acid analysis was performed on seeds and leaves obtained from primary transformants (Figure 5.7 and 5.8; Table 5.4 and 5.5). The variation in fatty acid profiles of different treatments was not significant (in most cases) due to high amount of variability amongst biological replicates. The influence of ACP 1 or 2 expression on UMFA biosynthesis could not be properly assessed due absence of UMFA when MAD was expressed. It is possible that UMFA were either produced at undetectable levels since these are primary transformants that may have low gene dosage resulting in lower transgene expression or UMFAs were not produced due to loss of transgene expression[207, 208]. Previous studies have also demonstrated limited accumulation of UMFAs in transgenic plants potentially due to lack of other enzyme components required for UMFA production [94]. Additionally, decreased $18:1^{\Delta 9}$ in the transgenic leaves expressing MAD alone and MAD with ACP 1 or 2 and decreased unsaturated fatty acids of seeds expressing MAD indicated a metabolic effect of MAD expression and may indicate the possibilty of beta-oxidation as previously repeated [94] (Figure 5.7 and 5.8; Table 5.4 and 5.5).
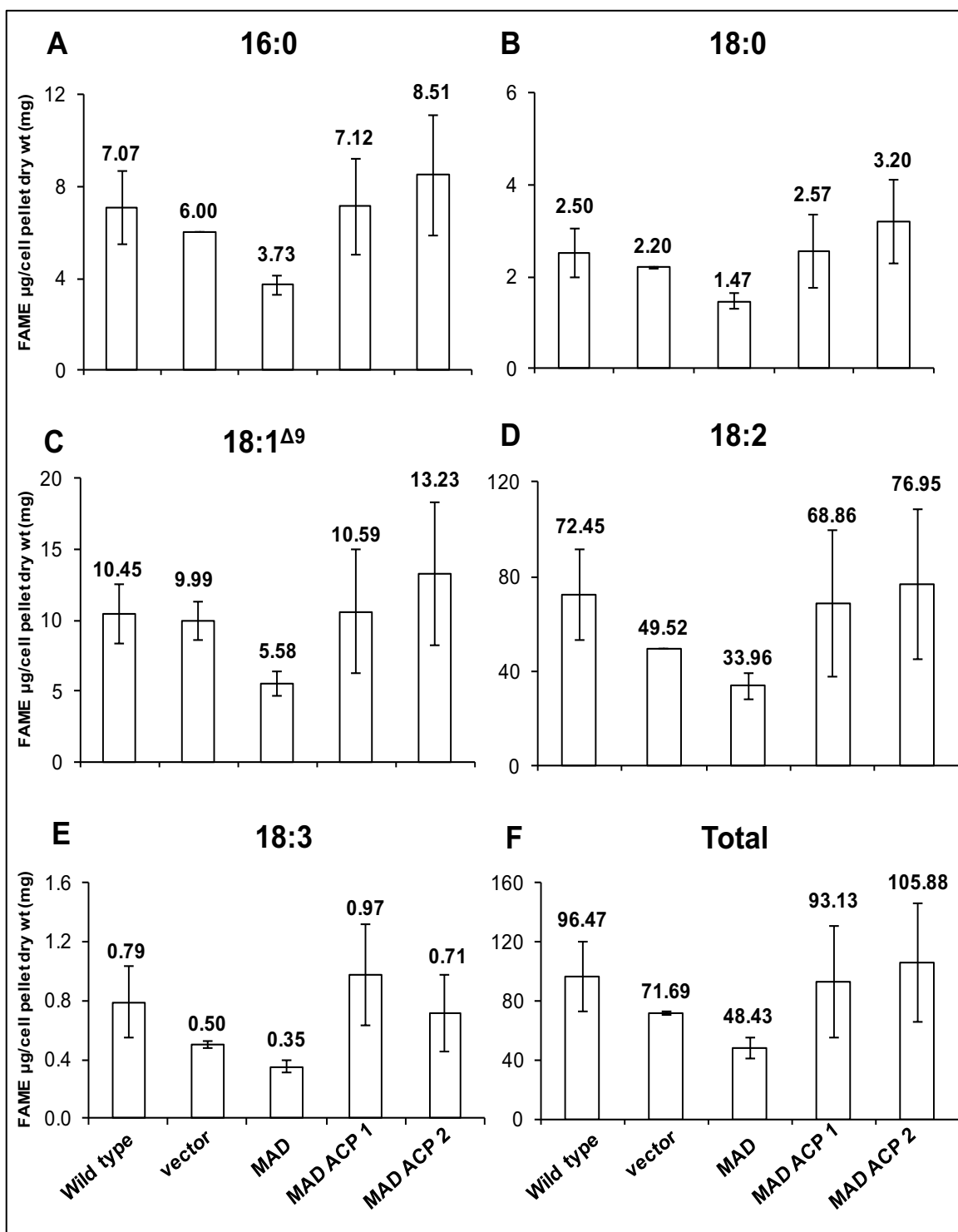
Figure 5.7. Fatty acid methyl ester content of tobacco seeds co-expressing geranium MAD with geranium ACP cDNAs. Average values (n=3) are indicated with error bars representing standard deviation of wild type (tobacoo) and pRI201-AN (vector alone), and transgenic plants expressing MAD, MAD- ACP 1 and MAD-ACP 2.

Table 5.4. FDR p-values for fatty acid content of tobacco seeds co-expressing geranium MAD with geranium ACP cDNAs.

FDR pvalues for comparison of each treatment

Fatty Acids	MAD vs. MAD-ACP 1	MAD vs. MAD-ACP 2	Vector vs. MAD	Wild Type vs. MAD	MAD-ACP 2 vs. MAD-ACP 1	Vector vs. MAD-ACP 1	Wild Type vs. MAD-ACP 1	Vector vs. MAD-ACP 2	Wild Type vs. MAD-ACP 2	Wild type vs. vector
C16:0	0.06	4.80E-04	0.97	0.07	0.98	0.99	0.99	0.56	0.99	0.99
C18:0	0.09	4.80E-04	0.97	0.08	0.98	0.99	0.99	0.56	0.99	0.99
C18:1$^{\Delta 9}$	0.12	5.20E-03	0.97	0.14	0.98	0.99	0.99	0.57	0.99	0.99
C18:2	0.12	0.02	0.97	0.08	0.98	0.99	0.99	0.57	0.99	0.99
C18:3	1.00E-03	0.09	0.97	0.07	0.98	0.13	0.99	0.57	0.99	0.99
Total	0.12	0.01	0.97	0.08	0.98	0.99	0.99	0.57	0.99	0.99

FDR p-values are for one to one comparison of each treatment using two factor mixed model analysis. Values <0.05 show significant difference in fatty acid amount between two genes.

Table 5.5. FDR p-values for fatty acid content of tobacco leaves co-expressing geranium MAD with geranium ACP cDNAs.

FDR pvalues for comparison of each treatment

Fatty Acids	MAD vs. MAD-ACP 1	MAD vs. MAD-ACP 2	Vector vs. MAD	Wild Type vs. MAD	MAD-ACP 2 vs. MAD-ACP 1	Vector vs. MAD-ACP 1	Wild Type vs. MAD-ACP 1	Vector vs. MAD-ACP 2	Wild Type vs. MAD-ACP 2	Wild type vs. vector
C16:0	0.48	0.89	0.99	0.14	0.99	0.17	0.99	0.79	0.99	0.13
C18:0	1.81E-05	0.92	0.99	0.037	2.95E-07	3.78E-06	0.83	0.99	0.01	0.02
C18:1$^{\Delta 9c}$	0.75	0.34	4.40E-05	8.58E-05	0.99	0.03	0.04	0.47	0.5	0.99
C18:2	0.98	0.89	0.99	0.46	0.99	0.97	0.99	0.99	0.99	0.99
C18:3	0.75	0.89	0.99	0.02	0.99	0.97	0.96	0.99	0.77	0.3
Total	0.75	0.92	0.99	0.16	0.99	0.87	0.99	0.99	0.99	0.3

FDR p-values are for one to one comparison of each treatment using two factor mixed model analysis. Values <0.05 show significant difference in fatty acid amount between two genes.
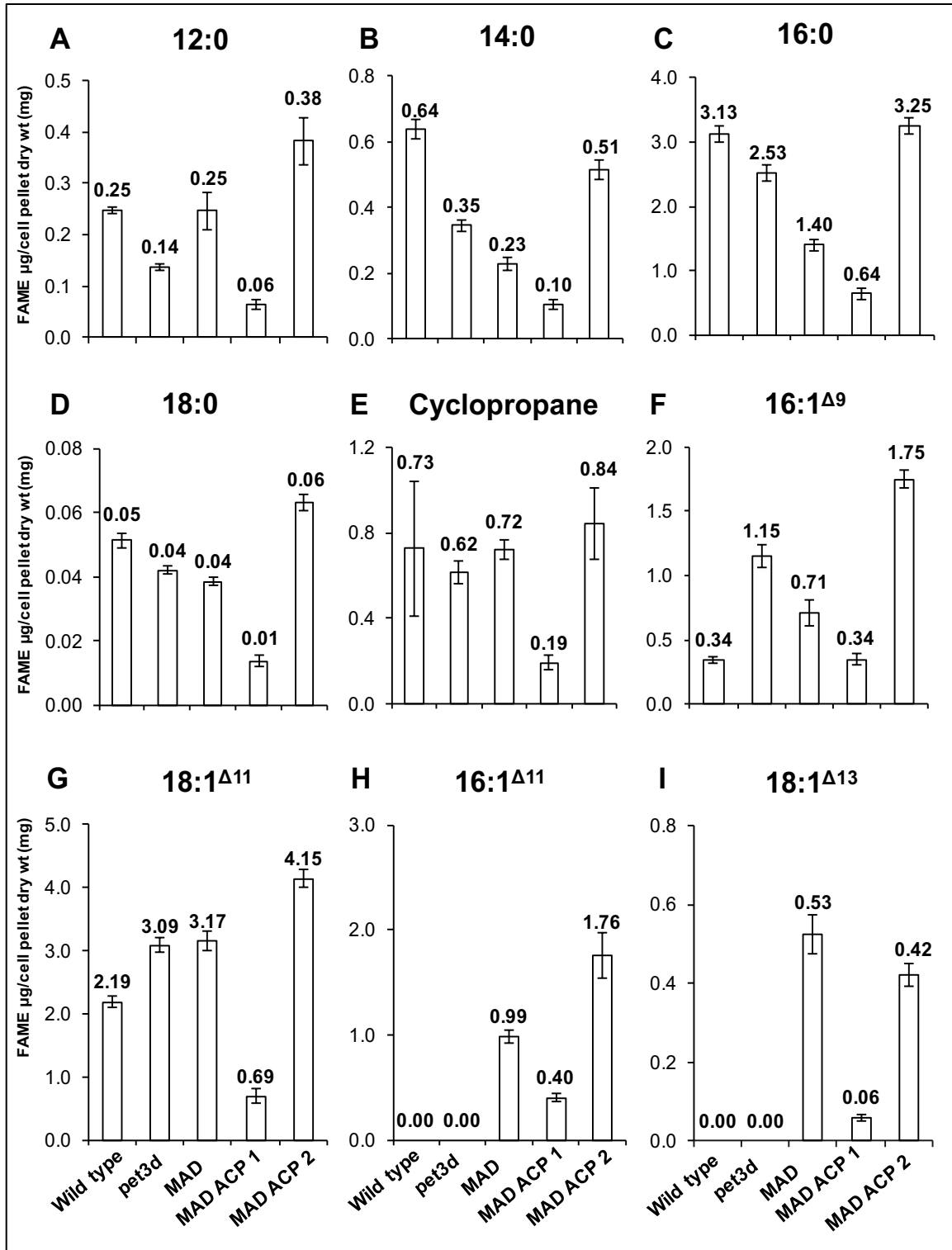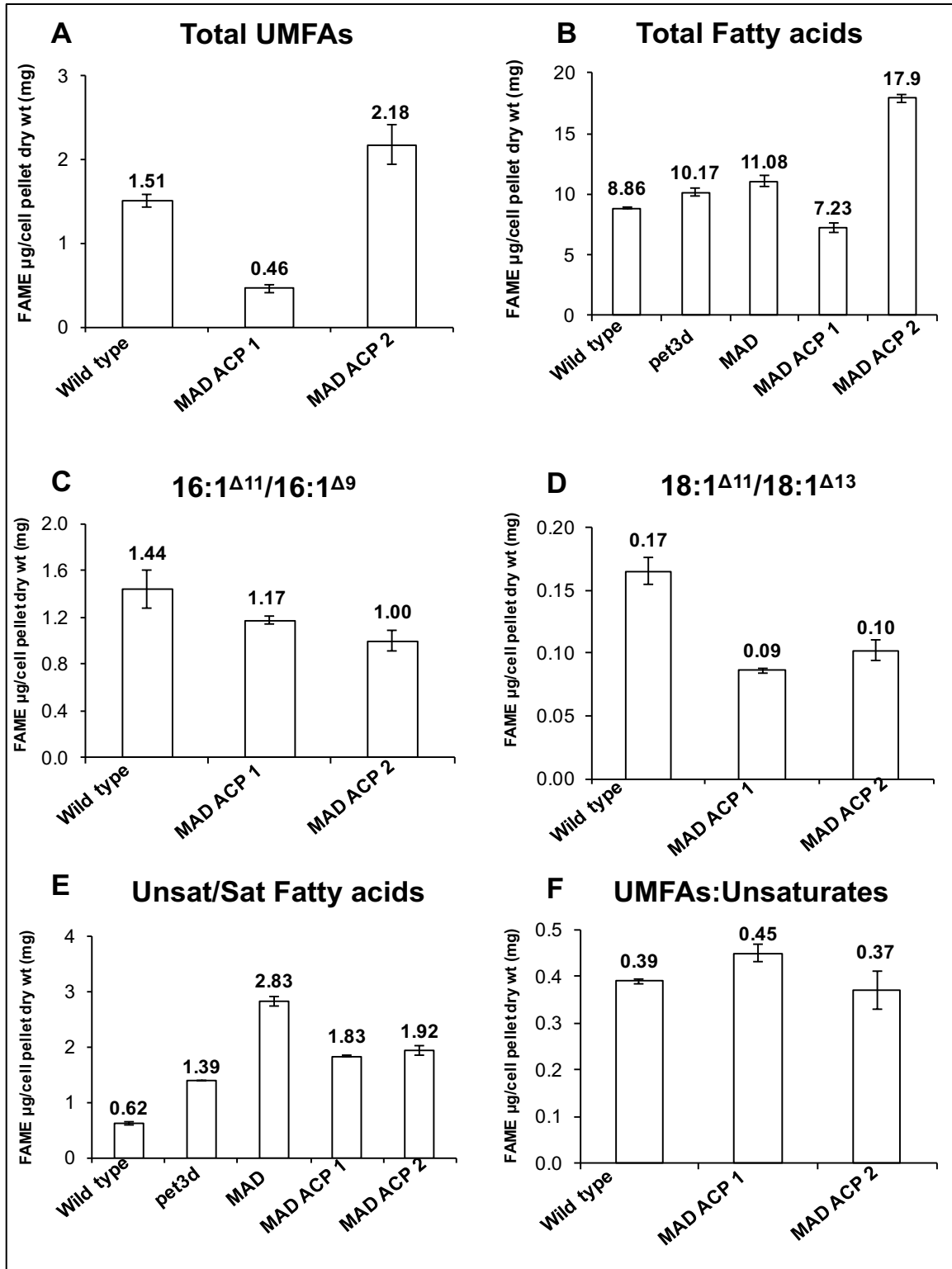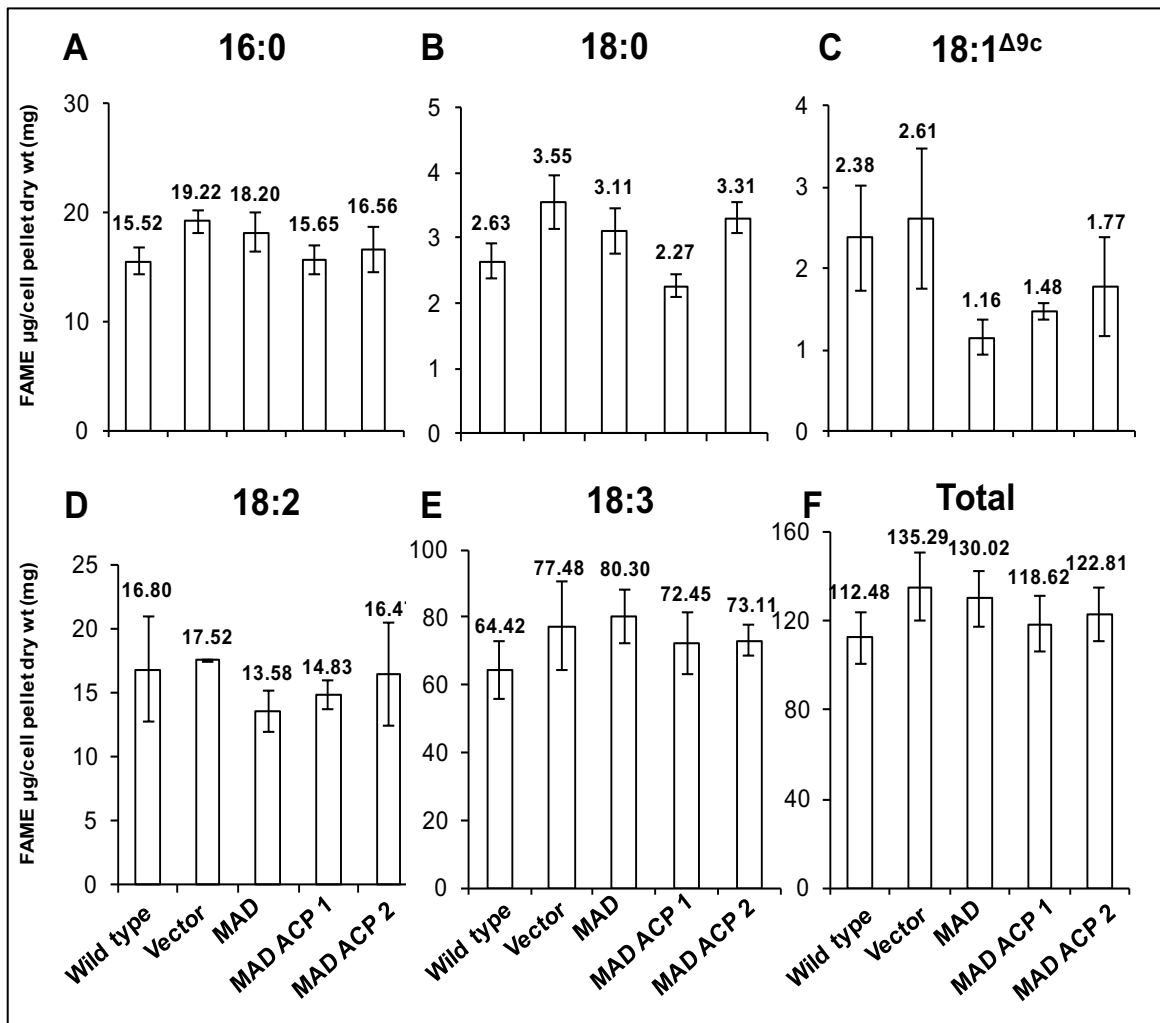
Figure 5.8. Fatty acid methyl ester content of tobacco leaves co-expressing geranium MAD with geranium ACP cDNAs. Average values (n=3) are indicated with error bars representing standard deviation of wild type (tobacoo) and pRI201-AN (vector alone), and transgenic plants expressing MAD, MAD- ACP 1 and MAD-ACP 2.

Conclusion

E. coli assays showed when ACP 1 was expressed in conjunction with MAD, it lowered both the mass and percentage (UMFA/total lipids) of UMFA compared to expression of MAD alone or MAD with ACP 2. This indicates ACP 1 may not be

associated to UMFA synthesis in geranium. However, ACP 2 was expressed in conjunction with MAD and increased the production (mass) of UMFAs as compared to expression of MAD alone. However the percentage of (UMFA/total lipids) of UMFA was not significantly different from MAD alone. ACP 2 is highly expressed in trichomes of geranium and its expression is correlated with changes in AnAc and UMFA production at different temperatures (Chapter 2, Chapter 3). ACP 2 is also phylogenetically closer to the *Coriandrum sativum* ACP that is involved in the biosynthesis of petroselinic acid $18:1^{\Delta 6}$, a specific UMFA (Chapter 2). Collectively, these results provide evidence for ACP 2 being a novel isoform involved in UMFA synthesis in trichomes of geranium.

In this study, transgenic tobacco plants expressing geranium MAD along with ACP 1 and 2 did not lead to a production of UMFAs or show any significant changes in fatty acid profile. Limited accumation of UMFAs in trangenic plants has been observed in literature and these results may be due to low gene dosage in the primary transformants or beta oxidation of fatty acids[5, 94, 209]. To confirm these possibilites, this experiment should be further investigated by evaluating the fatty acid profile of the seeds and leaves obtained from the next generation of primary transformants. It is possible that selection of homozygous progeny from next generation may increase the gene dosage and show production of UMFAs and help in understanding the roles of ACPs in UMFA production. Alternatively, ACP function could be assessed by *in vitro* biochemical

assays analyzing substrates specificity of ACPs or using another model plant like

Arabidopsis for transgenic assays.

CHAPTER 6

CONCLUSIONS AND FUTURE DIRECTIONS

Conclusions

Unusual monoenoic fatty acids (UMFAs) are interesting to the plant

biotechnology community due to their wide variety of applications as polymers,

fuels, nutraceuticals, medicine and renewable sources of energy[4]. The research

focus of this dissertation was to study the biosynthesis of specific UMFAs

($16:1^{\Delta 11}$ and $18:1^{\Delta 13}$) in trichomes (hair-like structures) of *Pelargonium ×*

hortorum (garden geranium). Production of these UMFAs in geranium is

particularly interesting because they are substrates for the production of a

secondary metabolite called anacardic acid (AnAc). More specifically, the

$16:1^{\Delta 11}$ and $18:1^{\Delta 13}$ UMFAs are precursors to the AnAc 22:1n5 and AnAc 24:1n5

congeners that confer pest resistance to garden geranium. More broadly, AnAc

are also known to have antifungal, antibacterial and anticancer activities and

have industrial applications[56].

Geranium glandular trichomes are highly specialized for production of both

UMFAs and AnAc and are readily isolated as pure cell preparations thus

providing an ideal model tissue for exploration of the underlying genetics and

biochemistry of UMFA and AnAc synthesis. To further investigate the biosynthesis of $16:1^{\Delta 11}$ and $18:1^{\Delta 13}$ in geranium trichomes, novel isoforms of the fatty acid biosynthesis enzymes including acyl carrier protein (ACP), 3-ketoacyl-ACP synthase (KAS) and acyl-ACP thioesterase (TE) were isolated and characterized. Expression profiles of these sequences were analyzed for the expected trichome specificity and for correlation of changes in $16:1^{\Delta 11}$ and $18:1^{\Delta 13}$ as well as changes in derived AnAc 22:1n5 and 24:1n5 in relation to changes in temperature.

Identification of FAS enzymes from geranium EST database and testing tissue specificity

Using diverse molecular approaches, a total of 14 FAS enzyme sequences were identified. These included 2 ACP, 3 FAT-A, 3 FAT-B, 4 KAS I, 1 KAS II and 1 KAS III (Appendix 6, Table 6.1). Phylogenetic analysis was used to determine potential similarity of these sequences to other FAS enzyme sequences known to be involved in UMFA biosynthesis in plants. These results suggested that ACP 2, KAS Ic, and FAT-A1 protein sequences were in the same clade as the respective enzymes ACP, KAS I and FAT-A from *Coriandrum sativum* that are involved in the biosynthesis of the UMFA petroselinic acid ($18:1^{\Delta 6}$). Thus these cDNA sequences could potentially be trichome specific isoforms that involved in the synthesis of $16:1^{\Delta 11}$ and $18:1^{\Delta 13}$.

Since the UMFAs are only produced in glandular trichomes, expression of specific genes involved in this process were expected to display the same tissue-specific pattern. This approach is supported by the identification of a highly trichome specific novel $\Delta 9$ 14:0-acyl carrier protein (ACP) desaturase (MAD) gene. The MAD gene is responsible for producing myristoleic acid ($14:1^{\Delta 9}$) that is elongated into $16:1^{\Delta 11}$ and $18:1^{\Delta 13}$ UMFAs. Thus, qRT-PCR was used to assess trichome specific expression of target FAS sequences. Results of qRT-PCR indicated that ACP 1, ACP 2, KAS Ia/b, KAS Ic, FAT-A1, and FAT-A2 are each highly expressed in trichomes compared to the bald pedicle, further suggesting their potential role in trichome UMFA metabolism.

Effect of temperature on production of AnAc, UMFAs and gene expression

RNA-SEQ and qRT-PCR were used to further investigate the association of specific FAS enzyme expression with biosynthesis of $16:1^{\Delta 11}$ and $18:1^{\Delta 13}$ and derived AnAc 22:1n5 and 24:1n5 at distinct temperatures. A *de novo* RNA transcriptome was generated from trichomes and bald pedicle tissue of geranium at 18°C and 23°C (3 biological replicates/tissue/condition) to assess the effect of temperature on gene expression and qRT-PCR was used to independently analyze expression at 18°C, 23°C and 28°C. RNA-SEQ resulted in 486398 contigs from the transcriptome and this provided valuable insights into genes associated with UMFA metabolism. Examples of these include validation of previously identified FAS genes, identification of complete sequences of FAS genes and identification of new FAS enzyme sequences. A few notable

139

examples of this RNA-Seq analysis were the identification of a new FAT-A that is highly expressed in trichomes and showed an increase in expression at 23°C and the identification of new sequences corresponding to reductases, hydroxylases, FAT-s, KASs, PKSs, and KCSs. Target ACPs, KASs, and TEs showed a trend for an increase in expression at 23°C that was further tested and validated through qRT-PCR.

In addition to trichome and bald pedicle samples at 18°C and 23°C, samples obtained at 28°C were also utilized for qRT-PCR analyses. UMFAs ($16:1^{\Delta11}$ and $18:1^{\Delta13}$) and AnAc (22:1n5 and 24:1n5) extracted from all the samples were analyzed by GC-FID and HPLC-DAD respectively. Results from these analyses indicated that not only is 23°C the optimal temperature for UMFA and AnAc synthesis compared to 18°C and 28°C but also indicated higher temperature negatively affects the production of metabolites. Production of UMFAs ($16:1^{\Delta11}$ and $18:1^{\Delta13}$) was positively correlated with production of AnAc (22:1n5 and 24:1n5) at all temperatures indicating a relationship between the substrate and the metabolite at a given temperature. Interesting the ratios of AnAc (24:1n5/ 22:1n5) and UMFAs ($18:1^{\Delta13}$/$16:1^{\Delta11}$) did not how corelation at 28°C and the amount AnAc 24:1n5 was more than AnAc 22:1 at all three temperatures whereas the amount AnAc 24:1n5 substrate $18:1^{\Delta13}$ was less compared to $16:1^{\Delta11}$ at all three temperatures. This indicates a possibility of substrate preference for $18:1^{\Delta13}$ by the downstream enzymes for production of AnAc 24:1n5. Finally, expression of ACP 1, ACP 2, KAS I-a/b, and KAS I-c were

significantly correlated with production of target metabolites at specific

temperatures (18°C, 23°C and 28°C), further validating their potential

involvement in UMFA metabolism.

De novo micro-RNA database

At the same time the *de novo* RNA-Seq transcriptome was assembled, a *de novo*

micro-RNA database was generated from trichomes at 18°C and 23°C and bald

pedicle at 23°C (1 replicate/ sample) to identify regulatory genetic factors

involved in UMFA metabolism. Detailed correlative analysis of miRNAs to

transcriptome and metabolites changes was beyond the scope of this

dissertation. However, these data are the first known catalog of miRNA

sequences in glandular trichomes and in geranium. These sequences can

provide a starting resource to further explore the regulatory roles of miRNA in

geranium trichome metabolism, physiology and development. A total of 337

micro-RNAs were identified in the *de novo* micro-RNA transcriptome and their

potential targets were predicted using a Plant Small RNA Target Analysis Server

(psRNA). We have also identified specific micro-RNAs that are highly expressed

in trichomes and can be chosen as candidates for further study based on target

predictions provided by the psRNA server.

Co-expression of geranium ACPs with MAD in *E. coli* and tobacco

To further elucidate potential function of ACPs in UMFA synthesis, both ACP 1

and ACP 2 were co-expressed with MAD in *E. coli* and tobacco. Primary

transformants of tobacco (both leaves and seeds) obtained from the transgenic

plant assay did not show production of UMFAs or any significant changes in the

fatty acid profile of transformants as compared to wild type plants. Limited

accumation of UMFAs in trangenic plants has been observed in literature and

these results may be due to low gene dosage in the primary transformants or

beta oxidation of fatty acids[5, 94, 209]. To confirm these possibilites, this experiment

should be further investigated by evaluating the fatty acid profile of the seeds

and leaves obtained from the next generation of primary transformants. It is

possible that selection of homozygous progeny from next generation may

increase the gene dosage and show production of UMFAs and help in

understanding the roles of ACPs in UMFA production. Another option is using a

different model plant like *Arabidopsis* for trangenic assays.

E. coli assay results indicated that ACP 1, when when expressed with MAD lead

to reduction in total fatty acids and ACP 2, when expressed with MAD, increased

the production of UMFAs and overall fatty acid biosynthesis as compared to

expression of MAD alone. Neither ACPs was found to affect the specific relative

proportion of $16:1^{\Delta 11}$ or $18:1^{\Delta 13}$ UMFAs ACP 2 is not only highly expressed in

trichomes of geranium but its expression is correlated with changes in AnAc and

UMFA production at different temperatures. Also, ACP 2 is phylogenetically

closer to *Coriandrum sativum* ACP that is involved in the biosynthesis of

petroselinic acid ($18:1^{\Delta 6}$, a specific UMFA). Collectively, these results provide

strong evidence for ACP 2 being a novel isoform involved in UMFA synthesis in trichomes of geranium.

Future Considerations

The *de novo* RNA and micro-RNA transcriptomes generated for geranium are novel platforms for identification of enzyme encoding sequences and transcription factors involved in trichome metabolism (UMFA and AnAc synthesis). Transcriptome data can be further analyzed using bioinformatics tools like IsoSVM to identify isoforms versus paralogs to study the metabolic regulation of genes based on tissue, stage or time specific plant development. Temperature studies done for this research can be extended to evaluate other environmental factors such as light intensity, moisture, photoperiod, humidity, temperature and soil quality to assess the influence of environmental condition on production of AnAc and UMFAs.

Studies invovling identification of novel trichome-specific ACP 2 that increases the production of UMFA when expressed with MAD can be extended to identifying other novel FAS enzymes involved in the same process. To further this work we have designed and optimized primers for expression of FAT-As and KASs in both *E. coli* and tobacco. Additionally, primers have also been designed and optimized for expression of enzymes like polyketide synthases and 3-

ketoacyl-CoA-synthases (that are potentially involved in AnAc synthesis) in plants.

Furthermore, small-scale expression and purification of ACP 1 and ACP 2 has been optimized and could serve to help generate acyl-ACP substrates for biochemical assays or for structural studies by X-ray crystallography. Such protein studies can provide valuable insights into substrate specificity, binding affinity, structure and enzymatic interactions of geranium ACPs and similar FAS enzymes, thus expanding knowledge about their role in UMFA synthesis. The information obtained from the future research discussed above can potentially help in demonstrating the utility of metabolic shuttling of specific enzyme isoforms, where a single gene, in the presence or absence of additional specific enzymes, could produce three distinct fatty acids ($14:1^{\Delta 9}$, $16:1^{\Delta 11}$ and $18:1^{\Delta 13}$). Furthermore, this will also help in producing UMFAs and/or AnAc at a higher level in commodity and transgenic crops with associated applications in the food, agriculture, medicinal, chemical and energy industries.

REFERENCES

1. John Ohlrogee, J.B. Lipid Biosynthesis. *The Plant Cell* **7** (1995).
2. Moire, L., Rezzonico, E., Goepfert, S. & Poirier, Y. Impact of unusual fatty acid synthesis on futile cycling through beta-oxidation and on gene expression in transgenic plants. *Plant Physiol* **134**, 432-442 (2004).
3. Anthony A. Millar, M.A.S.a.L.K. <All fatty acids are not equal-discrimination in plant membrane lipids.pdf>. *Trends in plant science* **5** (2000).
4. Napier, J.A. The production of unusual fatty acids in transgenic plants. *Annual review of plant biology* **58**, 295-319 (2007).
5. Roscoe, T., Maisonneuve, S. & Delseny, M. Production of unusual fatty acids in rapeseed. *Oléagineux, Corps gras, Lipides* **9**, 24-30 (2002).
6. Rogalski, M. & Carrer, H. Engineering plastid fatty acid biosynthesis to improve food quality and biofuel production in higher plants. *Plant Biotechnol J* **9**, 554-564 (2011).
7. Millar, A.A., Smith, M.A. & Kunst, L. All fatty acids are not equal: discrimination in plant membrane lipids. *Trends in plant science* **5**, 95-101 (2000).
8. Suh, M.C., Schultz, D.J. & Ohlrogge, J.B. What limits production of unusual monoenoic fatty acids in transgenic plants. *Planta* **215**, 584-595 (2002).
9. Suh, M.C., Schultz, D.J. & Ohlrogge, J. Isoforms of acyl carrier proteins involved in seed-specific fatty acid synthesis. *Plant Journal* **17**, 679-688 (1999).
10. Li-Beisson, Y., Nakamura, Y. & Harwood, J. Lipids: From Chemical Structures, Biosynthesis, and Analyses to Industrial Applications. *Subcell Biochem* **86**, 1-18 (2016).
11. Suh, M.C., Schultz, D.J. & Ohlrogge, J.B. What limits production of unusual monoenoic fatty acids in transgenic plants? *Planta* **215**, 584-595 (2002).
12. Bates, P.D., Stymne, S. & Ohlrogge, J. Biochemical pathways in seed oil synthesis. *Current opinion in plant biology* **16**, 358-364 (2013).
13. White SW, Z.J., Zhang YM, Rock The structural biology of type II fatty acid biosynthesis. *Annu Rev Biochem* **45** (2005).
14. Harwood, J.L. Recent advances in the biosynthesis of plant fatty acids. *Biochimica et Biophyscia acta* **1310**, 7-56 (1996).

15. Rawsthrone, S. Carbon flux and fatty acid synthesis in plants. *Progress in Lipid Research* **41**, 182-196 (2002).

16. Bates, P.D., Johnson, S.R., Cao, X., Li, J., Nam, J.W., Jaworski, J.G., Ohlrogge, J.B. & Browse, J. Fatty acid synthesis is inhibited by inefficient utilization of unusual fatty acids for glycerolipid assembly. *Proc Natl Acad Sci U S A* **111**, 1204-1209 (2014).

17. Byers, D.M. & Gong, H. Acyl carrier protein structure-function relationships in a conserved multifunctional protein family. *Biochemistry and cell biology = Biochimie et biologie cellulaire* **85**, 649-662 (2007).

18. Alenka Hlousek-Radojcic, D.P.-B.a.J.O. Expression of Constitutive and Tissue-Specific Acyl Carrier Protein Isoforms in Arabidopsis. *Plant Physiol* **98**, 206-214 (1992).

19. John B Ohlrogee, J.G.J. Regulation of fatty acid synthesis. *Annu Rev Plant Physiol. Plant Mol.Biol.* **48**, 109-136 (1997).

20. Thelen, J.J. & Ohlrogge, J.B. Metabolic engineering of fatty acid biosynthesis in plants. *Metabolic engineering* **4**, 12-21 (2002).

21. Shanklin, J.a.C., E.B. <DESATURATION AND RELATED MODIFICATIONS OF FATTY ACIDS.pdf>. *Plant Physiol* **96**, 382-389 (1998).

22. Dehesh, K., Edwards, P., Fillatti, J., Slabaugh, M. & Byrne, J. KAS IV- a 3-ketoacyl-ACP synthase from Cuphea sp. is a medium chain specific condensing enzyme. *Plant Cell* **15**, 383-390 (1998).

23. Dehesh, K., Edwards, P., Haynes, T., Cranmer, A. & Fillatti, J. Two novel Thioesterases are key determinants of the bimodal distribution of acyl chain length of cuphea palustris seed oil. *Plant Physiology* **110**, 383-390 (1996).

24. Voelker, T.e.a. Fatty Acid Biosynthesis Redirected to Medium Chains in Transgenic Oilseed Plants *Plant Physiol* **104**, 72-74 (1992).

25. J.B., O. Design of New Plant Products- Engineering of Fatty Acid Metabolism *Plant Physiol* **104**, 821-826 (1994).

26. Knudson DS, L.K., Nelsen JS, Bleibaum JL,Davis HM, Metz JG Cloning of a Coconut Endosperm cDNA Encoding a 1 -Acyl-sn-glycerol-3-Phosphate Acyltransferase That Accepts Medium-Chain-Length Substrates *Plant Physiol* **109**, 999-1006 (1995).

27. Knutzon, D.S., Hayes, T.R., Wyrick, A., Xiong, H., Davies, H.M. & Voelker, T.A. Lysophosphatidic acid acyltransferase from coconut endosperm mediates the insertion of laurate at the sn-2 position of triacylglycerols in lauric rapeseed oil and can increase total laurate levels. *Plant physiology* **120**, 739-746 (1999).

28. Yuan, Y., Liang,Y., Li,B., Zheng,Y., Luo,X. and Li,D. Cloning and Function Characterization of a beta-Ketoacyl-acyl-ACP Synthase I from Coconut (Cocos nucifera L.) Endosperm. *Plant molecular biology* **33**, 1131-1140 (2015).

29. Millar, A.A.a.K.L. Very-long-chain fatty acid biosynthesis is controlled through the expression and specificity of the condensing enzyme *Plant J* **12**, 121-131 (1997).

30. Broun, P.e.a. Catalytic plasticity of fatty acid modification enzymes underlying chemical diversity of plant lipids. *Science* **282**, 1351-1317 (1998).

31. Schultz, D.J., Suh, M.C. & Ohlrogge, J.B. Stearoyl-acyl carrier protein and unusual acyl-acyl carrier protein desaturase activities are differentially influenced by ferredoxin. *Plant Physiol* **124**, 681-692 (2000).

32. Dormann, P., Frentzen, M. & Ohlrogge, J. Specificities of the acyl-acyl carrier protein (ACP) thioesterase and glycerol-3-phosphate acyltransferase for octadecenoyl-ACP isomers. *Plant Physiol* **104**, 839-844 (1994).

33. Mekhedov, S., Cahoon, E.B. & Ohlrogge, J. An unusual seed-specific 3-ketoacyl-ACP synthase associated with the biosynthesis of petroselinic acid in coriander. *Plant molecular biology* **47**, 507-518 (2001).

34. Cahoon, E.B., Shanklin, J. & Ohlrogge, J.B. Expression of a coriander desaturase results in petroselinic acid production in transgenic tobacco. *Proceedings of the National Academy of Sciences* **89**, 11184-11188 (1992).

35. Cahoon, E.B. & Ohlrogge, J.B. Metabolic evidence for the involvement of a [delta] 4-palmitoyl-acyl carrier protein desaturase in petroselinic acid synthesis in coriander endosperm and transgenic tobacco cells. *Plant physiology* **104**, 827-837 (1994).

36. Schultz, D.J., Wickramsinghe, N. & Klinge, C. Anacardic acid biosynthesis and bioactivity. *Recent Advances in Plant Biochemistry*, 131-156 (2006).

37. Edgar B. Cahoon, E.-F.M., Kevin L. Stecca, Sarah E. Hall, & David C. Taylor, a.A.J.K. Production of Fatty Acid Components of Meadowfoam Oil in Somatic Soybean Embryos. *Plant Physiology,* **124**, 243–251 (2000).

38. Yu, D., Xu, F., Zeng, J. & Zhan, J. Type III polyketide synthases in natural product biosynthesis. *IUBMB life* **64**, 285-295 (2012).

39. Schneider, L.M., Adamski, N.M., Christensen, C.E., Stuart, D.B., Vautrin, S., Hansson, M., Uauy, C. & von Wettstein-Knowles, P. The Cer-cqu gene cluster determines three key players in a beta-diketone synthase polyketide pathway synthesizing aliphatics in epicuticular waxes. *J Exp Bot* (2016).

40. Dayan, F.E., A. M. Rimando, et al. "Sorgoleone.". *Phytochemistry* **71**, 1032-1039 ((2010).).

41. al, S.S.e. Aromatic and pyrone polyketide synthesis by a stilbene synthase from Rheum tataricum. *Phytochemistry* **62**, 313-323 (2003).

42. Broun, P.e.a. Accumulation of ricinoleic, lesquerolic and densipolic acids in seeds of transgenic Arabidopsis plants that express a fatty acid hydroxylase cDNA from castor bean. *Plant Physiol* **113**, 933-942 (1997).

43. van de Loo FJ, e.a. An oleate 12-hydroxylase from Ricinus communis L is a fatty acyl desaturase homolog. *Proc Natl Acad Sci U S A* **92**, 6473-6447 (1995).

44. Eccleston, V.S. & Ohlrogge, J.B. Expression of lauroyl–acyl carrier protein thioesterase in Brassica napus seeds induces pathways for both fatty acid

oxidation and biosynthesis and implies a set point for triacylglycerol accumulation. *The Plant Cell* **10**, 613-621 (1998).

45. Schultz, D.J., Medford, J.I., Cox-Foster, D.L., Grazzini, R., Craig, R. & Mumma, R.O. Anacardic acid in trichomes of *Pelargonium*: Biosynthesis, molecular biology and ecological effects. *Advances in Botanical Research* **32**, 176-192 (2000).

46. Grazzini, R. & Hesk, D. Distribution of anacardic acids associated with small pest resistance among cultivars of *Pelargonium x hortorum*. *Journal of the American Society of Horticultural Science* **120**, 343-346 (1995).

47. Schultz, D.J., Cahoon, E.B., Shanklin, J., Craig, R., Cox-Foster, D.L., Mumma, R.O. & Medford, J.I. Expression of a delta 9 14:0-acyl carrier protein fatty acid desaturase gene is necessary for the production of omega 5 anacardic acids found in pest-resistant geranium (*Pelargonium* x *hortorum*). *Proceedings of the National Academy of Sciences* **93**, 8771-8775 (1996).

48. Murphy, D., Fairbairn, D. & Bowra, S. (John Innes Centre Annual Report, 1999).

49. Li J, L.M., Wu PZ, Tian CE, Jiang HW and Wu GJ. & Molecular cloning and expression analysis of a gene encoding a putative beta-ketoacyl-acyl carrier protein (ACP) synthase III (KAS III) from Jatropha curcas. *Tree Physiol.* **28**, 921-927 (2008).

50. Li, M.J., Li,A.Q., Xia,H., Zhao,C.Z., Li,C.S., Wan,S.B., Bi,Y.P. and Wang,X.J. & Cloning and sequence analysis of putative type II fatty acid synthase genes from Arachis hypogaea L. *J. Biosci.* **34**, 227-238 (2009).

51. Cahoon E.B, B.C., Shanklin J, Ohlrogee JB cDNAs for isoforms of the delta-9 stearoyl-Acyl carrier protein from Thunbergia ala endosperm. *Plant Physiology* **106**, 807-808 (1994).

52. Cahoon E.B, C.A., Shanklin J, Ohlrogee JB Delta-6 hexadecenoic acid is synthesized by the activity of a soluble delat-6 Palmitoyl acyl acrrier protein deasturase in Thunbergia alata endosperm. *The Journal of biochemical Chemistry* **269**, 27519-27526 (1994).

53. Whittle E, C.E., Subrahmanyam S, Shanklin J A multifunctional acyl-acyl carrier protein desaturase frim Hedra helix L. (English Ivy) can synthesize 16- and 8-Carbon monoene and diene products. *The Journal of biochemical Chemistry* **280**, 28169-28176 (2005).

54. Guy JF, W.E., Kumaran D, Lindqvist Y, Shanklin The crystal structure of the Ivy delta-4 16:0 desaturase reveals structural detail of the oxidized active site and potential determinants of regioselectivity. *The Journal of biochemical Chemistry* **282**, 19863-19871 (2007).

55. Cahoon E.B, C.S., Shanklin J Characterization of a structurally and functionally divereged acyl-acyl protein desaturase from milkweed seed. *Plant molecular biology* **33**, 1005-1010 (1997).

56. D.J Schultz1, J.I.M., D Cox-Foster3, R.A Grazzini4, R Craig5, R.O Mumma3 Anacardic acids in trichomes of Pelargonium: Biosynthesis, molecular biology and ecological effects. *Advances in Botanical Research* **31**, 175-192 (2000).

57. Ellen H. Yerger*, R.A.G., David Hesk1, Diana L. Cox-Foster, Richard Craig, and Ralph 0. Mumma A Rapid Method for Isolating Glandular Trichomes. *Plant Physiol* **99**, 1-7 (1992).

58. D. Hesk, R.C., R. O. Mumma Comparison of anacardic acid biosynthetic capability between insect-resistant and-susceptible geraniums. *Journal of Chemical Ecology* **18**, 1349-1364 (1992).

59. Kubo, I., S. Komatsu, et al. "Molluscicides from the Cashew Anacardium occidentale and Their Large-Scale Isolation." *Journal of Agricultural and Food Chemistry* **34**, 970-973 (1986).

60. Desouza, C.P., N. M. Mendes, et al. The use of cashew nut shell of Caju (ANACARDIUM-OCCIDENTALE) as alternative molluscicide *Revista Do Instituto De Medicina Tropical De Sao Paulo* **34**, 459-466 (1992).

61. Oliveira, M.S.C., S. M. de Morais, et al. Antioxidant, larvicidal and antiacetylcholinesterase activities of cashew nut shell liquid constituents. *Acta Tropica* **117**, 165-170. (2011).

62. Kubo, I., H. Muroi, et al. Structure-antibacterial acitivity relationships of anacardic acids. *Journal of Agricultural and Food Chemistry* **41**, 1016-1019 (1993).

63. Muroi, H.a.I.K. Bactericidal activity of Anacardic Acids against Streptococcus mutans and their potentiation. *Journal of Agricultural and Food Chemistry* **41**, 1780-1783 (1993).

64. Himejima, M.a.I.K. Antibacterial Agents from the Cashew Anacardium occidentale (Anacardiaceae) Nut Shell Oil. *Journal of Agricultural and Food Chemistry* **39**, 418-421 (1991).

65. Bouttier, S., J. Fourniat, et al. Beta-lactamase inhibitors from Anacardium occidentale. *Pharmaceutical Biology* **40**, 231-234 (2001).

66. Song, C., A. Kanthasamy, et al. Environmental Neurotoxic Pesticide Increases Histone Acetylation to Promote Apoptosis in Dopaminergic Neuronal Cells: Relevance to Epigenetic Mechanisms of Neurodegeneration. *Molecular Pharmacology* **77**, 621-632 (2010).

67. Kubo, I., M. Ochi, et al. Antitumor agents form the cashew (Anacardium occidentale) apple juice. *Journal of Agricultural and Food Chemistry* **41**, 1012-1015 (1993).

68. Sukumari-Ramesh, S., N. Singh, et al. Anacardic acid induces caspase-independent apoptosis and radiosensitizes pituitary adenoma cells. *Journal of Neurosurgery* **114**, 1681-1690. (2011).

69. Cui, L., J. Miao, et al. "Histone acetyltransferase inhibitor anacardic acid causes changes in global gene expression during in vitro Plasmodium falciparum development." *Eukaryotic Cell* **7**, 1200-1210. (2008).

70. Gellerman, J., Anderson, W. & Schlenk, H. Biosynthesis of anacardic acids from acetate in Ginkgo biloba. *Lipids* **9**, 722-725 (1974).

71. Ten Cate, R., P. Krawczyk, et al. Radiosensitizing effect of the histone acetyltransferase inhibitor anacardic acid on various mammalian cell lines. *Oncology Letters* **1**, 765-769. (2010).

72. Eliseeva, E.D., V. Valkov, et al. Characterization of novel inhibitors of histone acetyltransferases. *Molecular Cancer Therapeutics* **6**, 2391-2398 (2007).

73. Souto, J.A., M. Conte, et al. Synthesis of benzamides related to anacardic acid and their histone acetyltransferase (HAT) inhibitory activities. *Chemmedchem* **3**, 1435-1442. (2008).

74. Rea, A.I., J. M. Schmidt, et al. Cytotoxic activity of Ozoroa insignis from Zimbabwe. *Fitoterapia* **74**, 723-735 (2003).

75. Schultz, D.J., Cahoon E.B., Shanklin, J., Craig, R., Cox-Foster. D.L., Mumma, R.O. & Medford, J.I. Expression of a Δ^9 14:0-acyl carrier protein fatty acid desaturase gene is necessary for the production of ω^5 anacardic acids found in pest-resistant geranium (*Pelargonium x hortorum*). *Proceedings of the National Academy of Sciences* **93**, 8771-8775 (1996).

76. Schultz, D.J., Craig, R., Cox-Foster, D.L., Mumma, R.O. & Medford, J.I. RNA isolation from recalcitrant plant tissue. *Plant Mol Biol Reptr* **12**, 310-316 (1994).

77. Yandell, M. & Ence, D. A beginner's guide to eukaryotic genome annotation. *Nature reviews. Genetics* **13**, 329-342 (2012).

78. Nagaraj, S.H., Gasser, R.B. & Ranganathan, S. A hitchhiker's guide to expressed sequence tag (EST) analysis. *Briefings in bioinformatics* **8**, 6-21 (2007).

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

80. Stamatakis, A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**, 2688-2690 (2006).

81. França, L.T.C., Carrilho, E. & Kist, T.B.L. A review of DNA sequencing techniques. *Quarterly Reviews of Biophysics* **35** (2002).

82. Scotto-Lavino, E., Du, G. & Frohman, M.A. 5' end cDNA amplification using classic RACE. *Nat Protoc* **1**, 2555-2562 (2006).

83. Serrano Vega, M.J., Martinez-Force,E. and Garces Mancheno,R Cloning, characterization and structural model of a FatA-type thioesterase from sunflower seeds (Helianthus annuus L.). *Planta* **221**, 868-880 (2005).

84. Hawkins, D.J.a.K., J.C Characterization of acyl-ACP thioesterases of mangosteen (Garcinia mangostana) seed and high levels of stearate production in transgenic canola. *Plant J* **13**, 743-752 (1998).

85. Wu PZ, L.J., Wei Q, Zeng L, Chen YP, Li MR, Jiang HW and Wu GJ. Cloning and functional characterization of an acyl-acyl carrier protein thioesterase (JcFATB1) from *Jatropha curcas*. *Tree Physiol.* **29**, 1299-1305 (2009).

86. Dani KG, H.K., Ravikumar P and Kush A. & Structural and functional analyses of a saturated acyl ACP thioesterase, type B from immature seed tissue of *Jatropha curcas*. *Plant Bio* **13**, 456-461 (2011).

87. Li, M.-J., Li, A.-Q., Xia, H., Zhao, C.-Z., Li, C.-S., Wan, S.-B., Bi, Y.-P. & Wang, X.-J. Cloning and sequence analysis of putative type II fatty acid

synthase genes from Arachis hypogaea L. *Journal of Biosciences* **34**, 227-238 (2009).

88.   Li, J., Li, M.-R., Wu, P.-Z., Tian, C.-E., Jiang, H.-W. & Wu, G.-J. Molecular cloning and expression analysis of a gene encoding a putative β-ketoacyl-acyl carrier protein (ACP) synthase III (KAS III) from Jatropha curcas. *Tree physiology* **28**, 921-927 (2008).

89.   Mockaitis, K., Gilbert,D.G., Schmutz,J.J., Kuhn,D., Saski,C., Scheffler,B., Livingstone,D., Jenkins,J., Main,D., Feltus,A., Podicheti,R., Zheng,P., Ficklin,S., Haiminen,N., Royaert,S. May,G.D., Bretting,P., Shapiro,H., Schnell,R., Parida,L. and Motamayor,J.C. The genome sequence of the most widely cultivated cacao type and its use to identify candidate genes regulating pod color. *Genome biology* **14**, 53 (2013).

90.   Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D.J. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic acids research* **25**, 3389-3402 (1997).

91.   Larkin MA, B.G., Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ and Higgins DG ClustalW and ClustalX version 2. *Bioinformatics* **23**, 2947-2948 (2007).

92.   Goujon M, M.H., Li W, Valentin F, Squizzato S, Paern J, Lopez R A new bioinformatics analysis tools framework at EMBL-EMBL-EBI. *Nucleic acids research* **38**, W695-699 (2010).

93.   McWilliam H, L.W., Uludag M, Squizzato S, Park YM, Buso N, Cowley AP, Lopez Analysis Tool Web Services from the EMBL-EBI. *Nucleic acids research* **41**, W597-600 (2013).

94.   Suh, M.C., Schultz, D.J. & Ohlrogge, J.B. What limits production of unusual monoenoic fatty acids in transgenic plants. *Planta* **215**, 584-595 (2002).

95.   Hosamani, K.M. Unique Occurrence of Unusual Fatty Acids and Their Industrial Utilization. *Industrial & Engineering Chemistry Research* **35**, 326-331 (1996).

96.   Schultz D.J., Suh M.C. & Ohlrogge J.B. Stearoyl-acyl carrier protein and unusual acyl-acyl carrier protein desaturase activities are differentially influenced by ferredoxin. *Plant Physiology* **124**, 681-692 (2000).

97.   Walters, D., Harman, J., Craig, R. & Mumma, R. Effect of temperature on glandular trichome exudate composition and pest resistance in geraniums. *Entomologia experimentalis et applicata* **60**, 61-69 (1991).

98.   Richardson, M.L. Temperature influences the expression of resistance of soybean (*Glycine max*) to the soybean aphid (*Aphis glycines*). *Journal of Applied Entomology* **136**, 641-645 (2012).

99.   Walters, D., Harman, J., Graig, R. & Mumma, R. Effect of temperature on glandular trichome exudate composition and pest resistance in geraniums. *Entomologia Experimentalis et Applicata* **62**, 301-302 (1992).

100. Zobayed, S., Afreen, F. & Kozai, T. Temperature stress can alter the photosynthetic efficiency and secondary metabolite concentrations in St. John's wort. *Plant Physiology and Biochemistry* **43**, 977-984 (2005).

101. Gianoli, E. & Niemeyer, H.M. Characteristics of hydroxamic acid induction in wheat triggered by aphid infestation. *Journal of Chemical Ecology* **23**, 2695-2705 (1997).

102. Jaakola, L. & Hohtola, A. Effect of latitude on flavonoid biosynthesis in plants. *Plant, cell & environment* **33**, 1239-1247 (2010).

103. Bräutigam, A. & Gowik, U. What can next generation sequencing do for you? Next generation sequencing as a valuable tool in plant research. *Plant Biology* **12**, 831-841 (2010).

104. Illumina, I. BaseSpace User Guide. *Illumina* **15044182** (2014).

105. Illumina NextSeq(R) Series of Sequencing Systems. *Illumina* **770-2013-053** (2015).

106. Martin, J.A. & Wang, Z. Next-generation transcriptome assembly. *Nature reviews. Genetics* **12**, 671-682 (2011).

107. da Fonseca, R.R., Albrechtsen, A., Themudo, G.E., Ramos-Madrigal, J., Sibbesen, J.A., Maretty, L., Zepeda-Mendoza, M.L., Campos, P.F., Heller, R. & Pereira, R.J. Next-generation biology: Sequencing and data analysis approaches for non-model organisms. *Mar Genomics* (2016).

108. Andrews, S. FastQC A Quality Control Tool for High Throughput Sequence Data. http://bioinformatics.babraham.ac.uk/projects/fastqc/ (2014).

109. M. G. Grabherr, B.J.H., M. Yassour, J. Z. Levin, D. A. Thompson, I. Amit, et al Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology* **29**, 644-652 (2011).

110. S. Gotz, J.M.G.-G., J. Terol, T. D. Williams, S. H. Nagaraj, M. J. Nueda, et al. High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic acids research* **36**, 3420-3435 (2008).

111. Salzberg, B.L.a.S.L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-359 (2012).

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

113. Y. Li, X.W., T. Chen, F. Yao, C. Li, Q. Tang, et al. RNA-Seq Based De Novo Transcriptome Assembly and Gene Discovery of Cistanche deserticola Fleshy Stem. *PloS One* **10**, e0125722 (2015).

114. M. Salem, B.P., R. Al-Tobasei, F. Abdouni, G. H. Thorgaard, C. E. Rexroad, et al. Transcriptome assembly, gene annotation and tissue gene expression atlas of the rainbow trout. *PloS One* **10** (2015).

115. M. Ashburner, C.A.B., J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**, 25-29 (2000).

116. Consortium, U. UniProt: a hub for protein information. *Nucleic acids research* **43**, 204-212 (2015).

117. Consortium, G.O. The Gene Ontology (GO) database and informatics resource. *Nucleic acids research* **32**, D258-D261 (2004).

118. R. C. Gentleman, V.J.C., D. M. Bates, B. Bolstad, M. Dettling, S. Dudoit, et al. Bioconductor open software development for computational biology and bioinformatics. *Genome biology* **5**, R80 (2004).

119. Gentleman, S.F.a.R. Using GOstats to test gene lists for GO term association. *Bioinformatics* **23**, 257-258 (2007).

120. Bates, D., Maechler, M., Bolker, B. & Walker, S. lme4: Linear mixed-effects models using Eigen and S4. *R package version* **1** (2014).

121. Wilson, M. & De Boeck, P. in Explanatory item response models 43-74 (Springer, 2004).

122. Hothorn, T., Bretz, F. & Westfall, P. Simultaneous inference in general parametric models. *Biometrical journal* **50**, 346-363 (2008).

123. Storey, J.D. A direct approach to false discovery rates. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **64**, 479-498 (2002).

124. Storey, J.D. The positive false discovery rate: a Bayesian interpretation and the q-value. *Annals of statistics*, 2013-2035 (2003).

125. Storey, J.D. & Tibshirani, R. Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences* **100**, 9440-9445 (2003).

126. Storey, J.D., Taylor, J.E. & Siegmund, D. Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **66**, 187-205 (2004).

127. Rajkumar, A.P., Qvist, P., Lazarus, R., Lescai, F., Ju, J., Nyegaard, M., Mors, O., Borglum, A.D., Li, Q. & Christensen, J.H. Experimental validation of methods for differential gene expression analysis and sample pooling in RNA-seq. *BMC genomics* **16**, 548 (2015).

128. Fang, Z. & Cui, X. Design and validation issues in RNA-seq experiments. *Briefings in bioinformatics* **12**, 280-287 (2011).

129. Honaas, L.A., Wafula, E.K., Wickett, N.J., Der, J.P., Zhang, Y., Edger, P.P., Altman, N.S., Pires, J.C., Leebens-Mack, J.H. & dePamphilis, C.W. Selecting Superior De Novo Transcriptome Assemblies: Lessons Learned by Leveraging the Best Plant Genome. *PLoS One* **11**, e0146062 (2016).

130. Marguerat, S. & Bähler, J. RNA-seq: from technology to biology. *Cellular and molecular life sciences* **67**, 569-579 (2010).

131. Cocquet, J., Chong, A., Zhang, G. & Veitia, R.A. Reverse transcriptase template switching and false alternative transcripts. *Genomics* **88**, 127-131 (2006).

132. Lu, B., Zeng, Z. & Shi, T. Comparative study of de novo assembly and genome-guided assembly strategies for transcriptome reconstruction based on RNA-Seq. *Sci China Life Sci* **56**, 143-155 (2013).

133. Nagano, M., Uchimiya, H. & Kawai-Yamada, M. Plant sphingolipid fatty acid 2-hydroxylases have unique characters unlike their animal and fungus counterparts. *Plant signaling & behavior* **7**, 1388-1392 (2012).

134. Cartegni, L., Chew, S.L. & Krainer, A.R. Listening to silence and understanding nonsense: exonic mutations that affect splicing. *Nature Reviews Genetics* **3**, 285-298 (2002).

135. Spitzer, M., Lorkowski, S., Cullen, P., Sczyrba, A. & Fuellen, G. IsoSVM--distinguishing isoforms and paralogs on the protein level. *BMC Bioinformatics* **7**, 110 (2006).

136. Fitch, W.M. Distinguishing homologous from analogous proteins. *Systematic Biology* **19**, 99-113 (1970).

137. Budak, H. & Akpinar, B.A. Plant miRNAs: biogenesis, organization and origins. *Functional & Integrative Genomics* **15**, 523-531 (2015).

138. Budak, H., Khan, Z. & Kantar, M. History and current status of wheat miRNAs using next-generation sequencing and their roles in development and stress. *Brief Funct Genomics* **14**, 189-198 (2015).

139. Budak, H., Kantar, M., Bulut, R. & Akpinar, B.A. Stress responsive miRNAs and isomiRs in cereals. *Plant Science* **235**, 1-13 (2015).

140. Chen, X. microRNA biogenesis and function in plants. *FEBS Letters* **579**, 5923-5931 (2005).

141. Zhang, B., Pan, X., Cobb, G.P. & Anderson, T.A. Plant microRNA: a small regulatory molecule with big impact. *Dev Biol* **289**, 3-16 (2006).

142. Zhao, M., Meyers, B.C., Cai, C., Xu, W. & Ma, J. Evolutionary patterns and coevolutionary consequences of MIRNA genes and MicroRNA targets triggered by multiple mechanisms of genomic duplications in soybean. *The Plant Cell* **27**, 546-562 (2015).

143. Kidner, C.A. & Martienssen, R.A. The developmental role of microRNA in plants. *Current opinion in plant biology* **8**, 38-44 (2005).

144. Allen, E., Xie, Z., Gustafson, A.M., Sung, G.-H., Spatafora, J.W. & Carrington, J.C. Evolution of microRNA genes by inverted duplication of target gene sequences in Arabidopsis thaliana. *Nature genetics* **36**, 1282-1290 (2004).

145. Zhang, B., Wang, Q. & Pan, X. MicroRNAs and their regulatory roles in animals and plants. *J Cell Physiol* **210**, 279-289 (2007).

146. Mallory, A.C., Dugas, D.V., Bartel, D.P. & Bartel, B. MicroRNA regulation of NAC-domain targets is required for proper formation and separation of adjacent embryonic, vegetative, and floral organs. *Current Biology* **14**, 1035-1046 (2004).

147. Lauter, N., Kampani, A., Carlson, S., Goebel, M. & Moose, S.P. microRNA172 down-regulates glossy15 to promote vegetative phase change in maize. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 9412-9417 (2005).

148. Kim, J., Jung, J.H., Reyes, J.L., Kim, Y.S., Kim, S.Y., Chung, K.S., Kim, J.A., Lee, M., Lee, Y. & Narry Kim, V. microRNA-directed cleavage of ATHB15 mRNA regulates vascular development in Arabidopsis inflorescence stems. *The Plant Journal* **42**, 84-94 (2005).

149. Guo, H.-S., Xie, Q., Fei, J.-F. & Chua, N.-H. MicroRNA directs mRNA cleavage of the transcription factor NAC1 to downregulate auxin signals

for Arabidopsis lateral root development. *The Plant Cell* **17**, 1376-1386 (2005).

150. Chen, J., Li, W.X., Xie, D., Peng, J.R. & Ding, S.W. Viral virulence protein suppresses RNA silencing–mediated defense but upregulates the role of microRNA in host gene expression. *The Plant Cell* **16**, 1302-1313 (2004).

151. Pattanaik, S., Patra, B., Singh, S.K. & Yuan, L. An overview of the gene regulatory network controlling trichome development in the model plant *Arabidopsis*. *Front Plant Sci* **5**, 259 (2014).

152. Schwab, R., Ossowski, S., Riester, M., Warthmann, N. & Weigel, D. Highly specific gene silencing by artificial microRNAs in *Arabidopsis*. *Plant Cell* **18**, 1121-1133 (2006).

153. Kwak, P.B., Wang, Q.Q., Chen, X.S., Qiu, C.X. & Yang, Z.M. Enrichment of a set of microRNAs during the cotton fiber development. *BMC genomics* **10**, 457 (2009).

154. Xue, X.Y., Zhao, B., Chao, L.M., Chen, D.Y., Cui, W.R., Mao, Y.B., Wang, L.J. & Chen, X.Y. Interaction between two timing microRNAs controls trichome distribution in *Arabidopsis*. *PLoS Genet* **10**, e1004266 (2014).

155. Rubio-Somoza, I. & Weigel, D. MicroRNA networks and developmental plasticity in plants. *Trends Plant Sci* **16**, 258-264 (2011).

156. Yu, N., Cai, W.-J., Wang, S., Shan, C.-M., Wang, L.-J. & Chen, X.-Y. Temporal control of trichome distribution by microRNA156-targeted SPL genes in *Arabidopsis thaliana*. *The Plant Cell* **22**, 2322-2335 (2010).

157. Rhoades, M.W., Reinhart, B.J., Lim, L.P., Burge, C.B., Bartel, B. & Bartel, D.P. Prediction of plant microRNA targets. *Cell* **110**, 513-520 (2002).

158. Sacco, J. & Adeli, K. MicroRNAs: emerging roles in lipid and lipoprotein metabolism. *Current Opinion in Lipidology* **23** (2012).

159. Xie, F., Frazier, T.P. & Zhang, B. Identification and characterization of microRNAs and their targets in the bioenergy plant switchgrass (*Panicum virgatum*). *Planta* **232**, 417-434 (2010).

160. Fernandez-Hernando, C., Suarez, Y., Rayner, K.J. & Moore, K.J. MicroRNAs in lipid metabolism. *Curr Opin Lipidol* **22**, 86-92 (2011).

161. Xie, F., Frazier, T.P. & Zhang, B. Identification, characterization and expression analysis of MicroRNAs and their targets in the potato (*Solanum tuberosum*). *Gene* **473**, 8-22 (2011).

162. M. R. Friedlander, S.D.M., N. Li, W. Chen, and N. Rajewsky miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. *Nucleic acids research* **40**, 37-52 (2012).

163. P. J. Cock, C.J.F., N. Goto, M. L. Heuer, and P. M. The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. *Rice"Nucleic Acids Res* **38**, 1767-1771 (2010).

164. A. M. Bolger, M.L., and B. Usadel Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30** (2014).

165. Griffiths-Jones, S., Grocock, R.J., van Dongen, S., Bateman, A. & Enright, A.J. miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic acids research* **34**, D140-144 (2006).

166. Dai, X. & Zhao, P.X. psRNATarget: a plant small RNA target analysis server. *Nucleic acids research* **39**, W155-159 (2011).

167. A. Singh, S.S., K. C. Panigrahi, R. Reski, and A. K. Sarkar Balanced activity of microRNA166/165 and its target transcripts from the class III homeodomain-leucine zipper family regulates root growth in *Arabidopsis thaliana*. *Plant Cell* **33**, 945-953 (2014).

168. A. Carlsbecker, J.Y.L., C. J. Roberts, J. Dettmer, S. Lehesranta, J. Zhou, et al. Cell signalling by microRNA165/6 directs gene dose-dependent root cell fate. *Nature* **465**, 316-321 (2010).

169. Wang, Q.D.a.H. The role of HD-ZIP III transcription factors and miR165/166 in vascular development and secondary cell wall formation. *Plant Signal Behav* **10**, 1078955 (2015).

170. Q. Liu, X.Y., L. Pi, H. Wang, X. Cui, and H. Huang The ARGONAUTE10 gene modulates shoot apical meristem maintenance and establishment of leaf polarity by repressing miR165/166 in *Arabidopsis*. *Plant J* **58**, 27-40 (2009).

171. Zhang, Z.Z.a.X. Argonautes compete for miR165/166 to regulate shoot apical meristem development. *Current opinion in plant biology* **15**, 652-658 (2012).

172. Y. Zhou, M.H., H. Zhu, Z. Zhang, X. Guo, T. Li, et al. Spatiotemporal sequestration of miR165/166 by Arabidopsis Argonaute10 promotes shoot apical meristem maintenance. *Cell Rep* **10**, 1819-1827 (2015).

173. Millar, M.R.a.A.A. Specificity of plant microRNA target MIMICs: Cross-targeting of miR159 and miR319. *Plant Physiol* **180**, 45-48 (2015).

174. C. Schommer, J.M.D., E. G. Bresso, R. E. Rodriguez, and J. F. Palatnik. Repression of cell proliferation by miR319-regulated TCP4. *Mol Plant* **7**, 1533-1544 (2014).

175. A. Nag, S.K., and T. Jack miR319a targeting of TCP4 is critical for petal growth and development in *Arabidopsis*. *Proc Natl Acad Sci U S A* **106**, 22534-22539 (2009).

176. Millar, A.A. & Gubler, F. The Arabidopsis GAMYB-like genes, MYB33 and MYB65, are microRNA-regulated genes that redundantly facilitate anther development. *The Plant Cell* **17**, 705-721 (2005).

177. Chen, X. A microRNA as a translational repressor of APETALA2 in Arabidopsis flower development. *Science* **303**, 2022-2025 (2004).

178. Chen, X. Small RNAs and their roles in plant development. *Annu Rev Cell Dev Biol* **25**, 21-44 (2009).

179. Reinhart, B.J., Weinstein, E.G., Rhoades, M.W., Bartel, B. & Bartel, D.P. MicroRNAs in plants. *Genes & development* **16**, 1616-1626 (2002).

180. Aukerman, M.J. & Sakai, H. Regulation of flowering time and floral organ identity by a microRNA and its APETALA2-like target genes. *The Plant Cell* **15**, 2730-2741 (2003).

181. Laufs, P., Peaucelle, A., Morin, H. & Traas, J. MicroRNA regulation of the CUC genes is required for boundary size control in Arabidopsis meristems. *Development* **131**, 4311-4322 (2004).

182. Ori, N., Cohen, A.R., Etzioni, A., Brand, A., Yanai, O., Shleizer, S., Menda, N., Amsellem, Z., Efroni, I., Pekker, I., Alvarez, J.P., Blum, E., Zamir, D. & Eshed, Y. Regulation of LANCEOLATE by miR319 is required for compound-leaf development in tomato. *Nat Genet* **39**, 787-791 (2007).

183. Wang, X.-J., Reyes, J.L., Chua, N.-H. & Gaasterland, T. Prediction and identification of Arabidopsis thaliana microRNAs and their mRNA targets. *Genome biology* **5**, 1 (2004).

184. Allen, E., Xie, Z., Gustafson, A.M. & Carrington, J.C. microRNA-directed phasing during trans-acting siRNA biogenesis in plants. *Cell* **121**, 207-221 (2005).

185. Vaucheret, H., Vazquez, F., Crété, P. & Bartel, D.P. The action of ARGONAUTE1 in the miRNA pathway and its regulation by the miRNA pathway are crucial for plant development. *Genes & development* **18**, 1187-1197 (2004).

186. Xie, W., Weng, A. & Melzig, M.F. MicroRNAs as New Bioactive Components in Medicinal Plants. *Planta Med* (2016).

187. Joven, J., Espinel, E., Rull, A., Aragonès, G., Rodríguez-Gallego, E., Camps, J., Micol, V., Herranz-López, M., Menéndez, J.A. & Borrás, I. Plant-derived polyphenols regulate expression of miRNA paralogs miR-103/107 and miR-122 and prevent diet-induced fatty liver disease in hyperlipidemic mice. *Biochimica et Biophysica Acta (BBA)-General Subjects* **1820**, 894-899 (2012).

188. Yang, F., Cai, J., Yang, Y. & Liu, Z. Overexpression of microRNA828 reduces anthocyanin accumulation in *Arabidopsis*. *Plant Cell, Tissue and Organ Culture (PCTOC)* **115**, 159-167 (2013).

189. Maghuly, F., Ramkat, R.C. & Laimer, M. Virus versus host plant microRNAs: who determines the outcome of the interaction. *PloS one* **9**, e98263 (2014).

190. Curaba, J., Spriggs, A., Taylor, J., Li, Z. & Helliwell, C. miRNA regulation in the early development of barley seed. *BMC plant biology* **12**, 1 (2012).

191. Xie, F., Jones, D.C., Wang, Q., Sun, R. & Zhang, B. Small RNA sequencing identifies miRNA roles in ovule and fibre development. *Plant Biotechnol J* **13**, 355-369 (2015).

192. Din, M. & Barozai, M.Y.K. Profiling microRNAs and their targets in an important fleshy fruit: Tomato *(Solanum lycopersicum)*. *Gene* **535**, 198-203 (2014).

193. Liu, Y., Wang, L., Chen, D., Wu, X., Huang, D., Chen, L., Li, L., Deng, X. & Xu, Q. Genome-wide comparison of microRNAs and their targeted transcripts among leaf, flower and fruit of sweet orange. *BMC genomics* **15**, 1 (2014).

194. Zhu, Q.-w. & Luo, Y.-p. Identification of miRNAs and their targets in tea (Camellia sinensis). *Journal of Zhejiang University Science B* **14**, 916-923 (2013).

195. Zhang, Z., Yu, J., Li, D., Zhang, Z., Liu, F., Zhou, X., Wang, T., Ling, Y. & Su, Z. PMRD: plant microRNA database. *Nucleic acids research* **38**, D806-813 (2010).

196. Vashisht, I., Mishra, P., Pal, T., Chanumolu, S., Singh, T.R. & Chauhan, R.S. Mining NGS transcriptomes for miRNAs and dissecting their role in regulating growth, development, and secondary metabolites production in different organs of a medicinal herb, Picrorhiza kurroa. *Planta* **241**, 1255-1268 (2015).

197. Sayanova, O., Haslam, R., Venegas Caleron, M. & Napier, J.A. Cloning and characterization of unusual fatty acid desaturases from Anemone leveillei: identification of an acyl-coenzyme A C20 Delta5-desaturase responsible for the synthesis of sciadonic acid. *Plant Physiol* **144**, 455-467 (2007).

198. De Marchis, F., Valeri, M.C., Pompa, A., Bouveret, E., Alagna, F., Grisan, S., Stanzione, V., Mariotti, R., Cultrera, N., Baldoni, L. & Bellucci, M. Overexpression of the olive acyl carrier protein gene (OeACP1) produces alterations in fatty acid composition of tobacco leaves. *Transgenic research* **25**, 45-61 (2016).

199. DAVID J.SCHULTZ*, E.B.C., JOHN SHANKLINt, RICHARD CRAIGt, DLANA L. COX-FOSTER§, RALPH 0. MUMMA§, AND JUNE I.MEDFORDT1 Expression of a d9 14:0-acyl carrier protein fatty acid desaturase gene is necessary for the productionof 5anacardicacidsfound in pest-resistant geranium (Pelargonium xhortorum). *Plant Biology* **93**, 8771-8775 (1996).

200. Vemanna, R.S., Chandrashekar, B.K., Hanumantha Rao, H.M., Sathyanarayanagupta, S.K., Sarangi, K.S., Nataraja, K.N. & Udayakumar, M. A modified MultiSite gateway cloning strategy for consolidation of genes in plants. *Molecular biotechnology* **53**, 129-138 (2013).

201. Usadel, B., Obayashi, T., Mutwil, M., Giorgi, F., Bassel, G., Tanimoto, M., Chow, A., Steinhauser, D., Persson, S. & Provart, N. Co-expression tools for plant biology: opportunities for hypothesis generation and caveats. *Plant, Cell & Environment* **32**, 1633-1651 (2009).

202. Emanuelsson, O., Nielsen, H. & Von Heijne, G. ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. *Protein Science* **8**, 978-984 (1999).

203. Petersen, T.N., Brunak, S., von Heijne, G. & Nielsen, H. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature methods* **8**, 785-786 (2011).

204. Emanuelsson, O., Brunak, S., von Heijne, G. & Nielsen, H. Locating proteins in the cell using TargetP, SignalP and related tools. *Nature protocols* **2**, 953-971 (2007).

205. Burow, M.D., Chlan, C.A., Sen, P., Lisca, A. & Murai, N. High-frequency generation of transgenic tobacco plants after modified leaf disk cocultivation with Agrobacterium tumefaciens. *Plant Molecular Biology Reporter* **8**, 124-139 (1990).

206. Peng, Z., Li, L., Yang, L., Zhang, B., Chen, G. & Bi, Y. Overexpression of peanut diacylglycerol acyltransferase 2 in *Escherichia coli. PLoS One* **8**, e61363 (2013).

207. Kohli, A., González-Melendi, P., Abranches, R., Capell, T., Stoger, E. & Christou, P. The quest to understand the basis and mechanisms that control expression of introduced transgenes in crop plants. *Plant signaling & behavior* **1**, 185-195 (2006).

208. Beaujean, A., Sangwan, R., Hodges, M. & Sangwan-Norreel, B. Effect of ploidy and homozygosity on transgene expression in primary tobacco transformants and their androgenetic progenies. *Molecular and General Genetics MGG* **260**, 362-371 (1998).

209. Schultz, D.J., Cahoon, E.B., Shanklin, J., Craig, R., Cox-Foster, D.L., Mumma, R.O. & Medford, J.I. Expression of a delta 9 14:0-acyl carrier protein fatty acid desaturase gene is necessary for the production of omega 5 anacardic acids found in pest-resistant geranium (*Pelargonium* x *hortorum*). *Proc Natl Acad Sci U S A* **93**, 8771-8775 (1996).

EXPRESSION AND PURIFICATION OF TWO *PELARGONIUM × HORTORUM* ACYL CARRIER PROTEIN ISOFORMS IN *E. COLI.*

Summary

An important co-factor in fatty acid biosynthesis is acyl carrier protein (ACP). ACP is a conserved carrier of acyl intermediates during fatty acid synthesis. Two isoforms of the ACP's have been identified from a *Pelargonium × hortorum* glandular trichome EST database (see description in Chapter 2). The coding sequence of each isoforms was cloned in pet22b vector, expression was induced in Rosetta (DE3) PLySs cell line and the protein products were purified using Ni-NTA column chromatography. These recombinant ACP isoforms can be used to generate acyl-ACP substrates to determine fatty acid biosynthesis protein specificity of substrates based on ACP isoforms.

Material and Methods

Amplification of ACP 1 and ACP 2

Gene-specific primers with NdeI restriction site (5' primer) and XhoI restriction site (3' primer) were designed for each ACP isoform (Table IA.1).   ACP 1 primers are 5' - TTC CAG ATT CAT ATG TCG GCC AAA CCA GAG ACT GTG -

3' and 3' - ATG CTC TCG AGA GCA GGA GCC TTC TTC TC - 5', ACP 2

primers are 5'- TCG CCT CCA GGT CAT ATG TCG GCC AAA GCA GAG ACT

GTG- 3' and 3' TCT CCA AAC TCT CGA GAG CAT CCT TCT TCT CAA C- 5'.

Expression construct

PCR products for each target template were purified using GeneClean Kit III (MP

Biomedical, LLC) then cloned into pGem T Easy vector according to

manufacturer's directions (Promega). Each clone was sequenced in both

directions and verified to be without sequence errors. Subsequently, the NdeI

and XhoI digested fragments were generated by restriction digest, purified and

cloned into compatible sites of the pet22b vector. The presence of ACP inserts

in pet22b after cloning was identified via restriction digest.

Induction assay

For overexpression of recombinant ACP proteins in pet22b vector, constructs

were chemically transformed in Rosetta pLysS (DE3) cell line. Three hundred µl

of overnight culture (2 ml LB supplemented with Carbenicillin 0.1 mg/ml and

Chloramphenicol 0.030 mg/ml) was used to inoculate 30 ml of LB supplemented

with the same antibiotics and were grown at 37°C on a rotary platform shaker set

at 250 rpm. Protein expression was induced by addition of IPTG to 0.6 mM when

cells reached an optical density of 0.5-0.8 at 600 nm. Subsequently, expression

was further optimized in a similar manner at lower temperatures (25°C). After

four hours of induction, cells (1 ml) were collected by centrifugation and cell

pellets were re-suspended in 0.3 ml of Lysis buffer (50 mM Tris-HCl, 5% Glycerol

and 50 mM NaCl, 7.5 kU/ml Lysozyme, 0.5mM PMSF and 25 U/ml DNase, pH 7-

8). After lysis, the cells were sonicated on ice (Ultrasonic Homogenizer – Cole

Parmer Instrument Co.) for at least 2 times for 20 seconds with an interval of 30

sec. Protein preparations were made by centrifuging (3000 × g) the sample for

20 mins at 4°C. The supernatant was recovered as the soluble protein fraction

and pellet was kept as the insoluble protein fraction. Coomassie Plus Protein

Assay (Pierce, Rockford, IL) was used to estimate protein concentrations.

Protein was normalized in all the samples and then 30 μg of protein was loaded

onto a 4-15% gradient Tris-HCl SDS-PAGE gel to check for the expression of the

protein of interest. Gels were stained in methanol/water/acetic acid (30/60/10;

v/v/v) containing Coomassie brilliant blue R250. Gels were destained in multiple

successions of methanol/water/acetic acid (30/60/10; v/v/v) until protein bands

were visible.

Protein purification

Cells (5 ml) were harvested after 4 hours from induced and uninduced samples

then centrifuged (3000 × g) and stored as pellets at -80°C until subjected to

purification. Three buffers NPI-10, NPI-20 and NPI-500 were prepared in the lab

as per Ni-NTA Spin Column Purification Handbook (QIAGEN). Ni-NTA Spin

Columns (QIAGEN) was used for purification as described in the handbook for

Ni-NTA Spin Column Purification of 6xHis-Tagged Proteins under Native

Conditions from *E. coli* cell Lysates.

Results and Conclusion

ACP amplification and Induction of recombinant proteins

PCR amplification resulted in the expected fragment size of 250 bp, these

fragments were cloned into the expression vector and recombinant molecules

were verified by restriction digest (Figure A1A.1). These constructs were

transformed into Rosetta (DE3)PLySs cell line and induced with IPTG for over-

expression of ACPs. Both ACP proteins were observed in total protein samples

that were induced and absent in samples that were not induced (Figure A1A.2

and A1A.3). Both ACPs were part of the soluble fraction of the total protein

(Figure A1A.2).
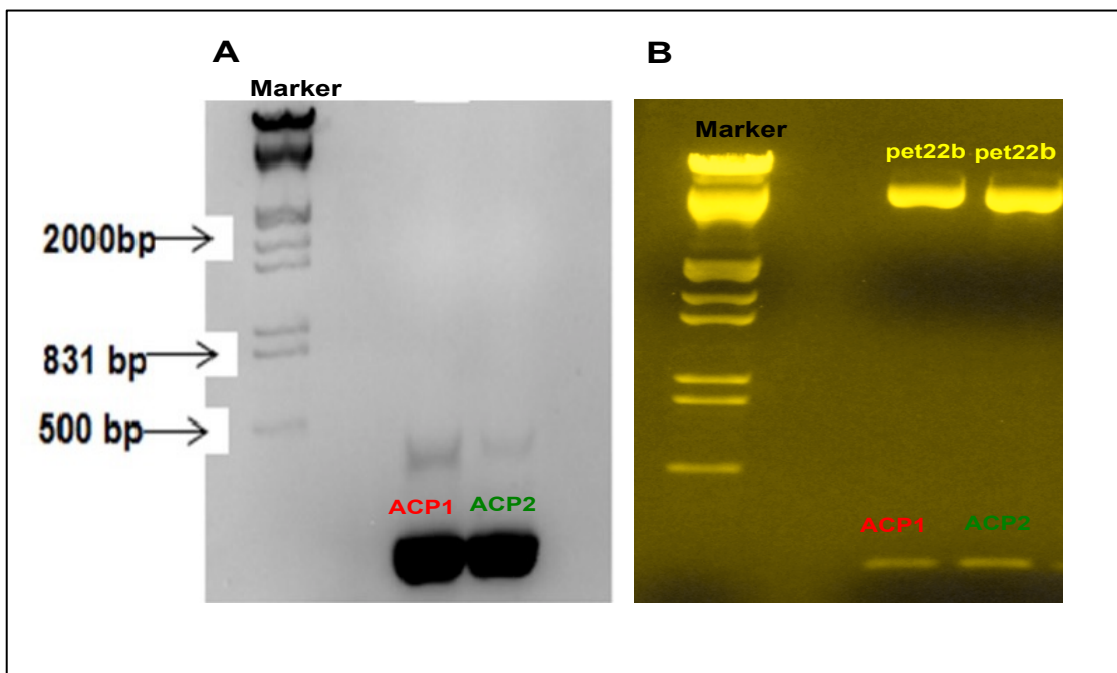
Figure A1A.1. Cloning of ACP 1 and 2 in pet22b vector. (A) PCR amplification of ACP 1 and ACP 2 from pBluescript construct. (B) Restrict Digest for verifying cloning of ACP in pet22b vector.
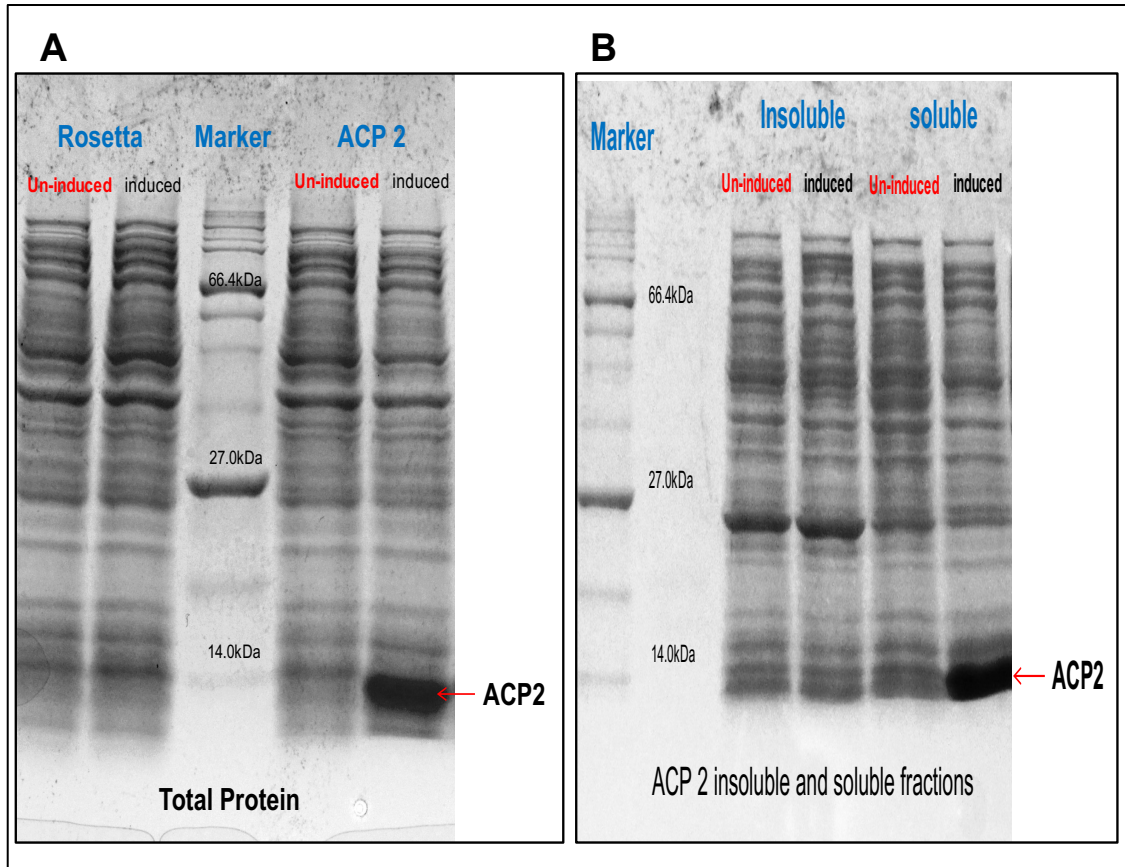
Figure A1A.2. Protein gel for ACP 2 overexpression. (A) ACP 2 protein band observed approximately at 13kDa in induced total protein sample. (B) ACP 2 protein band observed approximately at 13kDa in induced soluble protein sample.

Purification of ACP 1 and ACP 2

Affinity column chromatography yielded 1.2 mg ACP 2 and 0.75 mg of ACP 1.

Both ACP samples were considered to be pure since the wash fraction obtained

before the eluting pure protein was clear and no protein was detected using a

Bradfords assay (Figure A1A.4). The purification protocol led to recovery of 90

% of ACP 2 and 75% of ACP 1 from the protein bound to the column. (Figure

A1A.4)
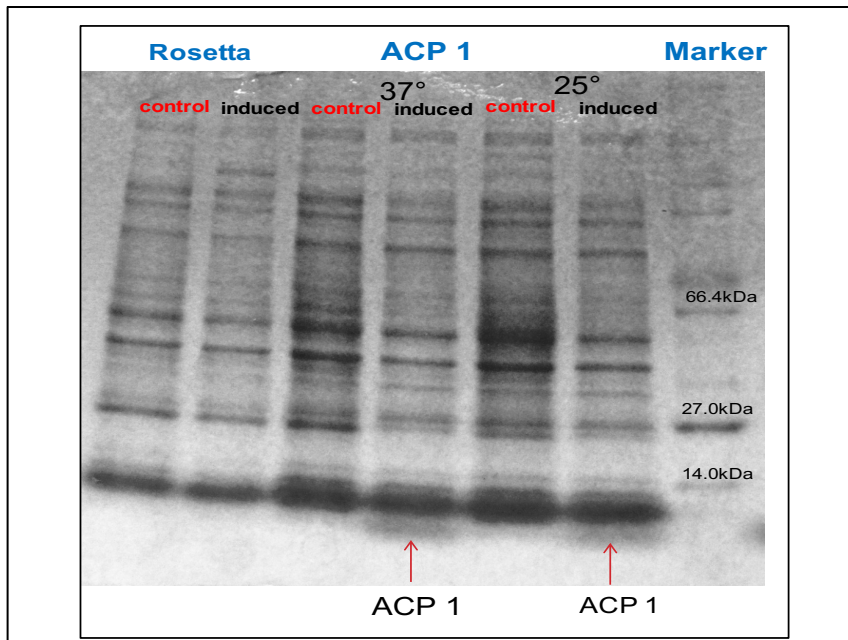
Figure A1A.3. Protein gel for ACP 1 overexpression. Faint ACP 1 protein band observed approximately at 13kDa in induced total protein sample for both 37°C and 25°C induction experiments.

Based on this experiment a large scale purification of these ACP proteins can be conducted and pure proteins can be used in either biochemical assays or x-ray crystallographic studies of the geranium ACPs. Such protein studies can provide valuable insights into substrate specificity, binding affinity, structure and enzymatic interactions of geranium ACPs and thus expanding knowledge about their role in UMFA synthesis.

A)

Protein	Amount
Total Protein	9.67mg
Insoluble	0.9mg
Soluble	8.77mg
Unbound	7.44mg
Bound	1.33mg
ACP 2	1.2mg

B)

Protein	Amount
Total	8.32mg
Insoluble	0.82mg
Soluble	6.9mg
Unbound	5.9mg
Bound	1 mg
ACP 1	0.75mg

C)
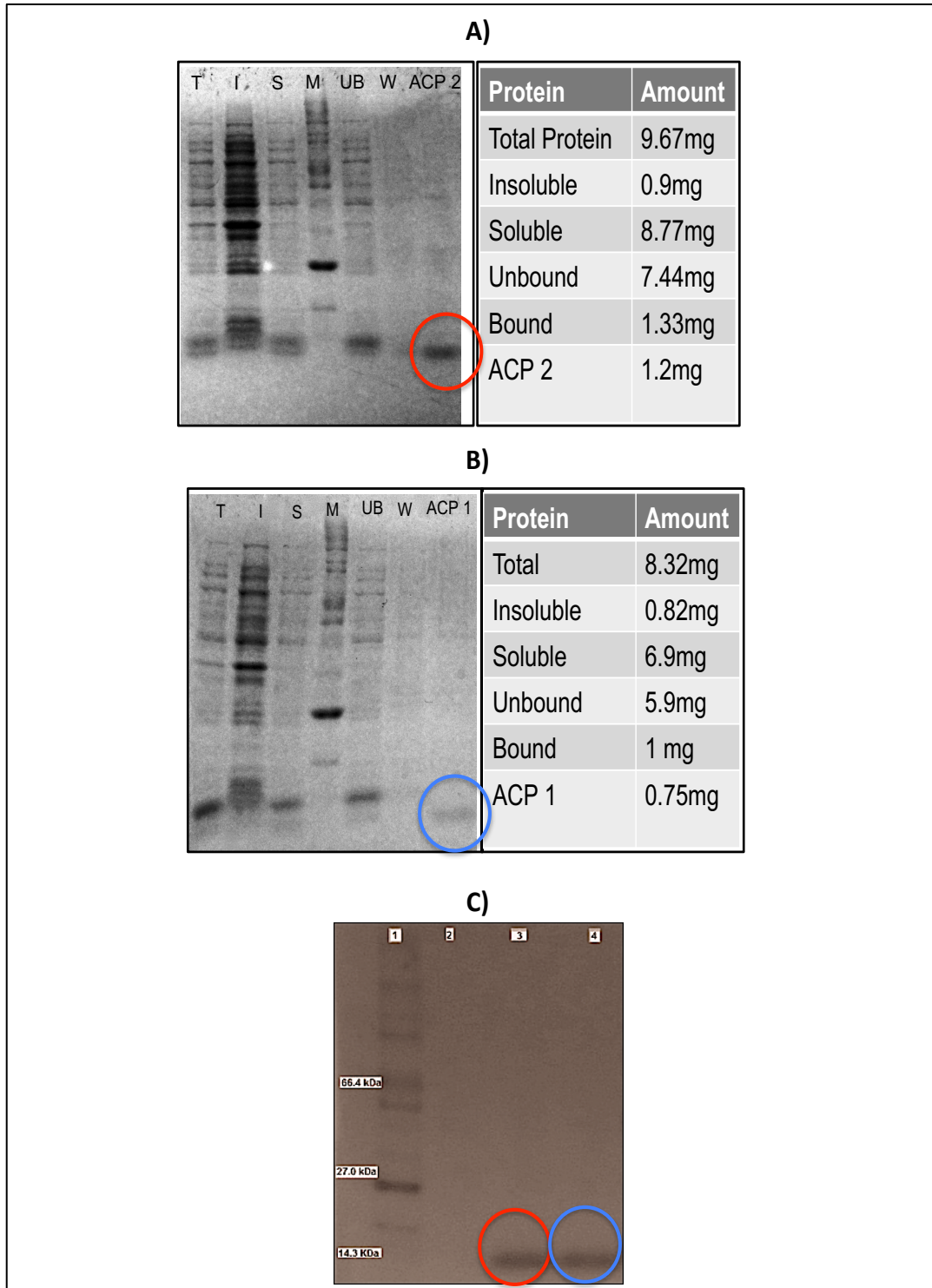
Figure A1A.4. Purified ACP 1 and ACP 2. (A) ACP 2 protein recovered from the column. (B) ACP 1 protein recovered from the column. (C) Purified ACP 2 and ACP 1 proteins obtained after final wash. T= Total protein sample, I = Insoluble fraction of protein, S = soluble fraction of protein, M= Marker, UB= Unbound protein, W= wash, Red Circle- pure ACP 2 protein And Blue Circle- pure ACP 1 protein.

APPENDIX IB

EXPRESSION OF *PELARGONIUM × HORTORUM* 3-KETOACYL-COA
SYNTHASE 2 IN TOBACCO LEAVES.

Summary

A 3-ketoacyl-CoA synthase (KCS 2) was isolated from geranium and was found
to be highly expressed in trichomes and thus was speculated to be potential
condensing enzyme invovled in anacardic acid biosynthesis. To directly
determine function and to test its involvement in anacardic acid biosynthesis,
KCS 2 was expressed in tobacco leaves via leaf disc transformation method. The
primary transformants (T_0 plant generation, n=3) were found to be sterile and
thus did not produce seed to screen next generation for further analysis (Figure
A2A.1). Biochemical analyses of tobacco leaves (T_0) via HPLC and GC did not
show any significant change of the fatty acid content or production of anacardic
acid like compound as compared to the wild type tobacco plants (Figure A2A.2)

Material and Methods

Gene specific primers that incorporated restriction sites for cloning into the plant
expression vector (Appendix 5, Table A5.1) were synthesized. The KCS 2
coding region was amplified from the EST clone of KCS 2 in pBluescript. The

PCR product was subsequently cloned into the NdeI (5') and SalI (3') sites of

pRI-201AN to generate the pRI-KCS2 plant expression construct. Directional

cloning was verified via restriction analysis and sequencing of the construct.

Plant transformation method, reverse transcription PCR for testing presence of

KCS2 in primary transfromants, GC and statistical analysis was done as

described in Chapter 5. HPLC analysis was done as described in Chapter 3.

Primers used for reverse transcription are

5'-TCCGGACGATAACGCCCAAGA-3' and 3'-AACCTTGTCCCCTTTCCTAAGC-

5'.

Results
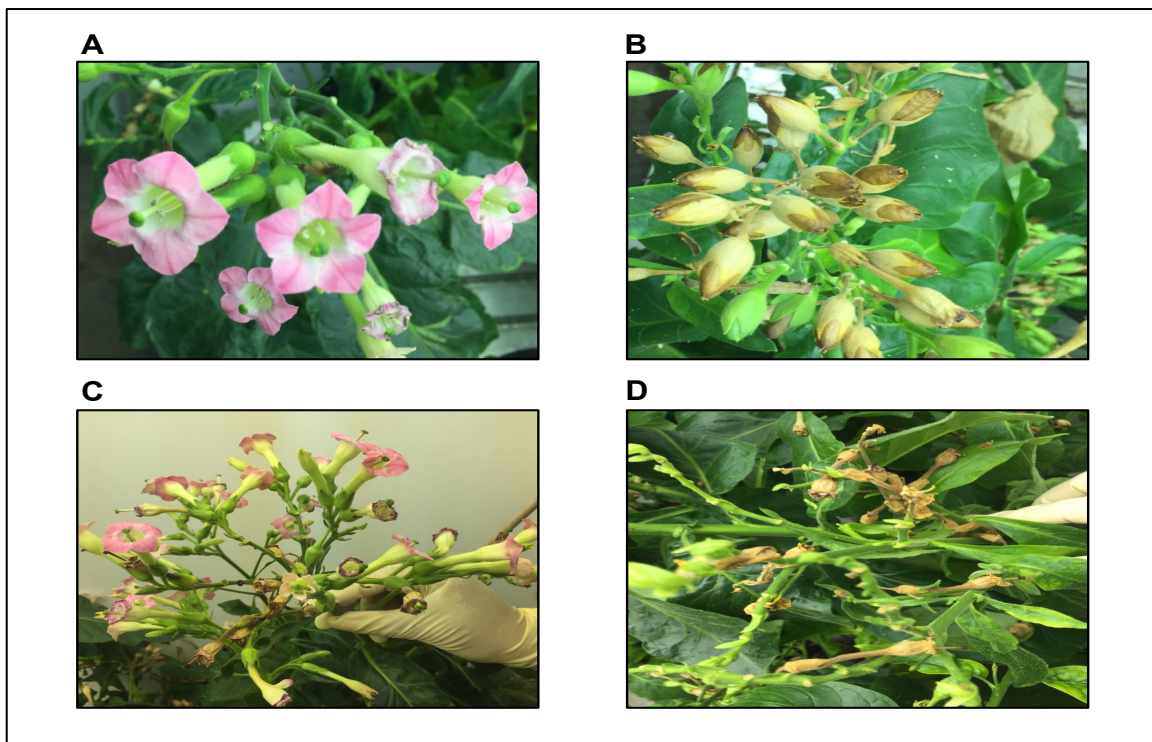
Figure A1B.1 Transgenic tobacco plants expressing geranium KCS 2
(A) Wild type flowers, (B) Wild type seeds, (C) KCS 2 primary transformants
flowers and (D) KCS 2 primary transformants sterile.

A C16:0

B C18:0

C C18:1$^{\Delta 9c}$

D C18:2

E C18:3

F Total

G

FDR p-values for comparison of each treatment

Fatty Acids	Vector vs. KCS 2	Wild Type vs. KCS 2	Wild Type vs. Vector
C16:0	0.05	1	0.10
C18:0	0.91	1	0.44
C18:1$^{\Delta 9c}$	0.91	1	0.98
C18:2	0.91	1	0.95
C18:3	0.35	1	0.44
Total	0.05	1	0.13
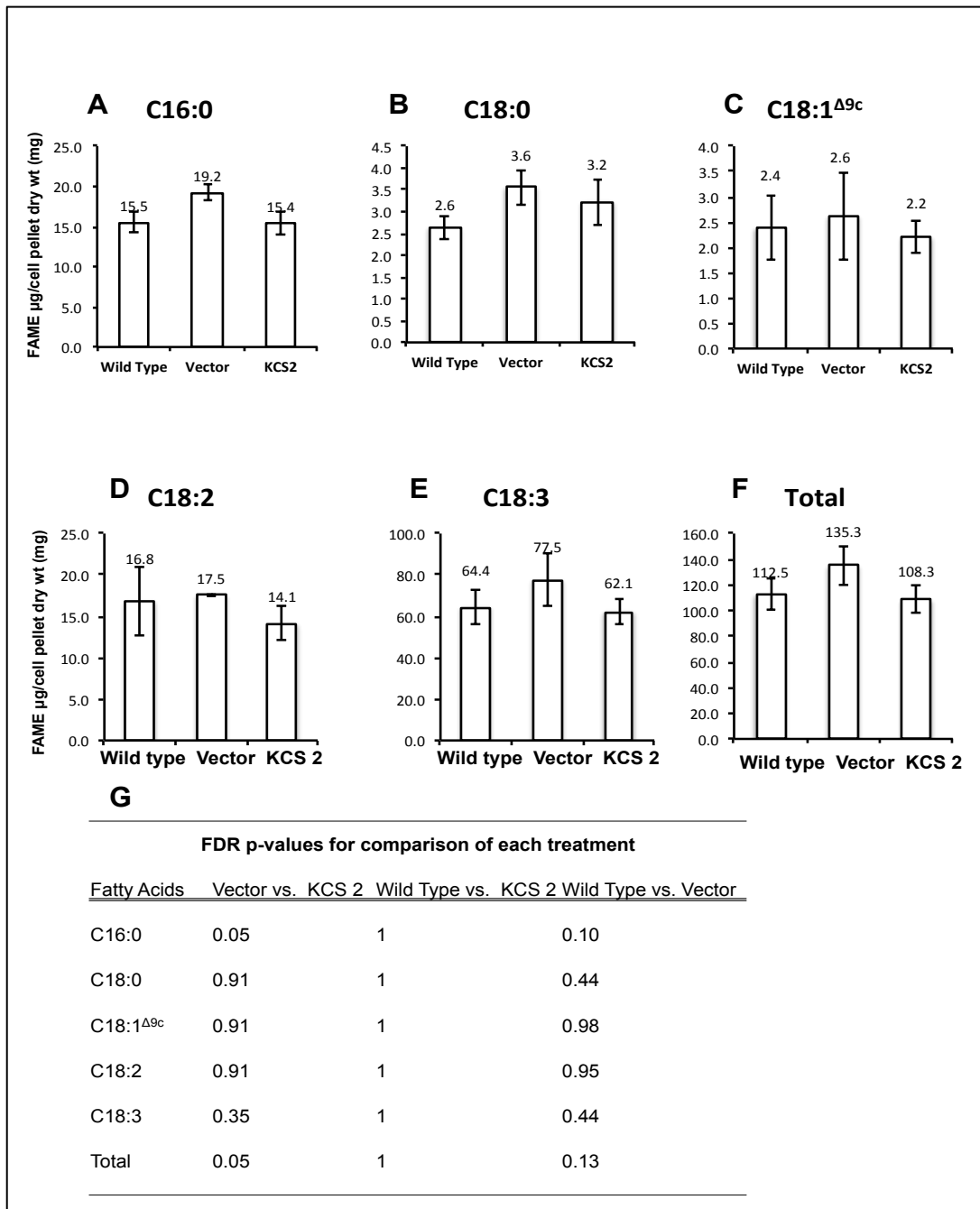
Figure A1B.2. Fatty acid methyl ester content of tobacco leaves expressing geranium KCS 2 cDNAs. Average values (n=3) are indicated with error bars representing standard deviation of wild type (tobacoo), pRI201-AN (vector alone) and transgenic plants expressing KCS 2. FDR p-values for one to one comparison of each treatment using two factor mixed model analysis. p-values <0.05 show significant difference in fatty acid amount between two genes.

APPENDIX 2

Table A2.1. Primer sequences for Thioesterase and β ketoacyl-ACP
Synthase EST clones used for primer walking analysis.

PRIMER desgin for Thioesterase EST clones		
EST CLONE #	PRIMER	SEQUENCE
03F05	5'	GGTGATGATGAACCAAGATACT
03F05	3'	GCCTGTAATCTAAGGTGATGGT
07A05	5'	CAGCTTGGAGTGATGTGATTGA
07A05	3'	TGCTGGTTCATGTCCAGATCAG
03C01	5'	ACCAAGATACCAGGCGACTTCAG
24B08	5'	GACCATTGCCAGTTTGTTGCAG
38EO6	5'	GGAAGAATTGGATGTAGACGCG
36C02	5'	GCCTGATTGAGTAGCATTGCACG
24B07	5'	GTCCCTGTAATGAGGCTGCATCT
40F05	5'	ATCCAAATCATCCCGTCCAGGA
30CO1	5'	CTGGTGTGCTTGTTGGAATGGGAT
03C01	5'	GTACCGTGTGGAGAAGAA
03C01	5'	GGATCTTCTAGCTTCGGT
21G11	5'	CCACACACTGGAGGCTCTTGTTA
21G11	5'	CTGACCCATATTGCCAAG
40F05	5'	CTCAGGGACTGTGATCCTA
03F05	5'	GTAGGAGCTAGATTGTCC
07A05	5'	CACACATAGGTGGCGTTTGA

PRIMER desgin for β ketoacyl-ACPsynthase EST clones		
EST CLONE #	PRIMER	SEQUENCE
02B12	5'	ATCGACCGATTCGACGCATCA
02B12	3'	ACTTCGTGTTGTTGCTTCGTAT
28B11	5'	TTGGAGCTGGACTGACTTG
28B11	3'	AGTGCCCAGGCTAGTCAGTAA
01B08	5'	GAGAGTCATCTGCACCTGGATTCT
24F04	5'	ACATCGACGGCAAAAACGACC
09EO3	5'	CGGTGTTTTCTGGTGGAGTTCAGA
15G03	5'	TGCCGGAGATTTGCTCTCCAA
22B11	5'	ATATGACCGATTCAAGGGCCGATG
20H06	5'	GCATATCTTCCTTGAAGCCCCA
02B12	3'	AGAAGAAAGGCCTCCGGATT
28B11	5'	ATCTGGGTTTCGCTCCAATG
28B11	5'	GACGGTGATGGCCAAAGACA
28B11	5'	GAATCCCATTACATTTCACC
09EO3	5'	ACAAACATGGGCTCTGCATTGC
09EO3	5'	CCGGTGATCTAGCTGAGGTAAATG
01B08	5'	GTAGCGCTGTCTAGTTCGTTTGAG
01B08	3'	TGCACATGCTACATCCACAC
20H06	5'	TTCGAACACTGATGCGCATC

Table A2.2. Details of sub-cloning for Thioesterase and β ketoacyl-ACP synthase.

GENE	EST Clones	Restriction Enyme	Liagted to vector
TE	03F05	BamHI	pBluescript SK minus
TE	24B07	BamHI	pBluescript SK minus
TE	24B08	NdeI	pGEM®-T Easy Vector
TE	07A05	BamHI	pBluescript SK minus
TE	07A05	ApaI	pBluescript SK minus
TE	38EO6	NdeI	pGEM®-T Easy Vector
TE	36C02	NdeI	pGEM®-T Easy Vector
KAS	24F04	EcoRI and BamHI	pBluescript SK minus
KAS	28B11	SacI and BamHI	pBluescript SK minus

Table A2.3. Primer sequences for Thioesterase and β ketoacyl-ACP synthase EST clones used for RACE

GENE	Primer	Sequence
FAT-B1	3'	GGA CCC AAT AAT CCC ACT GCA GCT GAC AA
FAT-B2	3'	GCT GCT ACT GCC GCC TCA TCA TT
FAT-B3	3'	GCG GTC CGA TCA GCG CCT ATC TCA TAA
KAS Ic	3'	ACA CCG GCA TCT TCT AGA CTC TTG
KAS II	3'	TGG ATG TGG CTC GGT CAT GTG ATA

.

Table A2.4. Primer sequences for SYBR Green Assay

Genes	Primer	Sequence
MAD	5'	TGA AGA GAA ATA GAA GAT GGG TGT TC
	3'	TCG TCT GGA AGC TCT CTC GTC CTC TCC TTA
SAD	5'	GCA AAG CAA TGG CTC TGA AGC TGA ATG CC
	3'	TGA CTT GGT CGT AGA ATC CAT CGG ACT CG
ω3	5'	CAG CTT ATG GCC ATG ACG ATT CCG ATT TCG
	3'	AAA CTC TTG TAG GTC TTC TCC GTC AAC GGA
ACTIN	5'	ACC TAC AAC TCA ATC ATG AAG TGT GAC GTG
	3'	GGA AAT ACG AAC TCA CCA CCA TAG CAC TTG
KAS Ia/b	5'	CGA CAC CTA CTA CGA AAA GCT C
	3'	CAA CGA GGC AGT ACC GGA GAG AAT C
KAS 1c	5'	GTG GGC ACA ATT CAG TGG TT
	3'	GTC TTC AAT ACT CGT CAC G
KAS III	5'	GGT TTC TCG CTC CTT CAG TTC
	3'	ACT AGC CTG GGC ACT CGA GAT T

Table A2.5. Primer sequences for TaqMan assay.

Gene	Primer/Probe	Sequences
SAD	5'	GAAACCGAATCGCAGAGCAA
	3'	GGAAAGTGGTGGTGGTGGA
	Probe	AATGGCTCTGAAGCTGAA
Actin	5'	CGCATCCCTCAGCACCTT
	3'	GGCCGGACTCGTCGTATTC
	Probe	CAGCAGATGTGGATTT
MAD	5'	TTCCAGTAGTAGCATCTGCTGCTT
	3'	CCATTACACCAGTTACTCCAACTTTTC
	Probe	CATTTCCAAGGTTAATCAT
ω3	5'	GCTGCGTTTCGCTTGGA
	3'	GCCCTGGACAGCCCAGTAG
	Probe	TTGGCTTGTTTGGCC
ACP 1	5'	AGAAGAAGGCTCCTGCTTAAATTAGC
	3'	ACAAATTATAAGCAAGTCCACTAGGCAAA
	Probe	ATGGAGAAGAAGGATTTTG
ACP 2	5'	TTAAAATAGTTTGGAGAGGCAACTG
	3'	ATAGAAACAGGTGAATTGTGCATGTA
	Probe	CCGAGACAGTATACTAC
FAT-A1	5'	TGTCTCTCGCAGCCATTTCA
	3'	CTGGTTGCGAGAGAGAGAAGAGAA
	Probe	TCGCCGAAGAACAA
FAT-A2	5'	TCGAAAACCACTCTCCATTTCTC
	3'	GTTTTGGAGATGGACGATTTGA
	Probe	CCCAAACTCACTCTCA
FAT-A3	5'	CAGGTTTAATGGTACCCCGGTT
	3'	CATAGGGGCCCGTGACC
	Probe	CCCAGAATTTATAGACCCATT

APPENDIX 3
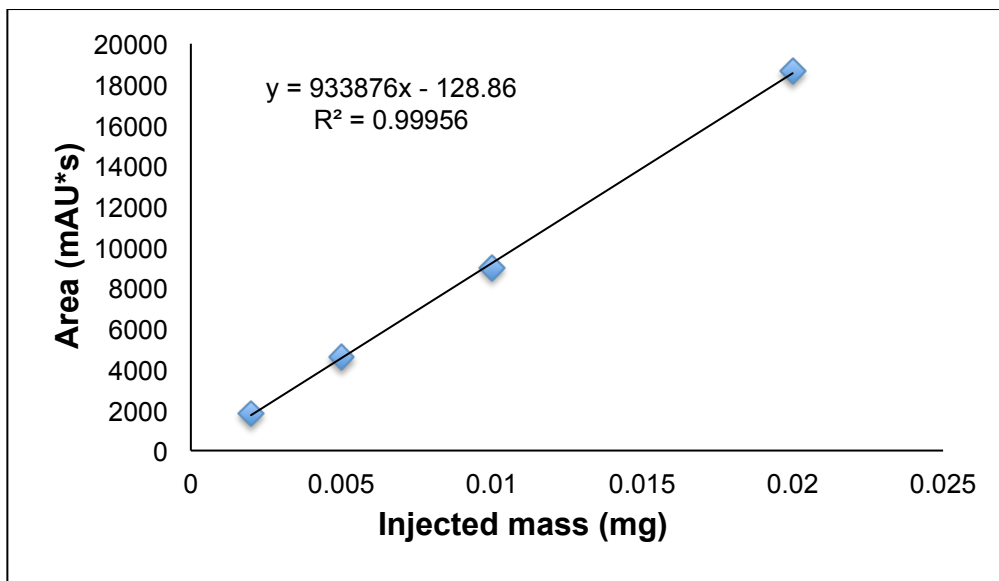
Figure 3A.1. Standard curve for 22:1n5 anacardic acid.
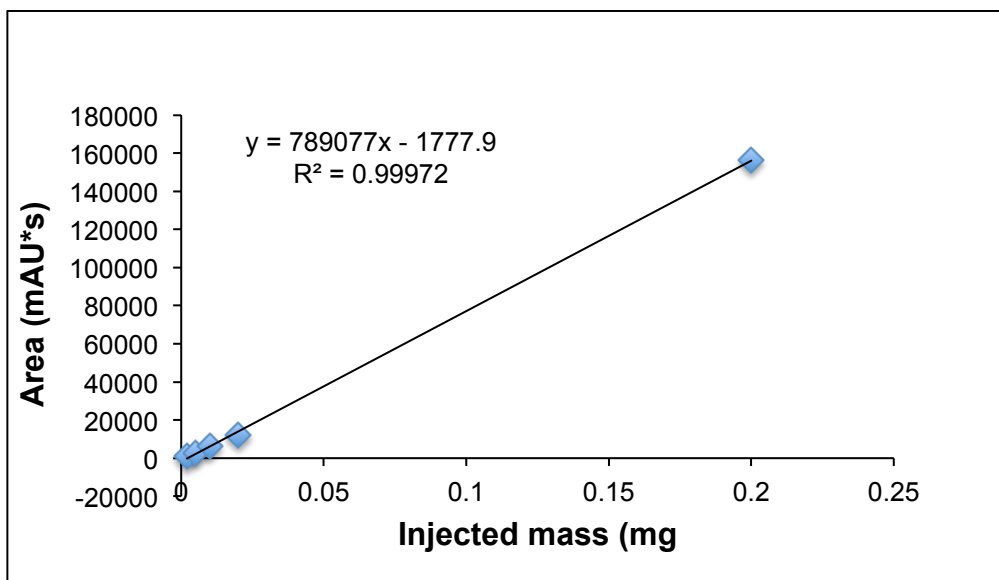
Figure 3A.2. Standard curve for 24:1n5 anacardic acid.

Table A3.1. List of GC standards

Fatty Acid Standards	Chemical Name	Retention Time (minutes)
C8:0	Methyl octanoate 1.9 wt. %	7.89
C10:0	Methyl decanoate 3.2 wt. %	8.31
C12:0	Methyl dodecanoate 6.4 wt. %	9.96
C13:0	Methyl tridecanoate 3.2 wt. %	11.12
C14:0	Methyl myristate 3.2 wt. %	12.53
C14:1$^{\Delta 9}$	Methyl myristoleate 1.9 wt. %	14.01
C15:0	Methyl pentadecanoate 1.9 wt. %	14.21
C16:0	Methyl palmitate 13 wt. %	16.17
C16:1	Methyl palmitoleate 6.4 wt. %	17.70
C17:0	Methyl heptadecanoate 3.2 wt. %	18.26
C18:0	Methyl stearate 6.5 wt. %	20.62
C18:1$^{\Delta 9t}$	Methyl oleate 19.6 wt. %	21.68
C18:1$^{\Delta 9c}$	Methyl elaidate 2.6 wt. %	22.24
C18:2$^{\Delta 6c}$	Methyl linoleate 13 wt. %	24.60
C20:0	Methyl arachidate 1.9 wt. %	25.57
C20:1	Methyl *cis*-11-eicosenoate 1.9 wt. %	27.28
C18:3$^{\Delta 3}$	Methyl linolenate 6.4 wt. %	27.57
C22:0	Methyl behenate 1.9 wt. %	31.61
C22:1$^{\Delta 9}$	Methyl *cis*-13-docosenoate 1.9 wt. %	33.82
C18:1$^{\Delta 11}$	cis vaccenic acid	22.70

Table A3.2. Correlation data for all the candidate genes, AnAc and UMFAs

Comparison Name 1	Comparison Name 2	correlation	pvalue	FDRpvalue	qvalues
AA22.mean	AA24.mean	0.969	0.000	0.002	0.001
C16c1Delta11	C18c1Delta13	0.978	0.000	0.002	0.001
AA22.mean	C16c1Delta11	0.925	0.001	0.010	0.003
AA22.mean	C18c1Delta13	0.956	0.000	0.003	0.001
AA24.mean	C16c1Delta11	0.935	0.001	0.007	0.003
AA24.mean	C18c1Delta13	0.968	0.000	0.002	0.001
AA22.mean	2^(-deltaCTMADTrichm.mean)	0.563	0.146	0.206	0.071
AA22.mean	2^(-deltaCTSADTrichm.mean)	0.719	0.044	0.103	0.035
AA22.mean	2^(-deltaCTw3Trichm.mean)	0.873	0.005	0.039	0.013
AA22.mean	2^(-deltaCTACP1Trichm.mean)	0.785	0.021	0.072	0.025
AA22.mean	2^(-deltaCTACP2Trichm.mean)	0.751	0.032	0.089	0.031
AA22.mean	2^(-deltaCTFATA1Trichm.mean)	0.084	0.843	0.852	0.295
AA22.mean	2^(-deltaCTFATA2Trichm.mean)	-0.750	0.032	0.089	0.031
AA22.mean	2^(-DeltaCTMAD_SybrTrichm.mean)	-0.665	0.072	0.138	0.048
AA22.mean	2^(-DeltaCTSAD_SybrTrichm.mean)	0.685	0.061	0.126	0.043
AA22.mean	2^(-DeltaCTw3_SybrTrichm.mean)	-0.734	0.038	0.092	0.032
AA22.mean	2^(-DeltaCTKASIab_SybrTrichm.mean)	0.658	0.076	0.138	0.048
AA22.mean	2^(-DeltaCTKASIc_SybrTrichm.mean)	0.743	0.035	0.089	0.031
AA22.mean	2^(-DeltaCTKASIII_SybrTrichm.mean)	0.141	0.738	0.793	0.274
AA24.mean	2^(-deltaCTMADTrichm.mean)	0.530	0.177	0.223	0.077
AA24.mean	2^(-deltaCTSADTrichm.mean)	0.577	0.134	0.200	0.069
AA24.mean	2^(-deltaCTw3Trichm.mean)	0.859	0.006	0.045	0.016
AA24.mean	2^(-deltaCTACP1Trichm.mean)	0.811	0.015	0.071	0.025
AA24.mean	2^(-deltaCTACP2Trichm.mean)	0.788	0.020	0.072	0.025
AA24.mean	2^(-deltaCTFATA1Trichm.mean)	-0.098	0.817	0.846	0.292
AA24.mean	2^(-deltaCTFATA2Trichm.mean)	-0.752	0.032	0.089	0.031
AA24.mean	2^(-DeltaCTMAD_SybrTrichm.mean)	-0.645	0.084	0.147	0.051
AA24.mean	2^(-DeltaCTSAD_SybrTrichm.mean)	0.586	0.127	0.193	0.067
AA24.mean	2^(-DeltaCTw3_SybrTrichm.mean)	-0.688	0.060	0.126	0.043
AA24.mean	2^(-DeltaCTKASIab_SybrTrichm.mean)	0.545	0.162	0.211	0.073
AA24.mean	2^(-DeltaCTKASIc_SybrTrichm.mean)	0.638	0.089	0.147	0.051
AA24.mean	2^(-DeltaCTKASIII_SybrTrichm.mean)	0.079	0.852	0.852	0.295
C16c1Delta11	2^(-deltaCTMADTrichm.mean)	0.347	0.400	0.464	0.161
C16c1Delta11	2^(-deltaCTSADTrichm.mean)	0.464	0.247	0.298	0.103
C16c1Delta11	2^(-deltaCTw3Trichm.mean)	0.811	0.015	0.071	0.025
C16c1Delta11	2^(-deltaCTACP1Trichm.mean)	0.843	0.008	0.055	0.019
C16c1Delta11	2^(-deltaCTACP2Trichm.mean)	0.824	0.012	0.069	0.024
C16c1Delta11	2^(-deltaCTFATA1Trichm.mean)	-0.134	0.752	0.793	0.274
C16c1Delta11	2^(-deltaCTFATA2Trichm.mean)	-0.713	0.047	0.105	0.036
C16c1Delta11	2^(-DeltaCTMAD_SybrTrichm.mean)	-0.632	0.093	0.147	0.051
C16c1Delta11	2^(-DeltaCTSAD_SybrTrichm.mean)	0.566	0.144	0.206	0.071
C16c1Delta11	2^(-DeltaCTw3_SybrTrichm.mean)	-0.674	0.067	0.133	0.046
C16c1Delta11	2^(-DeltaCTKASIab_SybrTrichm.mean)	0.556	0.152	0.206	0.071
C16c1Delta11	2^(-DeltaCTKASIc_SybrTrichm.mean)	0.631	0.093	0.147	0.051
C16c1Delta11	2^(-DeltaCTKASIII_SybrTrichm.mean)	0.158	0.709	0.776	0.268
C18c1Delta13	2^(-deltaCTMADTrichm.mean)	0.362	0.378	0.448	0.155
C18c1Delta13	2^(-deltaCTSADTrichm.mean)	0.504	0.203	0.251	0.087
C18c1Delta13	2^(-deltaCTw3Trichm.mean)	0.789	0.020	0.072	0.025
C18c1Delta13	2^(-deltaCTACP1Trichm.mean)	0.796	0.018	0.072	0.025
C18c1Delta13	2^(-deltaCTACP2Trichm.mean)	0.785	0.021	0.072	0.025
C18c1Delta13	2^(-deltaCTFATA1Trichm.mean)	-0.171	0.686	0.767	0.265
C18c1Delta13	2^(-deltaCTFATA2Trichm.mean)	-0.769	0.026	0.083	0.029
C18c1Delta13	2^(-DeltaCTMAD_SybrTrichm.mean)	-0.662	0.074	0.138	0.048
C18c1Delta13	2^(-DeltaCTSAD_SybrTrichm.mean)	0.556	0.153	0.206	0.071
C18c1Delta13	2^(-DeltaCTw3_SybrTrichm.mean)	-0.742	0.035	0.089	0.031
C18c1Delta13	2^(-DeltaCTKASIab_SybrTrichm.mean)	0.544	0.164	0.211	0.073
C18c1Delta13	2^(-DeltaCTKASIc_SybrTrichm.mean)	0.637	0.089	0.147	0.051
C18c1Delta13	2^(-DeltaCTKASIII_SybrTrichm.mean)	0.170	0.688	0.767	0.265

APPENDIX 4

1Trichome-18C_S13

Quality scores across all bases (Sanger / Illumina 1.9 encoding)

Position in read (bp)

4Trichome-23C_S14

Quality scores across all bases (Sanger / Illumina 1.9 encoding)

Position in read (bp)

7Pedicle-18C_S15

176
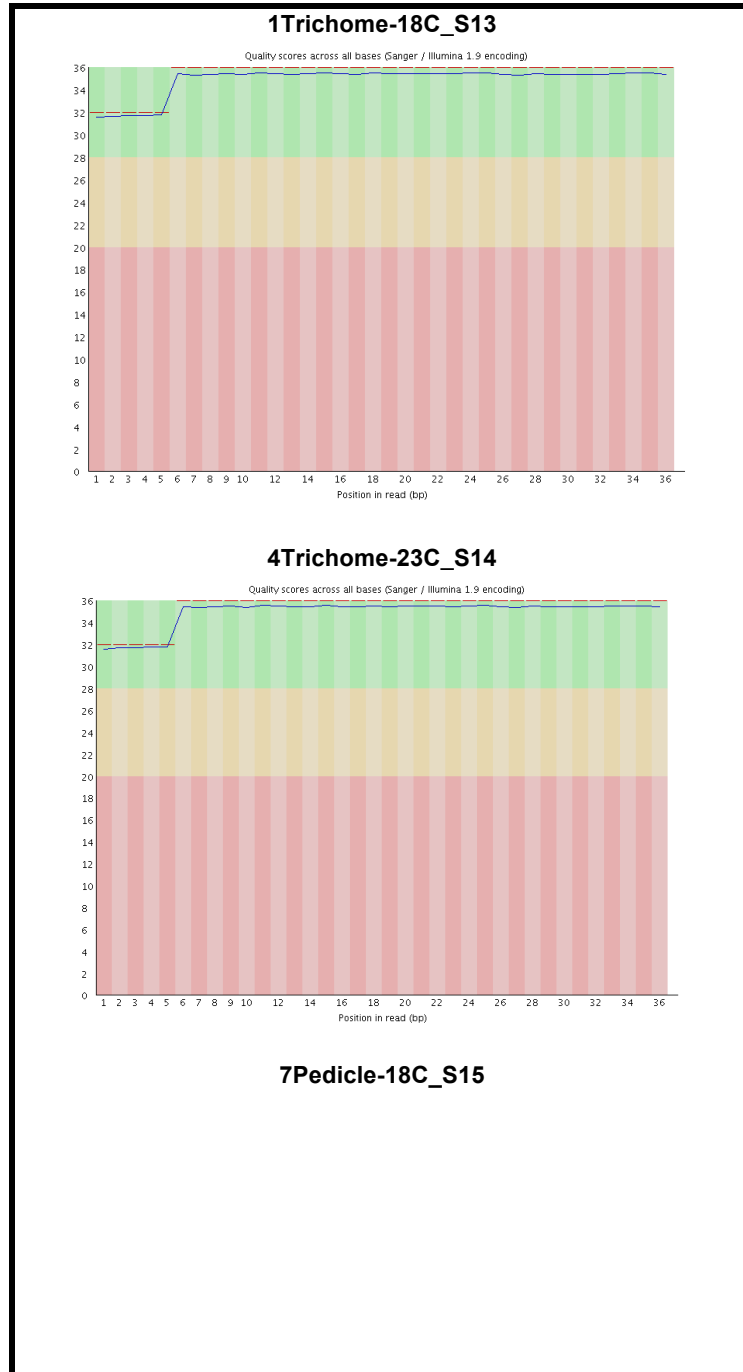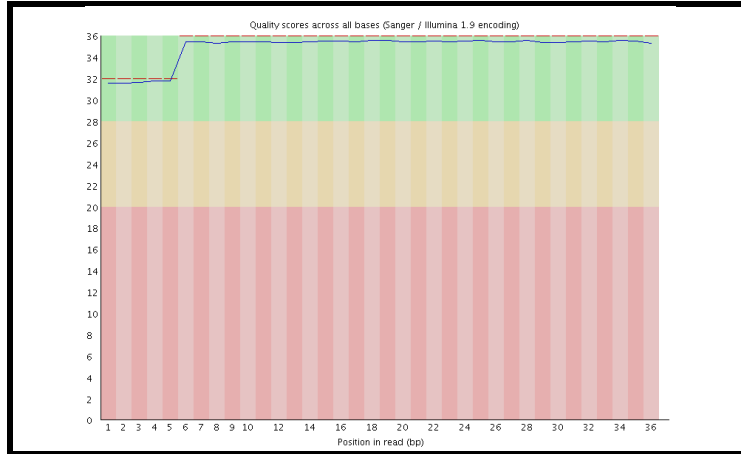
Figure A4.1. Quality Score for untrimmed miRNAs.

Table A4.1. Description of TruSeq small RNA adapter and primer sequences

Name	Sequence
RNA 5' Adapter	GTTCAGAGTTCTACAGTCCGACGATC GATCGTCGGACTGTAGAACTCTGAAC
RNA 3' Adapter	TGGAATTCTCGGGTGCCAAGG CCTTGGCACCCGAGAATTCCA
RNA RT Primer	AATGATACGGCGACCACCGAGATCTACACGTTCAGAGTTCTACAGTCCGA TCGGACTGTAGAACTCTGAACGTGTAGATCTCGGTGGTCGCCGTATCATT
RNA PCR Primer 13	CAAGCAGAAGACGGCATACGAGATTTGACTGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA TGGAATTCTCGGGTGCCAAGGAACTCCAGTCACAGTCAAATCTCGTATGCCGTCTTCTGCTTG
RNA PCR Primer 14	CAAGCAGAAGACGGCATACGAGATGGAACTGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA TGGAATTCTCGGGTGCCAAGGAACTCCAGTCACAGTTCCATCTCGTATGCCGTCTTCTGCTTG
RNA PCR Primer 15	CAAGCAGAAGACGGCATACGAGATTGACATGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA TGGAATTCTCGGGTGCCAAGGAACTCCAGTCACATGTCAATCTCGTATGCCGTCTTCTGCTTG

Table A4.2. Summary of sequence analysis reads.

Sample	Raw Reads	After Primer/ Adapter Trimming	After Size Trimming
1Trichome-18C_S13	12,894,586	10,947,284	3,723,972
4Trichome-23C_S14	13,838,346	13,047,087	5,302,277
7Pedicle-18C_S15	15,824,800	7,857,527	3,912,781
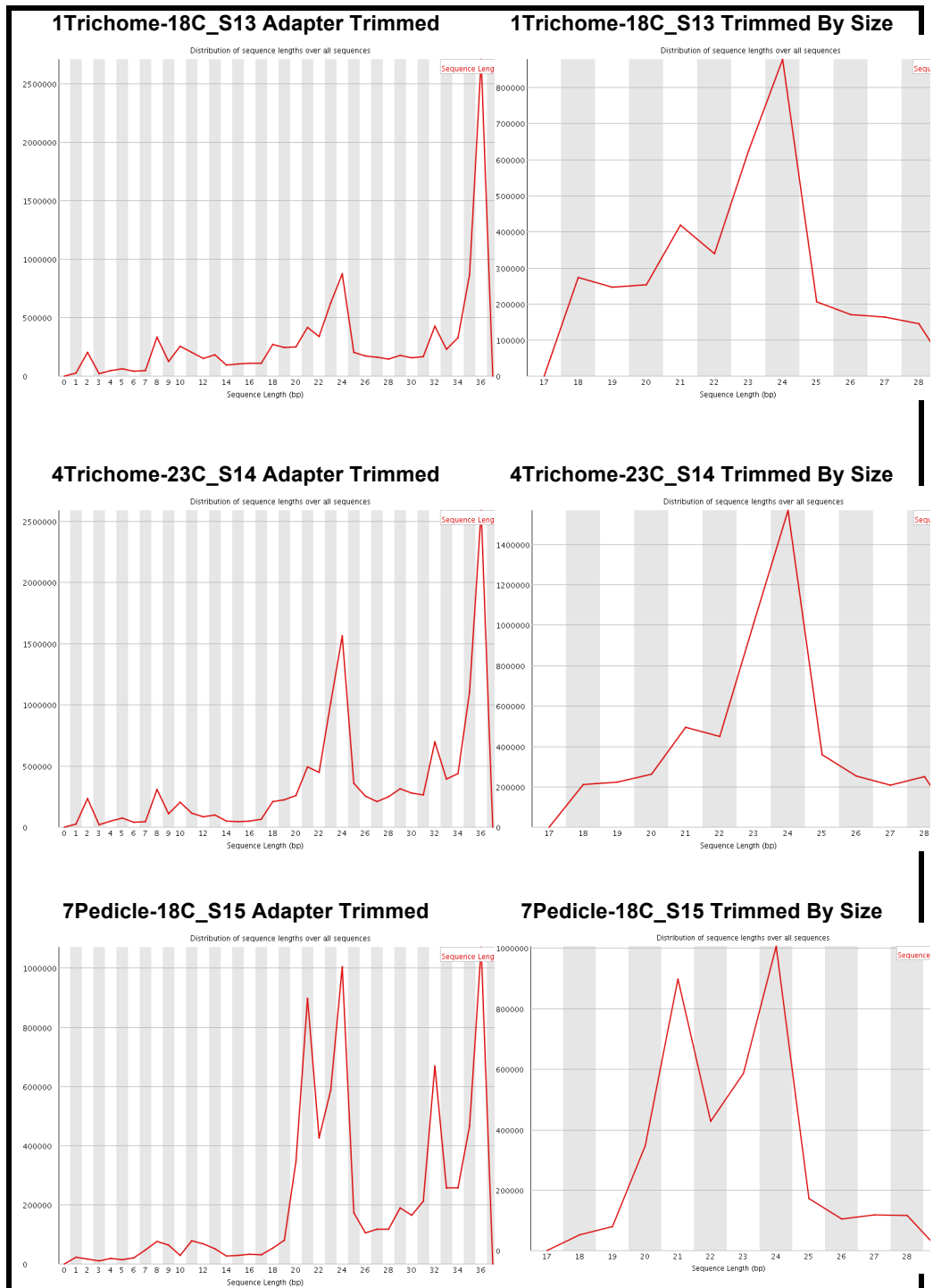
Figure A4.2. Sequence length distribution for trimmed sequences

Table A4.3: *Pelargonium × hortorum* miRNA sequences.

>pxh-miR-425-5p	>pxh-miR-319	>pxh-miR-2276-3p
AUGACACGAUCACUCCCGUUGA	UUGGACUGAAGGGAGCUCCCU	UCUGCAAGUGUCAGAGGCGAGG
>pxh-miR-1246	>pxh-miR-183-5p	>pxh-miR-501-3p
AAUGGAUUUUUGGAGCAGG	UAUGGCACUGGUAGAAUUCACUGA	AAUGCACCUGGGCAAGGAUUCA
>pxh-miR-9	>pxh-miR-34a-5p	>pxh-miR-6476a
UCUUUGGUUAUCUAGCUGUAUGA	AGGCAGUGUAGUUAGCUGAUUGC	UCAGUGGAGAUGAAACAUGA
>pxh-miR-1307	>pxh-miR-128	>pxh-miR-29a
ACUCGGCGUGGCGUCGGUCGUG	UCACAGUGAACCGGUCUCUUUU	CUAGCACCAUCUGAAAUCGGUU
>pxh-miR-1260	>pxh-miR-345	>pxh-miR-584
AUCCCACCGCUGCCACCA	GCUGACUCCUAGUCCAGGGCUCG	UUAUGGUUUGCCUGGGACUGA
>pxh-miR-335-3p	>pxh-miR-34-5p	>pxh-miR-21a-5p
UUUUUCAUUAUUGCUCCUGACC	UGGCAGUGUGGUUAGCUGGUUG	UAGCUUAUCAGACUGAUGUUGA
>pxh-miR-238	>pxh-miR-7501	>pxh-miR-15
UUUGUACUCCGAUGCCAUUCAGA	AUAUCUGAUUCUGACACGAAAAAA	UAGCAGCACAUCAUGGUUUACA
>pxh-miR-170-5p	>pxh-miR-5532	>pxh-miR-489-3p
UAUUGGCCUGGUUCACUCAGA	AUGGAAUAUAUGACAAAGGUGG	GUGACAUCACAUAUACGGCAGC
>pxh-miR-190	>pxh-miR-92a-3p	>pxh-miR-103-3p
UGAUAUGUUUGAUAUAUUAGG	UAUUGCACUUGUCCCGGCCUGU	AGCAGCAUUGUACAGGGCUAUG
>pxh-miR-8716	>pxh-miR-156b	>pxh-miR-532
AUGUGUCAAAAUGUGAGGCUGUCA	UGACAGAAGAGAGUGAGCAC	CAUGCCUUGAGUGUAGGACCGU
>pxh-miR-410	>pxh-miR-195	>pxh-miR-2478
AAUAUAACACAGAUGGCCUGU	UAGCAGCACAGAAAUAUUGGCA	GUAUCCCACUUCUGACACCA
>pxh-miR-828	>pxh-miR-146-5p	>pxh-miR-1895
UCUUGCUCAAAUGAGUAUUCCA	UGAGAACUGAAUUCCAUAGGCU	CCCCCGAGGAGGACGAGGAGGA
>pxh-miR-93	>pxh-miR-172	>pxh-miR-1448
AAAGUGCUGUUCGUGCAGGUAG	UGAGAAUCUUGAUGAUGCUGCAU	CUUUCCAACGCCUCCCAUAC
>pxh-miR-281-5p	>pxh-miR-702-3p	>pxh-let-7g
AAGAGAGCUUAUCCGUAGACAG	UGCCCACCCUUUACCCCACUCCA	UGAGGUAGUAGUUUGUACAGUU
>pxh-miR-15a	>pxh-miR-5083	>pxh-miR-397-5p
UAGCAGCACAUAAUGGUUUGUG	AGACUACAAUUAUCUGAUCA	UCAUUGAGUGCAGCGUUGAUG
>pxh-miR-877	>pxh-miR-31	>pxh-miR-454-3p
GUAGAGGAGAUGGCGCAGGG	AGGCAAGAUGCUGGCAUAGCU	UAGUGCAAUAUUGCUUAUAGGGU
>pxh-miR-541-5p	>pxh-miR-589	>pxh-miR-744
AAGGGAUUCUGAUGUUGGUCACACU	UGAGAACCACGUCUGCUCUGA	UGCGGGGCUAGGGCUAACAGCA
>pxh-miR-516b	>pxh-miR-378-3p	>pxh-miR-20
AUCUGGAGGUAAGAAGCACUUU	ACUGGACUUGGAGUCAGAAGGC	UAAAGUGCUUAUAGUGCAGGUA
>pxh-miR-374-5p	>pxh-miR-8-3p	>pxh-miR-199a-5p
AUAUAAUACAACCUGCUAAGUG	UAAUACUGUCAGGUAAAGAUGUC	CCCAGUGUUCAGACUAGCUGUUC
>pxh-miR-827	>pxh-miR-738	>pxh-miR-142-5p
UUAGAUGACCAUCAACAAACA	GCUACGGCCCGCGUCGGGA	CAUAAAGUAGAAAGCACUAC
>pxh-miR-125	>pxh-miR-4171-5p	>pxh-miR-3931-3p
UCCCUGAGACCCUAACUUGUGA	UGACUCUCUUAAGGAAGCCA	UACUUUGAGUCGGUACGAAUCA
>pxh-miR-301a-3p	>pxh-miR-92-3p	>pxh-let-7b-5p
CAGUGCAAUAGUAUUGUCAAAGC	UAUUGCACUCGUCCCGGCCUGA	UGAGGUAGUAGGUUGUGUGGU
>pxh-miR-6478	>pxh-miR-3120	>pxh-miR-20a-5p
CCGACCUUAGCUCAGUUGGUG	CACAGCAAGUGUAGACAGGCA	UAAAGUGCUUAUAGUGCAGGUA
>pxh-miR-6134	>pxh-miR-5181-3p	>pxh-miR-182
UGAGGUAGUAGGAUGUAGA	CACUUAUUUUGGACCGGAGGG	UUUGGCAAUGGUAGAACUCACACU
>pxh-miR-138	>pxh-miR-548d-5p	>pxh-miR-4792
AGCUGGUGUUGUGAAUCAGGC	AAAAGUAAUUGUGGUUUUUGCC	CGGUGAGCGCUCGCUGGC

>pxh-miR-30a-5p	>pxh-miR-484	>pxh-miR-941
UGUAAACAUCCUCGACUGGAAG	UCAGGCUCAGUCCCCUCCCGAU	CACCCGGCUGUGUGCACAUGUGC
>pxh-miR-130a	>pxh-miR-398	>pxh-miR-167-5p
CAGUGCAAUGUUAAAAGGGCAU	UGUGUUCUCAGGUCGCCCCUG	UGAAGCUGCCAGCAUGAUCUU
>pxh-miR-275	>pxh-miR-9226	>pxh-miR-5298d
UCAGGUACCUGAAGUAGCGCGCG	UCAAGUCCCUGUUCGGGCGCCG	UGGAGAUGAUAUGAAGAUGAAAAA
>pxh-miR-550-5p	>pxh-miR-28-3p	>pxh-miR-429
AGUGCCUGAGGGAGUAAGAGCCC	CACUAGAUUGUGAGCUCCUGGA	UAAUACUGUCUGGUAAAACCG
>pxh-miR-157a-5p	>pxh-miR-171	>pxh-miR-8109
UUGACAGAAGAUAGAGAGCAC	UGAUUGAGCCGUGCCAAUAUC	GCGCCGCGUGCCGGCCGCGGG
>pxh-miR-169	>pxh-miR-5139	>pxh-miR-625-3p
UAGCCAAGGAUGACUUGCCUA	AAACCUGGCUCUGAUACCA	GACUAUAGAACUUUCCCCCUCA
>pxh-miR-424-3p	>pxh-miR-654-5p	>pxh-miR-156aa
CAAAACGUGAGGCGCUGCUAU	AAAGGUGGUGGGCUGCGGAGCAUG	AUUGGAGUGAAGGGAGCU
>pxh-miR-23-3p	>pxh-miR-2111-5p	>pxh-miR-156
AUCACAUUGCCAGGGAUUACC	UAAUCUGCAUCCUGAGGUUUA	CUGACAGAAGAGAGUGAGCAC
>pxh-miR-203	>pxh-miR-653-5p	>pxh-miR-423-3p
GUGAAAUGUUUAGGACCACUAG	GUGUUGAAACAAUCUCUACUG	AAGCUCGGUCUGAGGCCCCUCAGU
>pxh-miR-1128	>pxh-miR-9-5p	>pxh-miR-252-5p
UACUACUCCCUCCGUCCCAAA	UCUUUGGUUAUCUAGCUGUA	CUAAGUAGUAGUGCCGCAGGUAA
>pxh-miR-16a-5p	>pxh-miR-5106	>pxh-miR-322-3p
UAGCAGCACGUAAAUAUUGGCG	AGGUCUGUAGCUCAGUUGGCAGA	AAACAUGAAGCGCUGCAACAC
>pxh-miR-1507-3p	>pxh-let-7e	>pxh-miR-124b-3p
CCUCGUUCCAUACAUCAUCUAG	UGAGGUAGGAGGUUGUAUAGU	UAAGGCACGCGGUGAAUGCUGA
>pxh-miR-5168-3p	>pxh-miR-135-5p	>pxh-miR-27-3p
UCGGACCAGGCUUCAAUCCCU	UAUGGCUUUUUAUUCCUAUGUGA	UUCACAGUGGCUAAGUUCUGC
>pxh-miR-3607-3p	>pxh-miR-192	>pxh-miR-153
ACUGUAAACGCUUUCUGAUG	CUGACCUAUGAAUUGACAGCCAG	UUGCAUAGUCACAAAGUGAUG
>pxh-let-7f	>pxh-miR-221-3p	>pxh-miR-24
UGAGGUAGUAGAUUGUAUAGU	AGCUACAUUGUCUGCUGGGUUU	UGGCUCAGUUCAGCAGGAACAG
>pxh-miR-218	>pxh-miR-340	>pxh-miR-186
UUGUGCUUGAUCUAACCAUGU	UUAUAAAGCAAUGAGACUGAUU	CAAAGAAUUCUCCUUUUGGGCU
>pxh-miR-165a-5p	>pxh-miR-395	>pxh-miR-840-5p
GAAUGUUGUCUGGAUCGAGG	CUGAAGUGUUUGGGGGAACUC	ACACUAAAGGACCUAAACUAAC
>pxh-miR-3968	>pxh-miR-160	>pxh-miR-858
CGAAUCCCACUCCAGACACCA	UGCCUGGCUCCCUGUAUGCCA	UCUCGUUGUCUGUUCGACCUU
>pxh-miR-6483	>pxh-miR-4332	>pxh-miR-8649
UAUUGUAGAAAUUUUCAGGAUC	CACGGCCGCCGCCGGGCGCC	ACACUGUGGAAGUGGAUCUCUCUC
>pxh-miR-21	>pxh-miR-181a	>pxh-miR-7767-5p
UAGCUUAUCAGACUGAUGUUGA	AACAUUCAACGCUGUCGGUGA	CCCCAAGCUGAGAGCUCUCCC
>pxh-miR-63-3p	>pxh-miR-500-3p	>pxh-let-7
UAUGACACUGAAGCGAGUUGGAAA	AUGCACCUGGGCAAGGAUUCU	UGAGGUAGUAGGUUGUAUAGUU
>pxh-miR-2778a-5p	>pxh-miR-629-5p	>pxh-miR-101
GUUUUUUUGCAUAUCCUGCA	UGGGUUUACGUUGGGAGAACU	UACAGUACUGUGAUAACUGAAG
>pxh-miR-200a-3p	>pxh-miR-8155	>pxh-miR-660
UAACACUGUCUGGUAACGAUGU	UAACCUGGCUCUGAUACCA	UACCCAUUGCAUAUCGGAGUUG
>pxh-miR-2916	>pxh-miR-3934-5p	>pxh-miR-3018
UGGGGACUCGAAGACGAUCAUAU	UCAGGUGUGGAAACUGAGGCAG	AAAGAAUAGAAAAUCGAAGGUG
>pxh-miR-235-3p	>pxh-miR-396	>pxh-miR-8175
UAUUGCACUCGUCCCGGCCUGA	UUCCACAGCUUUCUUGAACUU	GAUCCCCGGCAACGGCGCCA

Table A4.3 continued

>pxh-miR-4505	>pxh-miR-769	>pxh-miR-4451
AGGCUGGGCUGGGACGGA	UGAGACCUCUGGGUUCUGAGC	UGGUAGAGCUGAGGACA
>pxh-miR-6813-5p	>pxh-miR-141	>pxh-miR-100
CAGGGGCUGGGGUUUCAGGUUCU	UAACACUGUCUGGUAAAGAUGG	AACCCGUAGAUCCGAACUUGUG
>pxh-miR-26	>pxh-miR-24a-3p	>pxh-miR-2170
UUCAAGUAAUCCAGGAUAGGCU	UGGCUCAGUUCAGCAGGAACAG	ACAGUGAAAUUUUGUAGAGA
>pxh-miR-477a	>pxh-miR-393	>pxh-miR-6167
ACUCUCCCUCAAGGGCUUCUG	UCCAAAGGGAUCGCAUUGAUCU	UACCCAGGUGGAAGCUUUGA
>pxh-miR-148-3p	>pxh-miR-193-5p	>pxh-miR-403-3p
UCAGUGCACUACAGAACUUUGU	UGGGUCUUUGCGGGCGAGAUG	UUAGAUUCACGCACAAACUCG
>pxh-miR-3954	>pxh-miR-159	>pxh-let-7d-5p
UGGACAGAGAAAUCACGGUCA	UUUGGAUUGAAGGGAGCUCUA	AGAGGUAGUAGGUUGCAUAGU
>pxh-miR-5205b	>pxh-miR-548ay-5p	>pxh-miR-149-5p
CUUAUAAUUAGGGACGGAGGGAGU	AAAAGUAAUUGUGGUUUUUGC	UCUGGCUCCGUGUCUUCACUCCC
>pxh-miR-1436	>pxh-miR-107	>pxh-miR-5072
ACAUUAUGGGACGGAGGGAGU	AGCAGCAUUGUACAGGGCUAUGA	CGAUUCCCCAGCGGAGUCGCCA
>pxh-miR-140-3p	>pxh-miR-486	>pxh-miR-4286
UACCACAGGGUAGAACCACGG	UCCUGUACUGAGCUGCCCCGAG	ACCCCACUCCUGGUACC
>pxh-miR-191-5p	>pxh-miR-5538	>pxh-miR-7158-3p
CAACGGAAUCCCAAAAGCAGCUG	ACUGAACUCAAUCACUUGCUGC	CUGAACUAGAGAUUGGGCCCA
>pxh-miR-330	>pxh-miR-142a-5p	>pxh-miR-5368
GCAAAGCACACGGCCUGCAGAGA	CAUAAAGUAGAAAGCACUACU	GGACAGUCUCAGGUAGACA
>pxh-miR-23a-3p	>pxh-miR-1839	>pxh-miR-324
AUCACAUUGCCAGGGAUUUCC	AAGGUAGAUAGAACAGGUCUUGUU	CGCAUCCCCUAGGGCAUUGGUGU
>pxh-miR-168	>pxh-miR-6788-5p	>pxh-miR-1582
UCGCUUGGUGCAGGUCGGGA	CUGGGAGAAGAGUGGUGAAGA	GAAAGAGAGCCAGAACACAG
>pxh-miR-4504	>pxh-miR-394	>pxh-miR-6240
UGUGACAAUAGAGAUGAACAUG	UUGGCAUUCUGUCCACCUCC	CCAAAGCAUCGCGAAGGCCCACGGCG
>pxh-miR-1910-5p	>pxh-miR-421	>pxh-miR-16
CCAGUCCUGUGCCUGCCGCCU	AUCAACAGACAUUAAUUGGGCGC	UAGCAGCACGUAAAUAUUGGCG
>pxh-miR-195a-5p	>pxh-miR-6300	>pxh-miR-5119
UAGCAGCACAGAAAUAUUGGC	GUCGUUGUAGUAUAGUGG	CAUCUCAUCCUGGGGCUGG
>pxh-miR-1692	>pxh-miR-10a-5p	>pxh-miR-5735-3p
UGUAGCUCAGUUGGUAGAGU	UACCCUGUAGAACCGAAUUUGU	UGGACAACAGGAUAAUGGCGU
>pxh-let-7a	>pxh-miR-7475-5p	>pxh-miR-5049c
UGAGGUAGUAGGUUGUAUAGUU	CCGCCGCCGCCGCGCCCUCC	AGACAAUUAUUUUGGGACGGAGG
>pxh-miR-130	>pxh-miR-22	>pxh-miR-1468-5p
CAGUGCAAUGUUAAAAGGGCAUUGG	AAGCUGCCAGUUGAAGAACUGU	CUCCGUUUGCCUGUUUCGCUG
>pxh-miR-263	>pxh-miR-652	>pxh-miR-1863
AAUGGCACUGGAAGAAUUCACGG	AAUGGCGCCACUAGGGUUGUG	AGCUCUGAUACCAUGUUAGAUU
>pxh-miR-7	>pxh-miR-7528	>pxh-miR-126a-3p
UGGAAGACUAGUGAUUUUAUUGUU	CCGAAAUGCUAAUCUGAAGCUU	UCGUACCGUGAGUAAUAAUGCG
>pxh-miR-8938	>pxh-miR-339	>pxh-miR-548a
UUGCGUUCUGGCGGCUGCUGCUUCC	UGAGCGCCUCGACGACAGAGCCG	AAAAGUAAUUGUGGUUUUUGC
>pxh-miR-166	>pxh-miR-196	>pxh-miR-10
UCGGACCAGGCUUCAUUCCCC	UAGGUAGUUUCAUGUUGUUGGG	UACCCUGUAGAUCCGAAUUUGU
>pxh-miR-1303	>pxh-miR-3630-3p	>pxh-miR-375
UUUAGAGACGGGGUCUUGCUCU	UUUGGGAAUCUCUCUGAUGCAC	UUUUGUUCGUUCGGCUCGCGUGA
>pxh-miR-1421w-3p	>pxh-miR-342-3p	>pxh-miR-81a
UGCAUAGUGGGUGAUCUUCUU	UCUCACACAGAAAUCGCACCCGUC	UGAGAUCAUUGUGAAAGCUAUU

Table A4.3 continued

>pxh-miR-390-5p	>pxh-miR-31a-5p	>pxh-miR-151-3p
AAGCUCAGGAGGGAUAGCGCC	AGGCAAGAUGCUGGCAUAGCUG	CUAGACUGAAGCUCCUUGAGG
>pxh-miR-98-5p	>pxh-miR-750	>pxh-let-7k-5p
UGAGGUAGUAAGUUGUAUUGU	CCAGAUCUAACUCUUCCAGCUC	UGAGGUAGUAGAUUGAAUAGUU
>pxh-miR-222	>pxh-miR-301	>pxh-miR-279
AGCUACAUCUGGCUACUGGGUCUC	CAGUGCAAUAGUAUUGUCAAAGCA	UGACUAGAUCCACACUCAUCC
>pxh-miR-7122a	>pxh-miR-106	>pxh-miR-9a
UUAUACAGAGAAAUCACGGUCG	UACCGCACUGUGGGUACUUGCUGCU	UCUUUGGUUAUCUAGCUGUAUGA
>pxh-miR-210-5p	>pxh-miR-767	>pxh-miR-126-3p
CUGUGCGUGUGACAGCGGCUGA	UGCACCAUGGUUGUCUGAGCA	UCGUACCGUGAGUAAUAAUGCG
>pxh-miR-169b	>pxh-miR-146a-5p	>pxh-miR-384-5p
UAGCCAAGGAUGACUUGCCUG	UGAGAACUGAAUUCCAUAGGC	UUGGCAUUCUGUCCACCUCC
>pxh-miR-1957a	>pxh-miR-315	>pxh-miR-473
CAGUGGUAGAGCAUAUGAC	UUUUGAUUGUUGCUCAGAAGGC	ACUCUCCCUCAAGGGCUUCGC
>pxh-miR-850	>pxh-miR-3470b	>pxh-miR-482
UAAGAUCCGGACUACAACAAAG	UCACUCUGUAGACCAGGCUGG	UCUUUCCUACUCCUCCCAUUCC
>pxh-miR-4334-3p	>pxh-miR-6222-5p	>pxh-bantam
UCCCUGUCCUCCAGGAGCUC	CCUGUUUGGAUCAGCCAAGGC	UGAGAUCAUUGUGAAAGCUGAUU
>pxh-miR-1511	>pxh-miR-1127	>pxh-miR-184-3p
AACCAGGCUCUGAUACCAUG	AACUACUCCCUCCGUCCGAUA	ACUGGACGGAGAACUGAUAAGGGC
>pxh-miR-2355-3p	>pxh-miR-133	>pxh-miR-872-5p
AUUGUCCUUGCUGUUUGGAGAU	UUGGUCCCCUUCAACCAGCUGU	AAGGUUACUUGUUAGUUCAGG
>pxh-miR-4492	>pxh-let-7i-5p	>pxh-miR-3661
GGGGCUGGGCGCGCGCC	UGAGGUAGUAGUUUGUGCUGUU	UGACCUGGGACUCGGACAGCUG
>pxh-miR-8523	>pxh-miR-1120b-3p	>pxh-miR-8590
GAAAGAUGGUUAUCGUUU	UUCUUAUAUUGUGGGACAGAG	AUUCCGAUUUGUAGAAAAAAAAU
>pxh-miR-164	>pxh-miR-1180	>pxh-miR-894
UGGAGAAGCAGGGCACGUGCA	UUUCCGGCUCGCGUGGGUGUG	CGUUUCACGUCGGGUUCACC
>pxh-miR-535	>pxh-miR-3963	>pxh-miR-317
UGACAACGAGAGAGAGCACGC	UGUAUCCCACUUCUGACAC	UGAACACAGCUGGUGGUAUCUCAGU
>pxh-miR-574	>pxh-miR-6752-3p	>pxh-miR-205
CACGCUCAUGCACACACCCACA	UCCCUGCCCCCAUACUCCCAG	UCCUUCAUUCCACCGGAGUCUG
>pxh-miR-5054	>pxh-miR-6087	>pxh-miR-169aa
UCCCCACGGUCGGCGCCA	UGAGGCGGGGGGGGCGAGC	UAGCCAAGGAUGACUUGCCUG
>pxh-miR-408-3p	>pxh-miR-628-3p	>pxh-let-7j-5p
AUGCACUGCCUCUUCCCUGGC	UCUAGUAAGAGUGGCAGUCGA	UGAGGUAGUAGGUUGUAUAGUU
>pxh-miR-1133	>pxh-miR-320-3p	>pxh-miR-26a
CAUAUACUCCCUCCGUCCGAAA	AAAAGCUGGGUUGAGAGGGCGA	UUCAAGUAAUCCAGGAUAGGCU
>pxh-miR-25	>pxh-miR-5293	>pxh-miR-22a
CAUUGCACUUGUCUCGGUCUGA	GAUGAAGAAGUGGAAGGAAGAAGA	AAGCUGCCAGUUGAAGAACUGU
>pxh-miR-502	>pxh-miR-99	>pxh-miR-27a-3p
AAUGCACCUGGGCAAGGAUUCA	AACCCGUAGAUCCGAUCUUGU	UUCACAGUGGCUAAGUUCCGC
>pxh-miR-530-5p	>pxh-miR-194	>pxh-let-7c
UGCAUUUGCACCUGCACCUA	UGUAACAGCAACUCCAUGUGGA	UGAGGUAGUAGGUUGUAUGGUU
>pxh-miR-4995	>pxh-miR-395a-3p	>pxh-miR-3535
AGGCAGUGGCUUGGUUAAGGG	UGAAGUGUUUGGGGGAACUC	UGGAUAUGAUGACUGAUUACCUGAGA
>pxh-miR-276	>pxh-miR-191a	
UAGGAACUUCAUACCAUGCUC	CAACGGAAUCCCAAAAGCAGCUG	
>pxh-miR-399	>pxh-miR-467f	
UGCCAAAGGAGAGUUGCCCUA	AUAUACACACACACACCUACA	

Table A4.4. Normalized read counts for miRNA found in all three conditions.

miRNA	Trichome 18°C	Trichome 23°C	Bald Pedicle 23°C	TOTAL
pxh-miR-166	75254.24	138785.63	144453.31	358493.18
pxh-miR-319	53559.32	89219.33	143655.23	286433.88
pxh-miR-1128	11525.42	29739.78	16759.78	58024.98
pxh-miR-396	10169.49	23543.99	22346.37	56059.85
pxh-miR-390-5p	12881.36	22304.83	19154.03	54340.22
pxh-miR-393	10847.46	17348.20	12769.35	40965.01
pxh-miR-21a-5p	16271.19	7434.94	15961.69	39667.82
pxh-miR-21	16271.19	7434.94	15961.69	39667.82
pxh-miR-21a-5p	14915.25	7434.94	16759.78	39109.98
pxh-miR-21	14915.25	7434.94	16759.78	39109.98
pxh-miR-894	12203.39	19826.52	6384.68	38414.58
pxh-miR-168	8813.56	9913.26	15163.61	33890.43
pxh-miR-159	9491.53	12391.57	11971.27	33854.37
pxh-miR-165a-5p	6779.66	14869.89	11971.27	33620.82
pxh-miR-827	8813.56	19826.52	4788.51	33428.58
pxh-miR-5083	10169.49	19826.52	2394.25	32390.26
pxh-let-7f	11525.42	8674.10	10375.10	30574.63
pxh-miR-403-3p	4745.76	13630.73	11971.27	30347.76
pxh-let-7	12881.36	6195.79	10375.10	29452.24
pxh-let-7a	12881.36	6195.79	10375.10	29452.24
pxh-let-7j-5p	12881.36	6195.79	10375.10	29452.24
pxh-miR-3630-3p	8813.56	17348.20	2394.25	28556.02
pxh-miR-1448	4745.76	9913.26	12769.35	27428.38
pxh-miR-10a-5p	6101.69	3717.47	16759.78	26578.94
pxh-miR-10	7457.63	1239.16	17557.86	26254.65
pxh-miR-398	11525.42	7434.94	7182.76	26143.13
pxh-miR-397-5p	7457.63	11152.42	7182.76	25792.80
pxh-let-7	11525.42	4956.63	7980.85	24462.90
pxh-let-7a	11525.42	4956.63	7980.85	24462.90
pxh-miR-182	9491.53	4956.63	9577.02	24025.17
pxh-miR-156b	8135.59	8674.10	7182.76	23992.46
pxh-let-7	11525.42	4956.63	7182.76	23664.81
pxh-let-7j-5p	11525.42	4956.63	7182.76	23664.81
pxh-miR-156b	8135.59	8674.10	6384.68	23194.37
pxh-miR-156	8135.59	8674.10	6384.68	23194.37
pxh-miR-92a-3p	9491.53	6195.79	7182.76	22870.07
pxh-miR-5049c	4067.80	13630.73	4788.51	22487.04
pxh-miR-26	9491.53	4956.63	7980.85	22429.00
pxh-miR-26a	9491.53	4956.63	7980.85	22429.00
pxh-miR-26	9491.53	4956.63	7980.85	22429.00
pxh-miR-26a	9491.53	4956.63	7980.85	22429.00

Table A4.4 *continued*

miRNA	Trichome 18°C	Trichome 23°C	Bald Pedicle 23°C	TOTAL
pxh-miR-1133	2711.86	13630.73	5586.59	21929.19
pxh-miR-6478	8135.59	11152.42	2394.25	21682.26
pxh-miR-191-5p	10847.46	3717.47	4788.51	19353.44
pxh-miR-191-5p	10847.46	3717.47	4788.51	19353.44
pxh-miR-191a	10847.46	3717.47	4788.51	19353.44
pxh-miR-27-3p	7457.63	7434.94	3990.42	18882.99
pxh-miR-2916	4067.80	9913.26	4788.51	18769.56
pxh-miR-30a-5p	6101.69	6195.79	6384.68	18682.16
pxh-let-7i-5p	8135.59	2478.31	6384.68	16998.58
pxh-miR-5168-3p	1355.93	4956.63	10375.10	16687.66
pxh-miR-408-3p	6779.66	7434.94	2394.25	16608.86
pxh-miR-235-3p	6779.66	4956.63	4788.51	16524.80
pxh-miR-92-3p	6779.66	4956.63	4788.51	16524.80
pxh-miR-8175	7457.63	7434.94	1596.17	16488.74
pxh-miR-181a	9491.53	3717.47	3192.34	16401.34
pxh-miR-535	1355.93	4956.63	9577.02	15889.58
pxh-miR-2478	5423.73	8674.10	798.08	14895.92
pxh-miR-7122a	2033.90	8674.10	3192.34	13900.34
pxh-miR-27a-3p	6101.69	2478.31	3990.42	12570.43
pxh-miR-858	2711.86	6195.79	3192.34	12099.99
pxh-miR-320-3p	3389.83	4956.63	3192.34	11538.80
pxh-miR-151-3p	4067.80	2478.31	4788.51	11334.62
pxh-miR-103-3p	5423.73	2478.31	3192.34	11094.38
pxh-miR-107	5423.73	2478.31	3192.34	11094.38
pxh-miR-22	7457.63	1239.16	1596.17	10292.95
pxh-miR-22a	7457.63	1239.16	1596.17	10292.95
pxh-miR-22	7457.63	1239.16	1596.17	10292.95
pxh-miR-22a	7457.63	1239.16	1596.17	10292.95
pxh-miR-477a	2711.86	2478.31	4788.51	9978.69
pxh-miR-5368	5423.73	3717.47	798.08	9939.29
pxh-let-7e	3389.83	2478.31	3990.42	9858.57
pxh-let-7d-5p	5423.73	1239.16	3192.34	9855.22
pxh-miR-167-5p	2711.86	6195.79	798.08	9705.74
pxh-miR-171	2033.90	4956.63	2394.25	9384.78
pxh-miR-1910-5p	2033.90	4956.63	2394.25	9384.78
pxh-miR-221-3p	2711.86	4956.63	1596.17	9264.66
pxh-let-7b-5p	4067.80	2478.31	2394.25	8940.37
pxh-miR-172	2033.90	2478.31	3990.42	8502.64
pxh-miR-16a-5p	4745.76	1239.16	2394.25	8379.17
pxh-miR-16	4745.76	1239.16	2394.25	8379.17
pxh-miR-16a-5p	4745.76	1239.16	2394.25	8379.17
pxh-miR-16	4745.76	1239.16	2394.25	8379.17

Table A4.4 *continued*

miRNA	Trichome 18°C	Trichome 23°C	Bald Pedicle 23°C	TOTAL
pxh-miR-423-3p	4067.80	2478.31	1596.17	8142.28
pxh-miR-141	2033.90	1239.16	4788.51	8061.56
pxh-miR-170-5p	2033.90	4956.63	798.08	7788.61
pxh-miR-93	4067.80	1239.16	2394.25	7701.21
pxh-miR-375	2711.86	2478.31	2394.25	7584.43
pxh-miR-125	3389.83	2478.31	1596.17	7464.31
pxh-miR-157a-5p	2033.90	3717.47	1596.17	7347.54
pxh-miR-222	2711.86	3717.47	798.08	7227.42
pxh-miR-5072	1355.93	4956.63	798.08	7110.65
pxh-miR-5181-3p	1355.93	4956.63	798.08	7110.65
pxh-miR-160	1355.93	2478.31	3192.34	7026.59
pxh-miR-98-5p	4067.80	1239.16	1596.17	6903.12
pxh-miR-23a-3p	2711.86	2478.31	1596.17	6786.35
pxh-miR-8109	2033.90	3717.47	798.08	6549.46
pxh-miR-25	2711.86	1239.16	2394.25	6345.28
pxh-miR-24	3389.83	1239.16	1596.17	6225.16
pxh-miR-24a-3p	3389.83	1239.16	1596.17	6225.16
pxh-miR-24	3389.83	1239.16	1596.17	6225.16
pxh-miR-24a-3p	3389.83	1239.16	1596.17	6225.16
pxh-miR-100	4067.80	1239.16	798.08	6105.04
pxh-miR-164	1355.93	1239.16	3192.34	5787.43
pxh-miR-425-5p	2033.90	1239.16	2394.25	5667.31
pxh-miR-941	1355.93	2478.31	1596.17	5430.42
pxh-bantam	2033.90	2478.31	798.08	5310.30
pxh-miR-146-5p	2033.90	1239.16	1596.17	4869.22
pxh-miR-28-3p	2711.86	1239.16	798.08	4749.11
pxh-miR-183-5p	1355.93	1239.16	1596.17	4191.26
pxh-miR-301a-3p	1355.93	1239.16	1596.17	4191.26
pxh-miR-301	1355.93	1239.16	1596.17	4191.26
pxh-miR-301a-3p	1355.93	1239.16	1596.17	4191.26
pxh-miR-301	1355.93	1239.16	1596.17	4191.26
pxh-miR-101	2033.90	1239.16	798.08	4071.14
pxh-miR-1507-3p	677.97	2478.31	798.08	3954.37
pxh-miR-205	677.97	2478.31	798.08	3954.37
pxh-miR-584	677.97	2478.31	798.08	3954.37
pxh-miR-482	677.97	1239.16	1596.17	3513.29
pxh-miR-424-3p	1355.93	1239.16	798.08	3393.17
pxh-miR-486	1355.93	1239.16	798.08	3393.17
pxh-miR-81a	1355.93	1239.16	798.08	3393.17
pxh-miR-1421w-3p	677.97	1239.16	798.08	2715.21
pxh-miR-395	677.97	1239.16	798.08	2715.21
pxh-miR-395a-3p	677.97	1239.16	798.08	2715.21
pxh-miR-395	677.97	1239.16	798.08	2715.21
pxh-miR-395a-3p	677.97	1239.16	798.08	2715.21
pxh-miR-4995	677.97	1239.16	798.08	2715.21
pxh-miR-7528	677.97	1239.16	798.08	2715.21
pxh-miR-872-5p	677.97	1239.16	798.08	2715.21

APPENDIX 5

Table A5.1. Primer Sequences for *E. coli* and Tobacco Assay

End	Enzyme Site	Primer Sequence	Description
5'	NdeI	GCA TCG ACA TAT GTC GGC CAA ACC AGA GAC TGT G	ACP 1 in pet22b
3'	XhoI	CAT GTC TCG AGA GCA GGA GCC TTC TTC TC	
5'	NdeI	GCA TGA TCA TAT GTC GGC CAA AGC AGA GAC TGT G	ACP 2 in pet22b
3'	XhoI	ACG ATC TCG AGA GCA TCC TTC TTC TCA AC	
5'	SpeI	AGG GCC ACT AGT GAA GGA GAT TTC TAT GTC GGC CAA ACC	ACP 1 in pet3d-MAD
3'	NcoI	CAA AAT CCA TGG TCT CCA CTC CTT CTA ATT TAA GCA GG	
5'	XbaI	CAA GGG TCT AGA CGG AAG GAG GTT CTC ATG TCG GCC AAA GC	ACP 2 in pet3d-MAD
3'	NcoI	TGT CTC CCA TGG TTG CCT CTC CTT CCT ATT TTA AGC ATC C	
5'	NdeI	CAG ATC TCT CAT ATG GCT TCG TTC AC	ACP 1 in pRI201AN-MAD
3'	SalI	TCC ATC CGT CGA CAT TTA AGC AGG AGC CTT	
5'	NdeI	CCA ATC TAC ATA TGG CTT CCT TTA C	ACP 2 in pRI201AN-MAD
3'	SalI	CCT CTC CGT CGA CTT TTA AGC ATC CTT CTT	
5'	N/A	GCC ACG ACG GGC GTT CCT T	RT-PCR pRI201-AN Kanamycin
3'	N/A	GAG TAC GTG CTC GCT CGA T	RT-PCR pRI201-AN Kanamycin
5'	N/A	TGA GAG CCC GTG GGC AGT TT	RT-PCR pRI201-AN

			MAD
3'	N/A	GGC GAT GGT ACC GCA TAT TT	RT-PCR pRI201-AN MAD
5'	N/A	TAG CTC TTC CAG AGG GAA	RT-PCR pRI201-AN ACP 1
3'	N/A	CTT TCC TCT TCC ACG TTT	RT-PCR pRI201-AN ACP 1
5'	N/A	GGC TAT TCC AAC TGA TAC	RT-PCR pRI201-AN ACP 2
3'	N/A	CTT TCC TCT TCG ACA CTA	RT-PCR pRI201-AN ACP 2

Table A6.1. Sequences of fatty acid genes of *Pelargonium × hortorum.*

Gene designation	Complete nucleotide sequence (5' to 3')
*Pxh*ACP 1	cctcgtgccgcacacctttgtgatctccgccccaaatctcttctcttgcgcctccttctccaccaatatctcagatctct acaatggcttcgttcacagctaattctctttccctcacctccatctcctgctctttcacacctatcaaggcacctgccag gacctccagtcttaaatctgtttcattctccatcaacgggaatggcttttcatctcttaggttacgacaagggccatctc gcttccagatttcttgttcggccaaaccagagactgtggacaaggtgtgtgaaattgtgaagaagcaattagctctt ccagagggaactgaagtctcaggagattcaaagtttgctgcacttggagctgattctctcgacacggttgagattgt gatgggacttgaggaggaattcgggataaacgtggaagaggaaagtgctcagaacattgccaccgttcaagat gctgcagatctgattgagaagctggtggagaagaaggctcctgcttaaattagcatggatggagaagaaggattt tgattaaatttgcctagtggacttgcttataatttgtttttgttgttccctacatgtttaactggaacaaccctgccctgtgga tttttagcttagc
*Pxh*ACP 2	ctttgtgctctccgccccccatctctctctcctcctcttccccaccatctcagatttctctcctctcctctcctctctgctccaat ccaatctatcaatggcttcctttactcccaattctgtttctatgacctccatctcctgttccttgaggccgaacatggccc ctaccatgatctctggtatgaaatcagcatccttctccattaacaggaatggctttccatctcttaggttacaacaagg gtcatctcgcctccaggttctctgttcggccaaagcagagactgtggacaaggtgtgtgaaatagtgaggaagca attggctattccaactgatactgaggtctcaggagagtcaaagtttgctgcacttggagctgattctcttgacacggtt gagatagtgatgggacttgaggaagaatttgggattagtgtcgaagaggaaagtgctcagagtattgccaccgtt caagatgctgctgatctgattgagaagctcgttgagaagaaggatgcttaaaatagtttggagaggcaactgtcc gagacagtatactactatgaactttgattagtatctatattacatgcacaattcacctgtttctattggaggcatggcatt gcggcattctcttgttgtttctttatatacctagacgaatttctgtgccatgtttaatttaatttttatggtgagttgggatttag cacaaaagaaaaaaaaaaaaaaaaaaaaaaaaaggggg
*Pxh*FAT-A1	cccctttgatgtctggaaagaaagtttcgattttgatccaaagagtcacacccaaccacgtaatttcagaccaaacc cagtggcatttttgcctcctctgacccttttatcgtcttcctcggtcccatcgagtccctcctcgtcaattcccattcctctat aacaccaccaaaacccatgtgaattgtgatgtctctcgcagccatttcagctctcactcttgcttgacctcgccgaag aacaaccactctattctcttctctctctcgcaaccagaaatgttgaagctttcttgcaatgccacggactgtcagattc aagccctagctcaatgcagatctattgttaggcctcgccgccggaacggcgtctttgctgctcgccggtttctcggg cggctccaattgtgtctgtggtgtcggaccggacggtggaagttgttggcggcggcggttcggggagtttggcgga ccggctcagattggggagcttgaccgataatgggttgtcgtacacggaaaagtttattgtgaggtgttacgaggtcg ggattaacaagactgccactgtggagaccattgccaatttgcttcaggaagttggatgtaaccatgctcagagcgtt ggattttcaacagatgggtttgcaacaaccctcacgatgagaaagttgcatctcatatgggttactgctcgcatgca cattgaaatatacaaataccctgcttggagtgacgtgattgaaatagagacatggtgtcaaagtgaaggaagaat cggaactagacgtgattggattctgaaggactatggtactggtcaagttattgggagggctacaagcaagtgggt gatgatgaaccaagatactaggcgacttcagaaagttaatgatgatgtcagagatgagtatttggttttctgtccac gagaaccaagattagcatttccagagaagaacaatagcagcttaaggaaaaatatcaaagctcgaggatcctgc tcagtattccaggctgggacttatgcctagaagagctgatctggacatgaaccagcatgttaacaatgttgcatata ttggatgggttctggagagcatgcccaagaaatcattgacactcatgaactacaaaccatcaccttagattacag gcgggaatgccaacaagacgacatagtagattccctgactggcgtcgaacaaggcgagggcagtaaagcgct ttccaatcttcaaggagcaaacgggtatgctggagctgcaccagataagaaaaaagaccgccttcagttttttgcat ctattgagattggcaggcgacgggcttgaaataaacaggggggcgcactgagtggagaaaaaagccagctaga taagaaaaaaggcgacgaaagcgtagttgtttcttccaacgtttcttgcttgcttgtagttttgagagattttcttcgtttat ttttccttagaaagggtttttgctctgtattgatgtgcgtagtaggagctagattgtccatttatttaaaagcctctttgatcatc tgttgtgaaacttgaagtcactggtctttggaagattgtacttgaaagtaccaatccctcaaactcagatcactcatttg ttgtactattaattttctctttaagctaaatttgcgctgcagccgcgggggggatcca
*Pxh*FAT-A2	gcgccatgaatgtttttcctatttaagctcgaaaaccactctccatttctctcccaaactcactctcacttcaaatcgtcc

	atctccaaaaccctaaatgttgaacctttcgtgcaatgctactcaatctctagccctaaatgtctcttattcctctcgccg acgccgacgaaacggtgtcgttttttgctctccggtttctcgggccgggccggttctgtcggtcgagtcggaccgggt gagatcggggagaatggcggaggatgggttgtcgtttacggagaagtttgtcgtgaggagctatgaggtcggaat taacaaaactgccactgttgagaccattgccagtttgttgcaggaaactggatgtaaccacgtacaaagcacggg actttcaaccgatgggttcgggacaacccccatcatgaggaaactgcatctcatatgggtaacttctcgcatgcac atcgaaatatacaaatacccagcttggagtgatgtgattgaaatagagacgtgggcccaaggtgaaggaagaa ttggatgtagacgcgattggattatgaaagagtatggttctggtcaagttatcgggagagcttcaagcaagtgggtg aggatgaaccaagataccaggcgacttcagaaaattaatgatgatatcagggacgaggtttcggttttctctccaa gagagttaagattagcatttccagaggcgaacaatagcagtttaaggaaaataccgaagctagaagatcccgct caatattctagactgggacttatgcctagaagagctgatctggacatgaaccagcacgttaacaatgtcgcctatat tggatgggttctagagggcttgcctcaagaagtcatcgacacccacgaattacaaaccatcaccttagattacag acgggaatgcaaacaggatgacatggtcgattccctcaccagccccgaactagccaagaataatggtacaac agtttcagctcttctcggatccaatgggtctgcttcggctgcaagagataagaaacaaaaccgtcttcagttttttgca cttactgagacaatcaggcgatgggcttgaaataaacagagggcgtaccgtgtggagaagaaaacccgcaag ataacgctctcgaacctcgttagtttgtattttttatgagtaacatgtgttttctatacatattacgttctccattgtatctacttc gatcacgtgtcaggttgagaagtgagtcaagtcgttgtgcgttacgaggatatgtgggagaagtgacagaggctc gattgaataactgtcagttccttgtattcaaactccgtatggatagttcaaatgttgtattttagtttttgccctctaggcac acataggtggcgtttgattgcatggatatgaaagtggattggatttcgaaatctattatattgtctaatccaatgtttgat gacaacgattgattggatttcgaaatccattatgtggttttttatccatcatatgatggataaaataatccaagggtacc ctcttgtattacgatggattggatatagaatcagatcttattcaagggaccagtgc
*Pxh*FAT-A3	aagggggggtttctgttttccccagggaaagggggtggcacaggtttaatggtaccccggtttccagccgacttgtaaa ggcgcccagaatttatagacccattgggggaaatgggtcacgggcccctatgttgacctttcgtgcatgctactcaat ctctagccctaaatgtctcttattcctctcgccgacgccgacgaaacggtgtcgttttttttgctctccggtttctcgggcc gggccggttctgtcggtcgagtcggaccgggtgagatcggggagaatggcggaggatgggttgtcgtttacgga gaagtttgtcgtgaggagctatgaggtcggaattaacaaaactgccactgttgagaccattgccagtttgttgcagg aaactggatgtaaccacgtacaaagcacgggactttcaaccgatgggttcgggacaacccccatcatgaggaa actgcatctcatatgggtaacttctcgcatgcacatcgaaatatacaaatacccagcttggagtgatgtgattgaaat agagacgtgggcccaaggtgaaggaagaattggatgtagacgcgattggattatgaaagagtatggttctggtc aagttatcgggagagcttcaagcaagtgggtgaggatgaaccaagataccaggcgacttcagaaaattaatga tgatatcagggacgaggtttcggttttctctccaagagagttaagattagcatttccagaggcgaacaatagcagttt aaggaaaataccgaagctagaagatcccgctcaatattctagactgggacttatgcctagaagagctgatctgg acatgaaccagcacgttaacaatgtcgcctatattggatgggttctagagggcttgcctcaagaagtcatcgacac ccacgaattacaaaccatcaccttagattacagacgggaatgcaaacaggatgacatggtcgattccctcacca gccccgaactagccaagaataatggtacaacagtttcagctcttctcggatccaatgggtctgcttcggctgcaag agataagaaacaaaaccgtcttcagttttttgcacttactgagacaatcaggcgatgggcttgaaataaacagagg gcgtaccgtgtggagaagaaaacccgcaagataacgctctcgaacctcgttagtttgtattttatgagtaacatgt gttttctatacatattacgttctccattgtatctacttcgatcacgtgtcaggttgagaagtgagtcaagtcgttgtgcgtt acgaggatatgtgggagaagtgacagaggctcgattgaataactgtcagttccttgtattcaaactccgtatggata gttcaaatgttgtattttagtttttgccctctaggcacacataggtggcgtttgattgcatggatatgaaagtggattgga tttcgaaatctattatattgtctaatccaatgtttgatgacaacgattgattggatttcgaaatccattatgtggtttttatcc atcatatgatggataaaataatccaagggtaccctcttgtattacgatggattggatc
*Pxh*FAT-B1	ggtaatattccttcttcttctaaaaaatataggtattgttatatcatggcagattgtcagctgcagtgggattattgggtcc attgtctaatgataaacaattcatcaagttatttgggggaataagtcgcaaaaacttatatactaggcggcgtctagt ttgtacaaaagttaagattcatggtgatctaacagagaagaagcagagtgagataaaagtgttggcaagaaag attgatccatttaaaggaaagttggttcagggtgggattgtttcaaccaatacttctcaattagatcataccaggttga caacgaggccaaagcttccatttgggccataatgaatattttacaggactcagcacttaaccacagtaggactac gggactattggctaatggttttgggtcatcaccagagatgaccaaaaggaagttggtatgggttgttcacacaaca catattggtggacaactatccttcatggggcgatgttgttcaggtagacaaatggttttgtgcatcacaaaagagc agtactttgcgttctgattggcttctctctgatcctaacacaggaaaacctttggttcgagcaagcactaagttagtgat gatgaacacggagacaagaaagtttccaagttaataaaagaagtcagagaggagttacagcctttatctcagg gactgtgatcctatcataaatgccaagaaacgaattcaacaacatgatgatcaggtggacgttcacacaacagat aattacatscgaacgggtttaattcctggacgggatgatttggatgtgaataaccatgtcaacaatgccaaatacat tgcatggattctagagagtgctccgtgctcgattatggagagctatcagctttcacgtgtgacattggaatacaaaa aggagtgtggtatagatagtgtggtgcagtctgtaaccagagttgttggaaacgaaagcaatgacaatggcgaa gagattgaaatgcaacacttgcttcgtctagagacgggccaagaagttgcaaggggaagaaccacatggaag

	cttaaacgacaaaaagatgttataagtttctttctgccacagaatcgaggatgatggagacttgaacctaagacct cacacgaaaaaacttaaacgacaaaaaggtgttaaaagtttctttccgccacagaatctgacatggtcgtgcaaa cattcgaagatgatgaagccccaagttatagaatacacgtgaaattttttgctagcaatgaactcaattgtcaaatca gacaaatcttagtgatgatttgttacttatttcaataaaaagcaggtaaaaacgacaataattaaatcacaaagg gg
*Pxh*FAT-B2	aaatgtacattgtctttcccaaagttaaatcaaattctcacctggaatatatcatcctagacagctaaggtaaggtaa aggtttgtaagaagaagaagaaatcaactcaataaagatattagagggagagatggagatggggacgaaga gtatacagcaagggcttaattacaatgggtttataacttgtcatctttctcctgaggaagacaaattagaaggaatat acttatattataaataaataaataaaagaagaaacatatataataaaattcaagaaagcaagtccgattttttctaca caagcatgtgacagcaacgagactttatgccagctgagaaagagcttacggatattaagcactttctgctggaatc tgacccatattgccaaggttagctttttgggcctccactgagttcttcccctcacaatttcagcaccttcctcgagtcga agcaggtgctggcactcaacaccaactggactccctatatagccaatacctgctgtgtcagagtcagacacggc agtcagggactgcagcacactgtccttcccacactccctcctatactccaaagtcattgaacaaagttcgtgactct ccaagataggcaatggtgcactctgcgagtatcagcattccggttaacaaaagatatgatatagcgacgcaaga tatcatattaccacaagtactcaatatacaaaaacggatgttaacaaatacaattcacaagttcagatggttaagtg ctgtaaagttccagtccagtccaatggttataaaattagaaggaaataaatgggaccatgcaggtcatagacaga tatgcacataaacaaactgcagagtataagttgaaacctttagtgcataactaaaataacagttgaactggaaat ataagtgaagttttcaaattgaccccagactattagaagatagcactagaaagtgagtattaaagaacaacttgc aacatgtacatgcatctttatcctgtttaaaacttttacttgtcaaagtaacatcatcttagtacaccacaactacaactt agttataaatgaaaccatattgaatgtagatttgttatccacccacttttttttataggagaatagaagattcgcatatta ggagctaatctaaaaaataagagcagttcactccagtcgaaatattgatccggtcaaagcaataaactgaaatgt gatccgagaagaaggaattatgaggccagaaaatttacctcaagaatccacccaatgtacttgacattgttaaca tgctgattgacatccaaatcactccatcgaggctgcaaataaaatttatatatcttagtggtcgcaaaaaaggtgtg cagtagtttttttttccaacacttgagagcgttgacttacagataatccttgcggacataatctgcagtgttgtcgtcaag tttaggtagttttctgctatcctcatccacaacaggggcagcatccaaaaaatgaggctctatttcccctcgaacttctt ctggcattttggataatctccttgtctctttattcatcattacccacacactgcacccacaaaaccatgaagaacacg ataagttcttctgataaaacaaaacagttgaagaataaaagacctaaagccataaagcatattttgaaggaaata aatctgtctgcttaaaaactcataaaagaaataaaatttaaacaaataaaaaactcaaccattgcgatacattgct cctacaattatggaaaaagcactaaaaaatgagaattcagtttgtgtaaaatcagatgagctagtttcattccttcca tacgacacagacgctaatgcgaaatttattttagcaacaaaataggttctatttaatcattaaaacataatttctacct ggaggctcttgttaaagtttcaccagtttacaatcacgaagaagccaatcacggcgcataccattctttccagatg cactaacccaagtgtctacttgaacaacatcacccctgaaacacagataattcagctcattaatctcaaatgattag tttacaacactctaaacaaactcacaactaaagtggttcttgcaaacaagtgaggccataataagtacttattcaat aatttaaataaaggtaaacctaacacagataacatcagctgtattgtcaggtcaatagaagggagaaatataca cattaaagtttgtggcttaccaagttgggtaacgatctactaaaacctgcatcctagtaacaacccatattaagttcttt ttgtacatttctggggttgcaccaaagccatcaccaagaagcccagcagtcttaacatggttaagagccgtttcctg cagccaaaaataacaacattcaagtaagttagttaagaaaagacatattttattaatatcgttgaaggagcacca aacaaaactttgcatgctaataaaatctcatgggcacctacaaagaaagagaggacattaaataacttatcatac ctgcaaatggttcattaacgtctctatagatgcggtccgatcagcgcctatctcataagacctaattgagaagttttgt cggaaaacaagaccatcctgaacaattcttcctataccaaaaggatccacaagcatgtcaggtcgctttgtttcca ttcaagcatcatccactgcttctctgcagccaagaatattgtggtgatagcagcaagaagcatgctccaatcaggc aattggttaataaatgtcctggggggcatgagttgttggcatgtcatcttcacccctcacgccatcaattgttgccaccc caacttttgtaccattaatcttaggcggagcttgggcacttgtcttgacctgcaacttagctggttttgatttaattcctcc aaaatttgcagacccaccaccaagtttggtggcctttgaaccagaatctgagggcgtggaagctacaggaaag aatgatgaggcggcagtagcagcaaccatgatgatattgaactgtcaagtcagtcaggctgccaactgaacag ctgtatccgtaaaatgctctaccttgaacagcaacagagaagaaagtagtgaaaacagctcaacaaatgcga gaatatcgtttgttctttttttccagcaactgcttccagtcagagggacagtatcgtcaatttggatcccagaaaccct attttccggtaagtgctagggtttgcgatctcctccaagcaaaattctcgttccaactttccaaacagaaattaaataa aattaatggtttaattaaatagaaaaaaagagggacgcaacagagagaggagaagagagtaattagggcca gcggtttgagtg
*Pxh*FAT-B3	aaatgtacattgtctttcccaaagttaaatcaaattctcacctggaatatatcatcctagacagctaaggtaaggtaa aggtttgtaagaagaagaagaaatcaactcaataaagatattagagggagagatggagatggggacgaaga gtatacagcaagggcttaattacaatgggtttataacttgtcatctttctcctgaggaagacaaattagaaggaatat acttatattataaataaataaataaaagaagaaacatatataataaaattcaagaaagcaagtccgattttttctaca caagcatgtgacagcaacgagactttatgccagctgagaaagagcttacggatattaagcactttctgctggaatc

	tgacccatattgccaaggttagcttttttgggcctccactgagttcttcccctcacaatttcagcaccttcctcgagtcgaagcaggtgctggcactcaacaccaactggactccctatatagccaatacctgctgtgtcagagtcagacacggcagtcagggactgcagcacactgtccttcccacactccctcctatactccaaagtcattgaacaaagttcgtgactctccaagataggcaatggtgcactctgcgagtatcagcattccggttaacaaaagatatgatatagcgacgcaagatatcatattaccacaagtactcaatatacaaaaacggatgttaacaaatacaattcacaagttcagatggttaagtgctgtaaagttccagtccagtccaatggttataaaattagaaggaaataaatgggaccatgcaggtcatagacagatatgcacataaacaaactgcagagtataagttgaaaccttagtgcataactaaaataacagttgaactggaaatataagtgaagttttcaaattgaccccagactattagaagatagcactagaaagtgagtattaaagaacaacttgcaacatgtacatgcatctttatcctgtttaaaacttttacttgtcaaagtaacatcatcttagtacaccacaactacaacttagttataaatgaaaccatattgaatgtagatttgttatccacccacttttttataggagaatagaagattcgcatattaggagctaatctaaaaaataagagcagttcactccagtcgaaatattgatccggtcaaagcaataaactgaaatgtgatccgagaagaaggaattatgaggccagaaaatttacctcaagaatccacccaatgtacttgacattgttaacatgctgattgacatccaaatcactccatcgaggagataatcctttgcggacataatctgcagtgttgtcgtcaagtttaggtagttttctgctatcctcatccacaacaggggcagcatccaaaaaatgaggctctatttcccctcgaacttcttctggcattttggataatctccttgtctctttattcatcattacccacacactggaggctcttgttaaagtttcaccagttttacaatcacgaagaagccaatcacggcgcataccattctttccagatgcactaacccaagtgtctacttgaacaacatcaccccaagttgggtaacgatctactaaaacctgcatcctagtaacaacccatattaagttctttttgtacatttctgggttgcaccaaagccatcaccaagaagcccagcagtcttaacatggttaagagccgtttcctgcaaatggttcattaacgtctctatagatgcggtccgatcagcgcctatctcataagacctaattgagaagtttgtcggaaaacaagaccatcctgaacaattcttcctataccaaaaggatccacaagcatgtcaggtcgctttgtttccattcaagcatcatccactgcttctctgcagccaagaatattgtggtgatagcagcaagaagcatgctccaatcaggcaattggttaataaatgtcctggggggcatgagttgttggcatgtcatcttcacccctcacgccatcaattgttgccaccccaactttgtaccattaatcttaggcggagcttgggcacttgtcttgacctgcaacttagctggttttgatttaattcctccaaaatttgcagacccaccaccaagtttggtggcctttgaaccagaatctgagggcgtggaagctacaggaaagaatgatgaggcggcagtagcagcaaccatgatgatattgaactgtcaagtcagtcaggctgccaactgaacagctgtatccgtaaaatgctctaccttgaacagcaacagagaagaaagtagtgaaaacagctcaacaaatgcgagaatatcgtttgttctttttttccagcaactgcttccagtcagagggacagtatcgtcaatttggatcccagaaaccccctattttccggtaagtgctagggtttgcgatctcctccaagcaaaattctcgttccaactttccaaacagaaattaaataaaattaatggtttaattaaaatagaaaaaaagagggacgcaacagagagaggagaagagagtaattagggccagcggtttgagtg
*Pxh*KAS Ia	ggtcgtaatttttttaaattgactcaccgagttcactctatatttatccgagtcatgcaatcaccaactctctggccatctccgatcgattcgcttcgcaaatcgccgccaaatgcggcgattctccggaaaatccggaggcctttcttctccgcccgagccgcctctaccgccgtctcggccccgaagcgcgagaccgatccgaagaagcgcgtggtgatcaccggcatgggcctcgtctccgtcttcggcaccgacgtcgacacctactacgaaaagctcctcgccggcgagagcgggatcagcttgatcgaccgattcgacgcatcaaaatttccgaccagattcggcggccagattcgcgggtttagcgccggagctacatcgacggcgaaaacgaccggcgactcgacgattctctccggtactgcctcgttgccgggaagagagcgattgaagatgccgatcctgccggagatttgctctccaagattgataaagagagagagctggtgtgcttgttggaatgggattgggtggtcccacagtgctttctggtggagttcagaaaatgatagagaagggtcacaaacaagtatctcccttttttgttccatttactttaacaaacatgggctctgcattgcttggtattgaacttgatttcagaggtcctaattattcaatctctgcggcttgtgctacatccaattttttgcttctactctgctgccaatcacatccgccgcggtgaggctgatttgatgattgctggagggactgaggcgatgatagttccgtttggattgggaggctttgctgcgtgtagggcactgtctcagagaaatgatgatcctcggacggcttcaaggccgtgggacaaagatcgagatgggtttgttatgggtgaaggtgctggagtattggtaatggaaagcttggaacatgcaatgaagcgaggtgcaccaattgttgctgagtatttgggaggtgcggtgactgtgatgctcatcatatgaccgatccaagggccgatgggcttggcgtctctaactgcattaaactttgtcttgcaaatgctggcgtgtcacctgaagaggttaattacataaatgcacatgctacttccactcttgccggtgatctagctgaggtaaatgctatcaaaaaggtattcaagaacacatccgggatcaagatcaacgcaacaaagtctatgatcgggcatagtctcggtgctgctgggggtttggaagccattgccacagtgaaagctataaacacaggctggcttcatcctaccataaatcaatttaatctggagcctgaagtcgaattcgacactgttgcaaatacgaagcaacaacacgaagtaaatgttgccatttcaaactcgttcggatttggcggacacaactctgtcgtggcgttttctgctttcaagccctaatcacgagaaggtttgagttcgttccatcctttcatggtcactgcatatctttcttgaagcccccattgtactttactgtataataatcgcgtcatgttggccaagtggcttaggcgcttatacctttagaatcattgtcatcaattgaactatcggtacatttagatcttattcaatatcgagtttttgtagcgtttaggaatcatcgtagggcgaagctttgcacgccacgagattgagttgttgagtttacacttagtttcagggcagttttgaaatagaagatgaaagctatgcgccctattgtgtaagcctatgtttggacccgaaatagaagggtcttttgtgtcaaaattgttttgtgggttatagtcttttctgatatgttgtaacattattattgtacttgtaacttgtaagctattagaaattttatatatggtagaacaaaaaa
*Pxh*KAS Ib	gctcaccgagttcactctctatttatccgagtcatgcaatcaccaactctccggccatctccgatcgattctccggaa

	aatccggaggcctttcttctccgtccgagccgcctccaccgccgtctcggcccccgaagcgcgagaccgatccga agaagcgcgtggtgatcaccggcatgggcctcgtctccgtcttcggcaccgacgtcgacacctactacgaaaag ctcctcgccggcgagagcgggatcagcttgatcgaccgattcgacgcatcaaaatttccgaccagattcggcgg ccagattcgcgggtttagcgccggaggctacatcgacggcaaaaacgaccggcgactcgacgattctctccggt actgcctcgtcgccgggaagagagcgattgaagatgccgatcctgccggtgatttgctctccaagattgataaag agagagctggtgtgcttgttggaatgggatttggtggtcccacggtgttttctggtggagttcagaaactgatagaga agggtcacaaacaagtatctcccttttttgttccattttctataacaaacatgggctctgcattgcttggtattgaacttga tttcagaggtcctaattattcaatctctgcggcttgtgctacatccaattttgcttctactctgctgccaatcacatccgc cgcggtgaggctgatttgatgattgctggagggactgaggcaatgatagttccgattggattgggaagctttgctgc gtgtagggcactgtctcagagaaatgatgatcctcggacggcttcaaggccgtgggacaaagatcgagatggg cttgttatgggtgaaggtgctggagtattggtaatggaaagcttggaacatgcaatgaagcgaggtgcaccaattg ttgctgagtatttgggaggtgcggtgaactgtgatgctcatcatatgaccgatccaagggccgatgggcttggcgtc tctaactgcattaaactttgtcttgcaaatgctggcgtgtcacctgaagaggttaattacataaatgcacatgctacttc cactcttgccggtgatctagctgaggtaaatgctatcaaaaaggtattcaagaacacatccgggatcaagatcaa cgcaacaaagtctatgatcgggcatagtctcggtgctgctgggggtttggaagccattgccacagtgaaagctata aacacaggctggcttcatcctaccataaatcaatttaatctggagcctgaagtcgaattcgacactgttgcaaatac gaagcaacagcacgaagtaaatgttgccatttcaaactcgttcggatttggcggacccaactctgtcgtggcgtttt ctgctttcaagccctaatcacgaagatgtttgagttcgttccatcctttcatggttactgcatatcttccttgaagcccca ttgtactttgttgtataataatcgcgtcatgttggccaagcggcttaggcgcttatacctttagaatcattgttatcaattg aactatctggmcatttagatcttattcaaaaaaaaaaaaaaaaaaaaaagagccagtc
*Pxh*KAS Ic	tccatactgaggcttcaccgcttgggggattagagccatctctgagttcccttcaaacttccaggcttcggaaaatgt aagagccaatccagggccaaatgagttctcccaactggttttttgctcttaagcaagaaatcggatcccaagaagc gagtagttatactgggatgggggcttgtttcagtttttggcaatgacaattgatgtattctatgagaaactttctgagtgga gacagcgggattagcctaattgacagggtttgatgcttcaagcttttctgtcccggtttggcggacagattcggggtttt ttcttcgaaaggctacattgatgggaagaatgatcgccgccttgatgattgctggaggtactgtttggttgctggtaag agggctcttgaagatggcaaccttggacctcaagtgcttgaaactatggacagaacaaaaattggagtgcttgtg ggaactggcatgggaggcttaacagctttcagcagtggagttgaatctctaatccaaaagggatacaagaagat atctccatttttcattccttattccataaccaatatgggttctgcattgttggccattgacactggcttaatggggcctaatt actccatttcaacagcttgtgcaactgcaaattactgcttttacgcagctgcgaatcacattcgaagaggcgaagct gatatcatggtagttggagggactgaggcagcaatcatgcctactgggggttggtgggttcatagcatgccgggctc tatctcaaagaaacgccgaaccccatagagcttcaaggccttgggacaaaaaccgcgatggctttgtcatggga gaaggcgctggagttctgataatggagagcttggagagtgcaatgaagagaggagccaatatagttgctgagt wtttgggaggtgctgtaacctgtgatgctcatcacatgactgatcctcgctcagacggccttgagwcgtttgcaactt gcataaccaagagtctagaagatgccggtgtttcccccgaagaggtgaactatgtgaattgccatgccacatcaa ccttagcaggggattkggctgaggttaatgcaatcaaaaaggttttcaaggacacttctgagatgaagatgaatg gaactaagtctatgattggacatggccttggagctgctggtgggttggaagctatcgcaaccattaaagcaatcaa cactggttggctgcacccaacaatcaaccaagatgatctggagccttcagttacaattgacacagtcccaaatgtg aagaagaggcatgaagttaatgttgctatctctaactcatttggctttggtgggcacaattcagtggttgtctttgctcct tttacaccctaagaagaacttgtgacaaaagtactcttctttttttgcattactgttactaaaatcattgattttgtgataat agaatgaacatgatagagataatgtgtctcatcagtatcatgtttgctcgtgacgagtattgaagacaatgtacctcctt taacctgtttcattatacgtcaacaccacgcgtgttgtaagcaaattcagagctcattgattaataaaatttctatgcta ttttgttccctcgtgcc
*Pxh*KAS Id	ctaagtgtaaactcaacaactcaatctcgtggcgtgcaaacttcgcccttcgatgattcctaaacgctacaaaaact cgatattgaataagatctaaatgtccagatagttcaattgataacaatgattctaaaggtataagcgcctaagccgc ttggccaacatgacgcgattattatacaacaaaagtacaatggggcttcaaggaagatatgcagtaaccatgaa aaggatggaacgaactcaaaccttctcgtgattagggcttgaaagcagaaaacgccacgacagagttgtgtcc gccaaatccgaacgagtttgaaatggcaacatttacttcgtgctgttgcttcgtatttgcaacagtgtcgaattcgactt caggctccagattaaattgatttatggtaggatgaagccagcctgtgtttatagctttcactgtggcaatggcttccaa acccccagcagcaccgagactatgcccgatcatagactttgttgcgttgatcttgatcccggatgtgttcttgaatac cttttgatagcatttacctcagctagatcaccggcaagagtggaagtagcatgtgcatttatgtaattaacctcttcag gtgacacgccagcatttgcaagacaaagtttaatgcagttagagacgccaagcccatcggcccttggatcggtc atatgatgagcatcacagttcaccgcacctcccaaatactcagcaacaattggtgcacctcgcttcattgcatgttc caagctttccattaccaatactccagcaccttcacccataacaaacccatctcgatctttgtcccacggccttgaag ccgtccgaggatcatcatttctctgagacagtgccctacacgcagcaaagcctcccaatccaatcggaactatcat cgcctcagtccctccagcaatcatcaaatcagtctcaccgcggcggatgtgtttggcagcagagtggaagcaaa

	agttggatgtagcacaagccgtagagattgaataattaggacctctgaaatcaagttcaataccaagcaatgcag agcccatgtttgttaaagaaaatggaacaaaaaagggagatacttgtttgtgacccttctctatcactttctgaactcc accagaaaacaccgtcggaccacccaatcccattccaacaagcacaccagctctctctttatcaatcttggagag caaatctccggcaggatcggcatcttcaat
*Pxh*KAS II	cctcactttaatgctgtctcttttggcaatggtgtttttttatatttataaatatcttgtgttatcaattttttgtgttgaaatgataa aatagcaaggacataatttatactccaaatttgtaagcgtgaagtttctcaattgtcaggaatgctagctcttgtactct ttgacttggttcttctgtaatcgtttggttgtaagatggatattttgatttcaatatagaattccggatcaaactttctgcaag agttcattggttcattaacacggggacggagtgagtttgggattggtgccgaaccaagtactaaacatttggcgtct atcgacatcaaccattcaggctcttctcagcagtgaagtattgtccggataccctcgcaagcgctattcagatcact gcaaatatgcccacccaatgtttacaagttaaaactattcttgttcatatgtttgaagaatatgtaacttgaggattatg attgcaaaattgaaacaaattcattgtctgaagcgtgattaccattctctacttgtacatcatgaaaaagaaaaaac aagagatagtacaatgataacaagaacatcaacaaggccaaattctacaatatttgttaatgtgatgtcacaaaa aagacaagacagtgttttaatacatggaaatggaagtcatgctcgctttctcttcatgtccagaccataggagcaca tcaacagatccgccgccgtatgcaaggccgcaagcgcattataatatagggtagcaaatgactactcccccttggc tacaatactaccattacacaggttggagtagcttgcaaaatctattctacttgtaaggaacaaatataattgatgagtt gtgtccgccaaacccaaatgaattagacagcgccaccttgatgtctattctttcttttttagggccaaccaacagattt gtatccacacccgcgtctggattttccaggttgatatttgggtgaacccacccctgttcgtattgcctgtactgtcgcaac tgcttccacggcacccgatgctcctagtagatgaccgatcatggattttgtagagttcactcttagctccggattctgg ccaaaacatttaacgagggctcgatactctttaaggtctccagatagtgtggatgtagcatgtgcatttatgtaattta catcttcctttgatactcctgcctgagccaacgcttttcaatgcagagaataacaccaagcccatctggatgtggct cggtcatgtgataagcatcgcatgtgaaacttccaccaagaaattctgcataaatatttgcacctctcctcttagcat gctctaattcttccaaaagcagaactccagctccttcacccataacaaatccatcacggttctgataaattacaaac aaaaagttcaaagtaaataacacacatgaaacagaaaaaaggaacaaaaccaatattacagtttgattttctgc cttgaataaacacatttacatgataaaaggaggaaagggtgaaaagaaagcacaagaaattgtaataacatta cactatcccaagggcgtgatgctttggtcgggtcactgttcctttgtgaaagagccctgcatgcaacaaaacctccc aaacctgcacccaaaattatgccaacttcattgcacatcaagttgactcatccttaaggaaatgtatataccagaa agaaagggataccaataggtataattgctgcatctgagccaccacaaagcattacatcctgcattttcgaacagttt tagaatacacatagcacgaaatgtaataaatttgaagtttaactatttgtttcagcaatattaataaagaattcaataa aaatatacaagaaaccactacaaattgcttggcttacgagcattaatgcaaaatacttacagcttcacctcgaatg atatggtttgcagcattcagtatacaaaagttgccagtagcacaagcagtggaaattgagtaatttgggcccatcc atcccagatccattgcaagcatggcagaacccatatttgtggtcgcaaaaggtacacagaagggattcattttcct atatgagatccttaaggcttcaattgcatcatgaaaaaccttcatgccacccatagctgaaccaatcaagacacca catttagttttatctaattcatccatcacatctccagtaatcccaccatctgccaaggctttcttccagcagtaagcatg taaagcatgaatttatccatcctcttggaaagtttggtgcaacccatccatcagtcgagaaagatttgatctctccag caatcctc
*Pxh*KAS III	gcacgagggatttttttcttcggaccctcctaggaaaaagcagcctctgattcgctctgtctcaggcctctgctcttcatt gattttatattcatttcctgaatccaaaatggtgtctagacgctcttgtttccgtcattttcttgttcctccacacagtcgctg agcaagcgagcgagctttgttgtgcgttgtcgctttcgtttcgaggatacaaaatcacatgtgctcggaatttgggag cttgcttcgtgtttagccagtgagtgttgtagtagaagttggagttaattaatggccagtgcatctgggtttctcgctcctt cagttccgagcctccggagctcgattgggtcttctgggtccgttttccgatctgggtttcgctccaatgttggagcttgc aaatgggtcgtctgctctggagtcagtgaaggcgcgactgggggttgcgcaatctcgagtgcccaggctagtcagt aaggggtgcaagttagttggatgtggctcggcaacaccaacacttcagatttcgaatgacgatcttgcaaaatac gttgaaaccaacgatgaatggatatcgtacgcactggcattcgtaatagacgagtacttacaggaaaagatagc ttgacaagtttagcagcagaggcagcgagaaaagcgcttgacatggcacaagttgatcctgatgacctggactt agtcttgatgtgtacatctactccggaggatcttttcggcagtgctcctcagatccaaaaagcactcggctgcaaag gacaaccgttgtcctatgatatcacagctgcatgtagcggatttgtattgggtctagtctcggccgcctgtcatattag gggaggcgggttcaaaaacgttctagttattggggctgatgcgctttctcgatacgttgattggaccgatagaggga cttgcattctctttggagatgctgctggtgccgtcgtagtacaggcctgtgatagcgaagacgatggcttatttgcttttg atttgcacagcgacggtgatggccaaagacacttaaatgcgagtatgagagaaactgaaacaaataaagagg gctctaatggttcagttttagacttccctcctaggcgctcctcgtattcttgcatccatatgaacgggaacgaggtttttc gctttgcttgccggtgtgttcctcagtcaattgaagctgcactcgacaaggccggcctcacctcctctaatatcgattg gttactcctcatcaggcgaatcagaggatcctcgatgcagtttccacccggttggaagtcccgaaagaaagggt gatatcaaatttggcaaattacggcaacacgagcgctgcatcgattccattggcactcgacgaagctgtccggag cgggaaggtgaaggcggggcagaccatcgcgactgccggtttttggagctggactgacttggggctctgctattct cagatggggatgaaactgtcctttcactaccatcaatggccaaaacagcagagaaaaagaagagaagaaga

	atgtgtttttgtccaatccttgtttcatgtgtattttccaagatacttgtttgtgttattttctttgaatcccattacatttcaccttta acaatcacaataaatctctattcttttcattgttctcatgacttaaaacacatgtattcttcattatattacaagtgaagta ggtttcaatctttcaaaaagcgasycgcccagtc

CURRICULUM VITAE

Richa Arunkumar Singhal
Department of Biology, Life Sciences Rm 139
University of Louisville
Louisville KY 40292
Phone: 502-802-6410
Email: rasing02@gmail.com

EDUCATIONAL QUALIFICATION

YEAR	DEGREE	INSTITUTE	Graduation
2009-2016	PhD in Biology	University of Louisville, College of Arts and Science, Department of Biology	(Expected completion 12/2016)
2005-2007	M.Sc. in Biophysics	Department of Biophysics and Biotechnology, Mumbai University.	Graduated with Distinction and ranked 2nd in the university
2002- 2005	B.Sc. In Biotechnology	Kischinchand Chellaram (K.C.) College, Mumbai University.	Graduated with Distinction

PROFESSIONAL QUALIFICATION

YEAR	DEGREE/ CERTIFICATE	INSTITUTE
2014	Entrepreneurship Academy	University of Louisville
2014	Grant Writing Academy	University of Louisville
2013	Graduate Teaching Academy	University of Louisville
2006	Diploma In Modern Instrumentation Technique	Western Regional Instrumentation Centre. (Mumbai University)

2004	Diploma In Plant tissue culture and Secondary metabolite production.		Padmashree Institute of Biological Sciences.(BANGLORE, India)

YEAR	POSITION	TEACHING	INSTITUTE
2009-2015	Graduate Teaching Assistant	Biology 104 (Introduction to biology), Biology 244 (Principles of Biology), Biology 331 (Genetics).	University of Louisville, Department of Biology
2007-2008	Lecturer and Lab Instructor	Biophysics and Biotechnology	University of Mumbai
2007-2008	Science Faculty	Basics of Physics, Chemistryand Biology	Sampath Classes, Mumbai (India)
2002-2008	Science Faculty	Basics of Physics, Chemistryand Biology	Shanti Tutorials, Mumbai (India)

RESEARCH EXPERIENCE

University of Louisville, Department of Biology, Louisville, KY   2009 -
Present
Ph.D. Student
Advisor: Dr. David J. Schultz
Project Title: Identification and characterization of genes involved in metabolism of n5 monoene precursors to n5 anacardic acids in the trichomes of Pelargonium x hortorum.

University of Mumbai, Department of Biophysics, Mumbai, India   2005-2007
Master's Student
Advisor: Dr. Prabhakar Dongre.
Project Title: Aloe polysaccharides and their versatile impact on health systems (Literature Review).

University of Mumbai, Department of Biotechnology, Mumbai, India   2002-2005
Bachelor's Student
Advisor: Dr. Jaya Shashtri
Project Title: Antimicrobial activity of Barbadensis miller a Aloe vera plant.

PRESENTATIONS/POSTERS

- **Oral Talk at KAS Nov 2015. Awarded 1st Place.**
  "Characterization and identification of fatty acid biosynthesis genes involved in metabolic channeling to anacardic acid biosynthesis". Richa A. Singhal[1] and David J. Schultz[1]
- **Poster Presentation at Research Louisville May 2014.**
  Cloning of ESR1 into pRI 201-AN and Transformation of Tobacco Cells
  Jan D. Zielonka[1], Richa A. Singhal[1], Carolyn Klinge[2], David J. Schultz[1]
- **Oral Talk at KAS Nov 2011. Awarded 3rd Place.**
  "EST Database approach to identify fatty acid biosynthesis genes involved in metabolic channeling to anacardic acid biosynthesis". Richa A. Singhal[1] and David J. Schultz[1]
- **Poster Presentation at Research Louisville Oct 2011.**
  "EST Database approach to identify fatty acid biosynthesis genes involved in metabolic channeling to anacardic acid biosynthesis". Richa A. Singhal[1]and David J. Schultz[1]

GRANTS AWARDED

- AandS Grant, (University Of Louisville Research Grant, 2013), $500
- GSC Grant, (University Of Louisville Research Grant 2013), $300
- IRIG Grant, (University Of Louisville Research Grant 2013), $5000

MANUSCRIPTS UNDER PREPARATION

1. Richa A. Singhal[1] and David J. Schultz[1]. Identification and expression analysis of fatty acid biosynthesis genes from Pelargonium x hortorum.

2. Richa A. Singhal[1] and David J. Schultz[1]. Co-expression of geranium acyl carrier protein with Δ^9 14:0-acp desaturase

3. Richa A. Singhal[1], David J. Schultz[1]., et. al. Micro-RNA database for Pelargonium x hortorum.

4. Richa A. Singhal[1], David J. Schultz[1]., et. al. Production of unusual monoenes, anacardic acid and gene expression affected by temperature in trichomes of Pelargonium x hortorum.

MEMBERSHIPS

- Kentucky Academy of Sciences.
- American Society of Plant Biologist.
- UT-KBRIN Bioinformatics.

ACADEMIC and PROFESSIONAL ACHIEVEMENTS

- Awarded a certificate for being an outstanding Professor nominated by students, Faculty Favorite 2013-2014. University Of Louisville.
- 1st Place Graduate Student Presentation Award at the 97th Annual Kentucky Academy of Science Meeting, 2011 (Cellular and Molecular Biology Section
- 3rd Place Graduate Student Presentation Award at the 97th Annual Kentucky Academy of Science Meeting, 2011 (Cellular and Molecular Biology Section)
- Stood 2nd in M.Sc. Part 1 and Part 2 (Biophysics) in Mumbai University.
- Awarded scholarship for academic excellence by Mumbai University.
- President of SGI USA group, an on campus RSO at University of Louisville, 2009-current.
- Fundraising Chair for Biology Student association, Fall 2014 - Fall 2015.

LABORATORY SKILLS

Molecular, Cellular and Microbiology:
PCR (basic, touchdown, gradient, qRTPCR).
Cloning (bacterial vectors, plant vectors).
Expression of genes in bacterial system.
Expression of genes in plant system.
Plant Tissue culture.
Genetic Transformations.
Protein purification.
DNA, RNA, protein and metabolite extractions.
Electrophoresis (AGE and PAGE).
Maintaining E.coli and agrobacterium cell lines.

Bioinformatics:
EST database analysis
Micro-RNA database analysis.
RNA sequencing database analysis.

Biochemistry:
High Pressure Liquid Chromatography.
Gas Chromatography.
Thin Layer Chromatography.