Two methylketone biosynthetic enzymes from wild tomatoes

By

Geng Yu

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Molecular, Cellular and Developmental Biology)
in the University of Michigan
2013

Doctoral Committee:

Professor Eran Pichersky, Chair
Professor Ursula Jakob
Professor Jianming Li
Professor Ronald W. Woodard

Dedication

To
My family for their sacrifice and love

Acknowledgements

I would like to express my sincere thanks to Dr. Eran Pichersky for his tremendous support and guidance throughout my Ph.D study. This dissertation would not be possible without the numerous hours he spent on reviewing and editing my writing. He has given me countless constructive suggestions to make my research progress quickly, the freedom to go to the conferences of my choice and the logic training for me to correctly analyze even the negative results and work more efficiently. I would also like to thank my thesis committee Drs. Ursula Jakob, Jianming Li, Ronald W. Woodard for their advice and comments during my doctoral research.

I enjoy my working experience with the current and previous Pichersky lab members and also the seminar discussions I had with the 4th floor plant research lab members. I want to thank specially the postdocs I have directly worked with: Drs. Ines Shauvinhold, Thuong Nguyen and Nazmul Bhuiyan, for their guidance and help at the beginning stage of my Ph.D study. Vassiliki Falara, Dagan Hammar, Yuki Matsuba, and Reza Sohrobi have made valuable contributions to my oral defense talk. My thanks also go to my collaborators: Dr. Anthony L. Schilmiller in Robert L. Last lab and Dr. Yongxia Guo in Joseph Noel lab who have worked with me on two different manuscripts. Also, the department of Molecular, Cellular and Developmental Biology has a fully supportive group of staff, including Mary Carr, Kelly Campbell, Sheila Dunn, Diane Durfy, Jacqueline Glebe, Edward Grant, Gary Phillips, Laura Moggo and Gregg Sobocinski, who provided various help in different ways during my Ph.D. study. I also want to thank

the members in our self-organized student-seminar group who helped to improve the oral presentation skill before my prelim and oral defense talks.

Finally, I would like to express my appreciation to my mother Liping Zhang, who had lost my father Tingqi Yu just one year before I was admitted to graduate school but still made the sacrifice to let me study abroad. In addition, my wife Jingya Wang has made great support for my study here. Their love and support have made my study in US much easier.

Table of contents

List of Figures

List of Tables

Abstract

Many plants make varying amounts of aliphatic 2-methylketones, but their mode of synthesis is unclear. High concentrations of 2-undecanone and 2-tridecanone in the trichomes of a wild tomato species *Solanum habrochaite*s have been found to be toxic to pests. Since cultivated tomato (*S. lycopersicum*) produces only low levels of methylketones, the creation of genetic hybrids that produce high levels of methylketones was previously attempted but was unsuccessful. Thus, elucidating methylketone biosynthesis may allow the generation of cultivated tomatoes capable of producing methylketones by genetic engineering.

A recent study from our laboratory suggested that plants produce methylketones through *de novo* fatty acid biosynthesis and identified *ShMKS1* (*S. habrochaites* methylketone synthase 1) as a gene involved in methylketone biosynthesis. Further genetic analysis of the segregated progenies of the interspecific crosses between *S. habrochaites* X *S. lycopersicum* identified another gene, *ShMKS2* (*S. habrochaites* methylketone synthase 2), which was shown to be necessary for methylketone production. I heterologously expressed of *ShMKS2* in *Escherichia coli* and observed the production of methylketones. Enzymatic studies of ShMKS1 and ShMKS2 showed that ShMKS2 has a thioesterase activity, hydrolyzing the thioester bond in *3*-ketoacyl-ACPs to produce the corresponding 3-keto acids, and that ShMKS1 has a decarboxylase activity towards 3-keto acids, resulting in the production of 2-methylketones. Both enzymes were found to be localized in plastids, the site of fatty acid biosynthesis in plants.

Constitutive co-expression of *ShMKS1* and *ShMKS2* led to the synthesis of methylketones in Arabidopsis and tobacco but also disrupted growth and development of the transgenic plants. Trichome-specific expression of these two genes in cultivated tomato, by using the promoters of *ShMKS1* and *ShMKS2*, did not result in detectable methylketone production. All transgenic plants expressing *ShMKS2* in leaves produced myristic acid, suggesting that ShMKS2 can hydrolyze myristoyl-ACP as well.

The specificity of two tomato trichome-specific promoter was found to be maintained in other species.

The results demonstrated the mechanism for methylketone biosynthesis in plants and that any plant species could be engineered to produce methylketones by co-expressing *ShMKS1* and *ShMKS2* in leaves.

Chapter 1 Introduction

1.1 BIOLOGICAL OCCURRENCE AND FUNCTIONS OF METHYLKETONES

1.1.1 Definition of methylketones and the scope of this research

Methylketones are ketones with at least one methyl-carbonyl moiety. Although different organic moieties can be linked covalently to the methyl-carbonyl, this dissertation will only discusses the aliphatic methylketones, as aliphatic methylketones are the major form of methylketones that originate in natural biological processes (Forney and Markovetz, 1971).

1.1.2 Methylketone occurrence and function in mammals

Acetone, the only methylketone detected in mammals, is produced during fatty acid catabolism in liver or kidney. Acetoacetate and β-hydroxybutyrate can coexist with acetone, forming what is known as “ketone bodies” (Campbell and Farrell, 2006). Under normal conditions acetone is maintained at low levels in the body by being secreted together with urine. Acetoacetate and β-hydroxybutyrate are finally converted to acetyl-CoA outside the liver, because the liver lacks the enzyme to oxidize the β-hydroxybutyrate to acetoacetate or convert the acetoacetate to acetoacetyl-CoA. Under abnormal conditions, such as high-fat catabolism resulting from starvation or diabetic disorders, acetone accumulates in the blood and can be detected in one’s breath and urine

(Kalapos, 2003). Acetone is also detected in mammals maintained on high-fat or carbohydrate-free diets.

The function of ketone body is to supply energy for tissues other than liver under high fat catabolism conditions, but acetone is a dead-end metabolite that cannot be catabolized further by mammals.

1.1.3 Methylketone occurrence and function in insects

Methylketones are also found in some insects, specifically in the exocrine glands of ants and honeybees. For example, 6-methylhept-5-en-2-one was detected in the ant species *Dolichoderus attelaboides* (Cavill and Hinterberger, 1960), 2-heptanone was detected in honeybees (Shearer and Boch, 1965), and 2-tridecanone, 2-pentadecanone and 2-heptadecanone were detected in the ant species *Acanthomyops claviger* (Regnier and Wilson, 1968).

Methylketones secreted by ants and honeybees are important chemical signals used to trigger behavior changes within the same species. For example, 2-heptanone secreted by ant *Conomyrma pyramica*, serves as an alerting and recruiting signal in low concentration while an alarming signal at high concentration (Blum and Warter, 1966). 2-Heptanone secreted by honeybees acts as a signal to recruit guard-bees or as a chemical marker or anesthetic for small arthropods during a bite (Papachristoforou et al., 2012). Some methylketones secreted by insects, like 2-tridecanone and 2-undecanone, have a pungent odor, possibly providing a defensive function for insects.

1.1.4 Methylketone occurrence and function in microbes

Dairy products, like blue cheese and butter, have methylketones in their aromas due to the microbes residing in them (Jolly and Kosikowski, 1975). The odious smell of rancid food is mainly due to the methylketones produced by the microbes degrading any remaining fat (Alford et al., 1971).

The physiologic role of methylketones for the methylketone-producing microbes has not been documented, and these compounds were regarded as dead-end metabolites. Methylketones are produced continuously from decaying plant remains but do not accumulate in the environment, potentially making these chemicals an important carbon and energy source for various microorganisms in the food chain (Forney and Markovetz, 1971).

1.1.5 Methylketone occurrence and function in plants

Methylketones were first documented in plants as principle components in the essential oils of common rue (*Ruta graveolens*) by C. Greville Williams in 1857 (Williams, 1857). Later, methylketones were found in lime (*Citrus aurantifolia*), clove (*Eugenia aromatica*), hop (*Humulus lupulus* L.), coconut (*Cocos nucifera*), cinnamon (*Cinnamomum verum*), matsubasa (*Shizandra nigramaxim*), palm (*Trachycarpus fortunei*), peanut (*Arachis hypogaea*), sunflower (*Helianthus annuus*) and a few other plant species (Watts, 1886; Buttery et al., 1963; Forney and Markovetz, 1971). A prominent methylketone producing plant is a wild tomato species, *Solanum habrochaites* sp. *glabratum* (formally called *Lycopersicon hirsutum* f. *glabratum*), which accumulates up to 8 mg/g fresh leaf weight of methylketones in its young plant leaves (Fridman et al., 2005).

Methylketones serve as odorants in some plants, which may function as chemical signals to attract pollinators or to repel herbivores. In the species *S. habrochaites* mentioned above, methylketones are made in large amounts in its gland hairs and are used as direct chemical toxins to kill insects making contact with its leaves (Williams et al., 1980).

1.2 BIOSYNTHESIS OF METHYLKETONES IN DIFFERENT ORGANISMS

1.2.1 Methylketone biosynthesis in mammals

In mammals, acetone is produced during fatty acid catabolism. Fatty acid catabolism occurs when carbohydrate sources are limited, which is common when fasting or maintaining a high-fat diet. While fasting, gluconeogenesis is activated to maintain blood-sugar levels, which consumes oxaloacetic acid. Reduced oxaloacetic acid levels prevent acetyl-CoA, generated from fatty acid catabolism, to enter the tricarboxylic acid cycle (TCA cycle), causing acetyl-CoA to accumulate. In high acetyl-CoA conditions, two acetyl-CoAs form acetoacetyl-CoA, catalyzed by thiolase. Then acetoacetyl-CoA is converted to β-hydroxy-β-methylglutaryl-CoA (HMG-CoA) by HMG-CoA synthase. Lastly the HMG-CoA is converted to acetoacetate and acetyl-CoA by HMG-CoA lyase. HMG-CoA synthetase and HMG-CoA lyase mainly exist in the mitochondria of hepatocytes, so liver is the major organ to generate keto bodies. Acetoacetate is quickly converted to β-hydroxybutyrate by the enzyme β-hydroxybutyrate dehydrogenase, which is highly active in the hepatocytes (Figure 1-1). Acetone is mostly produced by the spontaneous decarboxylation of acetoacetate and produced by the decarboxlase in mammals in small amount.

Figure 1-1. Keto body biosynthesis

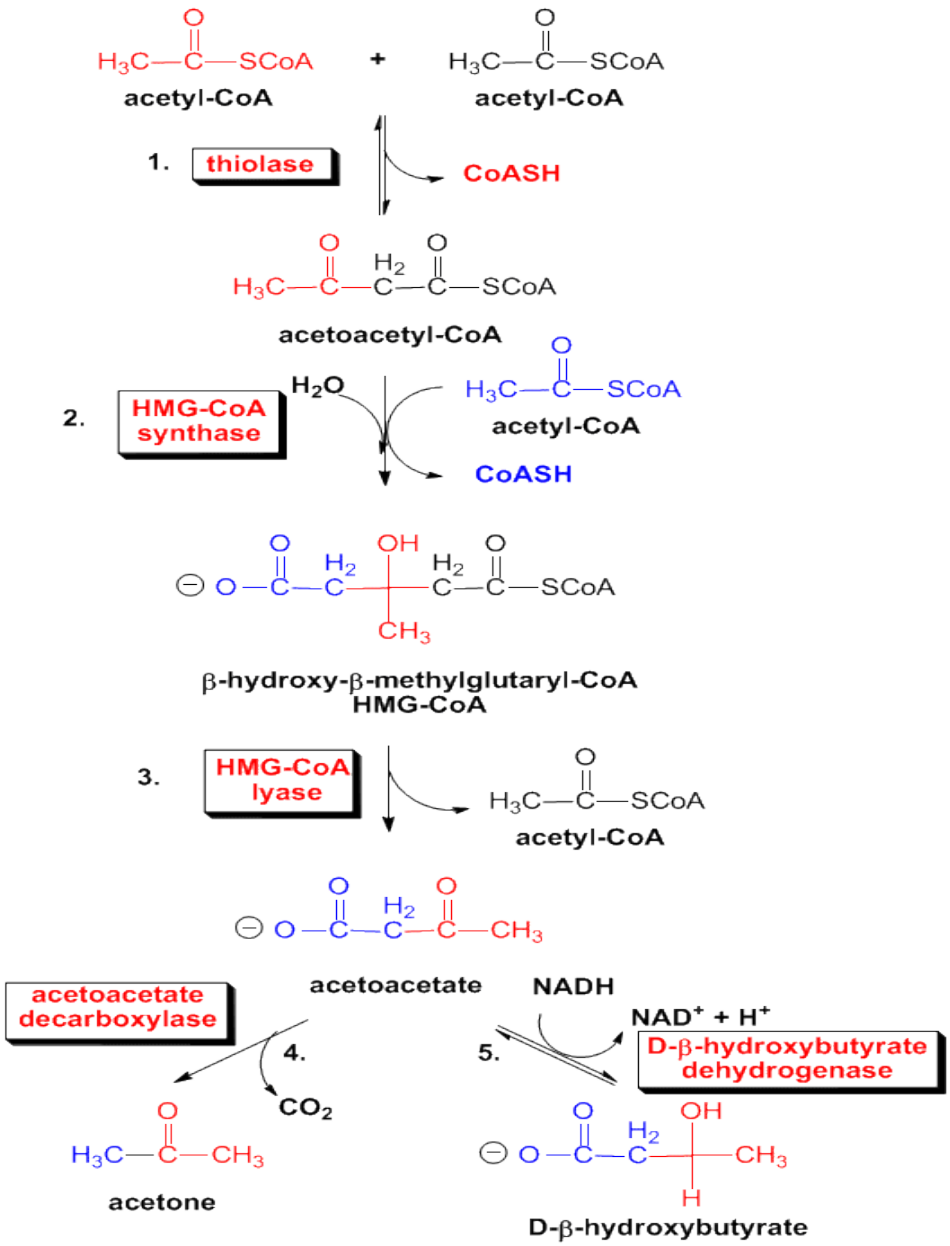

Figure 1-1. Keto body biosynthesis. Two acetyl-CoA are catalyzed by thiolase to form one acetoacetyl-CoA, then the acetoacetyl-CoA is catalyzed by HMG-CoA synthase to produce *β*-hydroxy-*β*-methylglutaryl-CoA (HMG-CoA), followed by acetoacetate production with the enzyme HMG-CoA lyase. Acetoacetate is decarboxylate to acetone enzymatically or spontaneously. Acetoacetate is also converted to hydroxybutyrate by the catalysis of hydroxybutyrate dehydrogenase. (Figure source: http://employees.csbsju.edu/hjakubowski/classes/ch112/pathways-charts/ketonebodysynth.htm)

1.2.2 Methylketone biosynthesis in insects

In insects, there is no reported study on reactions in methylketone biosynthesis. Since fatty acid metabolism is highly active in insects and most methylketones found in insects have a saturated straight carbon chain linked to the methyl-carbonyl functional group, therefore it has been postulated that insect methylketones were derived from fatty acids (Forney and Markovetz, 1971).

1.2.3 Methylketone biosynthesis in some fungi and bacteria species

Biochemical studies on some fungi species have provided insight to elucidate the methylketone biosynthesis mechanism. Pure cultures of *Penicillium* and several species of *Aspergillus,* isolated from rancid butter, have been shown to produce methylketones when cultured on specific fatty acids (Gehrig and Knight, 1958). In all trials, the methylketones produced were one carbon shorter than the fatty acids supplied to the culture (Gehrig and Knight, 1958), suggesting that *beta*-oxidation followed by a decarboxylation reaction was involved in the conversion of fatty acids to their corresponding methylketones. The mechanism of converting fatty acids to methylketones was postulated as abortive *β*–oxidation (Figure 1-2Figure 1-2), in which fatty acids are oxidized to 3-keto acids (Gehrig and Knight, 1958). Next, the 3-keto acids are decarboxylated to methylketones. 3-Keto acids are not stable by nature, and they undergo decarboxylation spontaneously. Thus, an enzyme may not be needed to mediate the decarboxylation. Since the degradation of fatty acids starts by linking fatty acids to co-enzyme A (CoA) cofactors, the decarboxylation reaction must occur after a thioesterase-mediated removal of the CoA from the acyl-CoAs. However, the esterase or thioesterase

capable of hydrolyzing the thioester bond of the *β*-ketoacyl-CoA has not been identified yet.

Figure 1-2. Methylketone generated by postulated abortive *β*-oxidation pathway

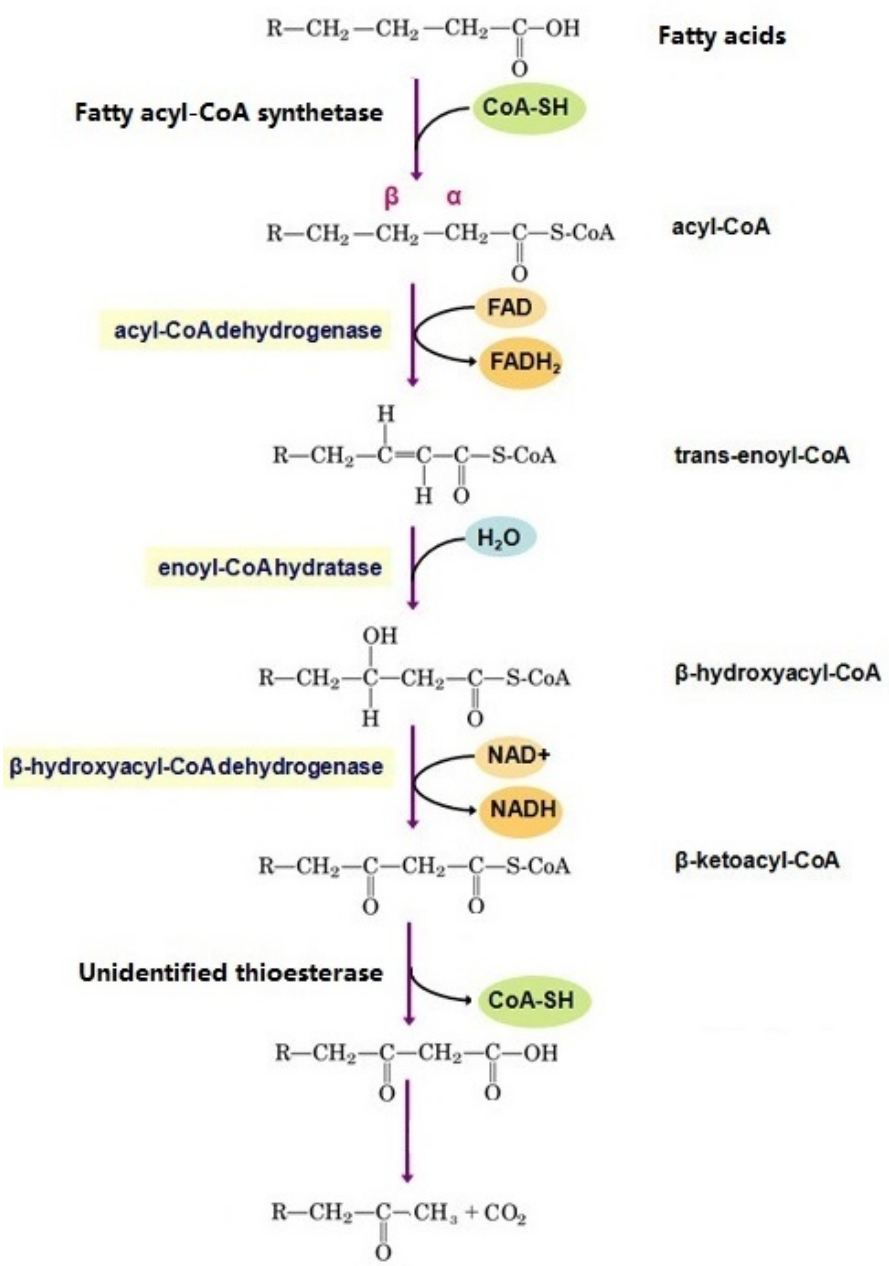

Figure 1-2. Methylketone generated by postulated abortive *β*-oxidation pathway. Fatty acids are activated by binding to a CoA moiety and oxidized to *β*-ketoacyl-CoA. After the CoA are hydrolyzed, the resulted 3-keto acids generate methylketones. (Figure adapted from http://www.nature.com/scitable/content/reactions-of-beta-oxidation-14897250)

Some bacterial species also convert alkanes to methylketones. First, some bacteria convert alkanes to methylketones (Lukins and Foster, 1963). When the bacterium *Mycobacterium smegmatis* was grown on *n*-butane, *n*-pentane or *n*-hexane as the sole carbon sources, it produced 2-butanone, 2-pentanone and 2-hexanone, respectively (Lukins and Foster, 1963). In another study, 2-, 3-, 4- and 5-decanone were produced by *Pseudomonas aeruginosa* growing on n-decane. 2-Decanone was produced in the highest concentration, indicating that the second carbon is preferentially oxidized (Fredricks, 1967). A mechanism for catalyzing alkanes to methylketones has been proposed but not yet established. In the proposed pathway, alkanes first gain a free radical on the carbon 2 position, which binds to oxygen to form 2-hydroperoxides. Next, these 2-hydroperoxides are reduced to secondary alcohols and finally oxidized to methylketones (Leadbetter and Foster, 1960).

1.2.4 Methylketone biosynthesis in plants

In plants, reports about the biosynthesis of methylketones have emerged only recently. A biochemical genomics analysis on the mass methylketone producing tomato *S. habrochaites* PI126449 suggested that methylketones might derive from fatty acid biosynthetic intermediates similar to the esterase-involved mechanism of fungi system. First, the RNA levels of several enzymes involved in fatty acid biosynthesis were found to be higher in the gland hairs of *S. habrochaites* PI126449 than in those of *S. habrochaites* LA1777 (Fridman et al., 2005), indicating that the plant may produce

methylketone through *de novo* fatty acid biosynthesis. Secondly, an esterase like enzyme, *Sh*MKS1 (*Solanum habrochaites* Methyl Ketone Synthase I) was identified, whose RNA, protein and enzymatic activity levels were found to correlate with methylketones levels and gland hair density in a variety of tomato accessions (Fridman et al., 2005). Thirdly, the purified ShMKS1 protein was found to be transported into isolated chloroplasts, in which the fatty acid biosynthesis of plants happens, consistent with the elevated expression of fatty acid biosynthetic enzymes (Fridman et al., 2005). Finally, ShMKS1 was also found to catalyze the decarboxylation reaction (Fridman et al., 2005). Taken together, ShMKS1 was suggested to be a bi-functional enzyme of esterase and decarboxylase, which hydrolyzes intermediates in the fatty acid biosynthetic pathway rather than in the fatty acid degradation pathway to produce 3-keto acids and then decarboxylates the resulted 3-keto acids to methylketones (Fridman et al., 2005).

However, the overall reaction rate of the purified enzyme with the synthesized substrate, β-ketomyristoyl-ACP, was very slow, which raised concerns as to whether other enzymes might be involved in producing methylketones *in vivo*. This dissertation is a follow-up on that question by examining the involvement of other enzymes in the biosynthesis of methylketones and providing further enzymatic characterization of *Sh*MKS1 and the additional enzymes.

1.3 FATTY ACID BIOSYNTHESIS IN PLANTS

Methylketones found in various organisms, including plants, typically have a linear saturated carbon backbone, suggesting they are derived from fatty acids. Plants were postulated to synthesize methylketones through *de novo* fatty acid biosynthesis

(Fridman et al., 2005), so a brief overview of saturated fatty acid biosynthesis in plants is introduced.

Fatty acid biosynthesis in plants mainly takes place in plastids, with a small portion occurring in mitochondria. The carbon sources for plant fatty acid biosynthesis are photosynthates in the form of acetyl-CoA, produced by plastid pyruvate dehydrogenase (Broun et al., 1999). The first committed step for fatty acid biosynthesis is to generate malonyl-CoA from acetyl-CoA by acetyl-CoA carboxylase (ACC, Figure 1-3). After malonyl-CoA was converted to malonyl-ACP (acyl carrier protein) by malonyl-CoA:ACP acyltransferase, malonyl-ACP will condense with one acetyl-CoA by the enzyme ketoacyl synthase III (KAS III) to generate a four carbon β-ketoacyl-ACP. ACP is used as the acyl carrier in all the following reactions; the β-ketoacyl-ACP will be converted first to 3-hydroxyacyl-ACP by β-ketoacyl-ACP reductase, then to enoyl-ACP by hydroxyacyl-ACP dehydrogenase and finally to a fully reduced acyl-ACP by a second reductase (enoyl-ACP reductase). The acyl-ACP will then condense another malonyl-ACP to complete the reaction cycle by generating a two-carbon longer β-ketoacyl-ACP. The new β-ketoacyl-ACP will go through the same reaction cycle for another six or seven times until it reaches 16 or 18 carbons long. The second condensation and the following five condensations were catalyzed by KAS I while the condensation between palmitoyl-ACP and malonyl-ACP was catalyzed by KAS II.

The terminal products of fatty acid biosynthesis, palmitoyl-ACP or stearoyl-ACP, are either incorporated into phosphatidylglycerol by acyltransferases inside the plastids or hydrolyzed by thioesterase to be exported out of plastids. Two classes of plant thioesterase, FATA and FATB (fatty acyl-ACP thioesterase A and B), exist in plants.

FATA prefers oleoyl-ACPs while FATB prefers saturated acyl-ACPs as substrates (Salas and Ohlrogge, 2002). However, a thioesterase that hydrolyzes *β*-ketoacyl-ACP has never been discovered before this study.

Figure 1-3. The reactions of fatty acid biosynthesis

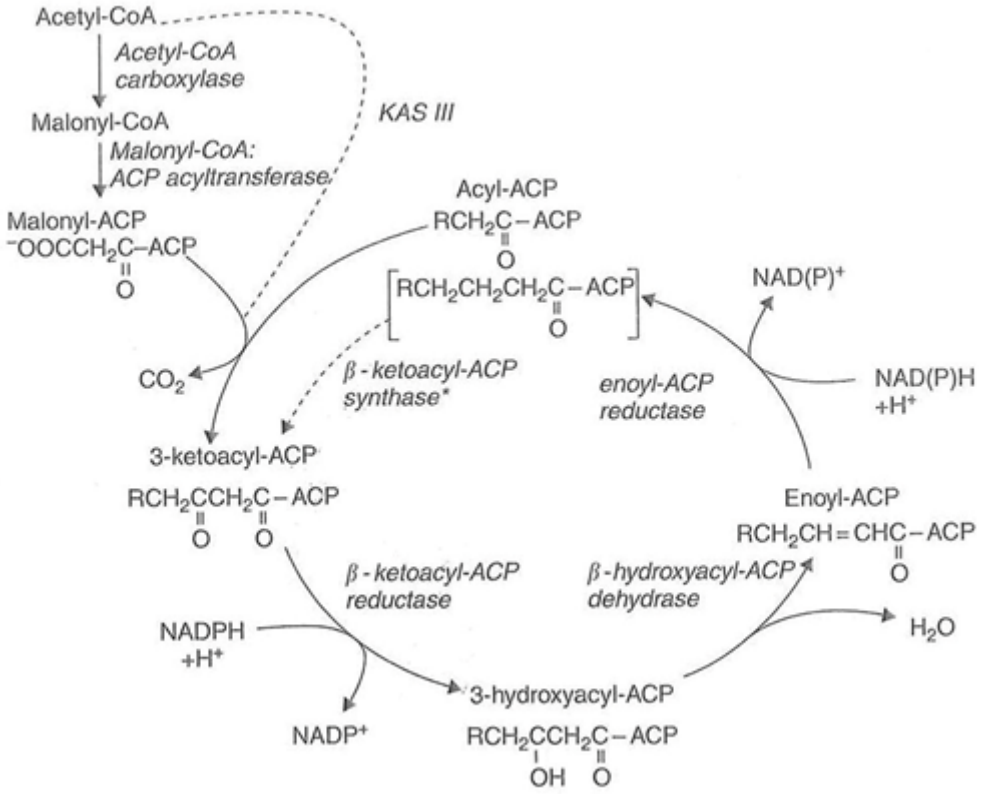

Figure 1-3. The reactions of fatty acid biosynthesis. Acetyl-CoA is activated to malonyl-CoA, then quickly converted to malonyl-ACP. Malonyl-ACP will go through condensation, reduction, hydroxylation and reduction cycle to become a two-carbon longer acyl-ACP, and the saturated acyl-ACP is ready for another condensation to go through another cycle. The iterative reaction will terminate when the carbon chain length reaches 16 or 18. (Figure taken from Gurr M.I., Harwood, J.L. and Frayn, K.N. Lipid Biochemistry [Fifth Edition, Blackwells Publishing, Oxford, 2002])

1.4 TRICHOME STUDY

Since most of this study into methylketone biosynthesis in tomato species will use trichome as a model system, a brief overview of trichome biology is presented below.

1.4.1 Introduction of trichomes

Trichomes are the protrusions of plant epidermis on the aerial part of plants (Schilmiller et al., 2010b). The term is derived from the Greek word “trichos”, meaning hairs (Glas et al., 2012). Trichomes can be found on all the aerial parts of the plants but are present mainly on the leaves and stems. Trichomes are extremely diverse in their morphology (Payne, 1978; Werker, 2000) and they exhibit species-specific shape, color and odor. Therefore, certain characteristics of trichomes are used to identify plant species (Reis et al., 2002).

The most common criterion to classify trichome is whether those hairs are glandular or non-glandular. Glandular trichome could have a single-cell or a multicellular gland on the top of the trichome stalk, which can be detached from the stalk easily. Glandular trichomes are found in one third of vascular land plants (Schilmiller et al., 2010b). For example, tomato (*Solanum lycopersicum*), tobacco (*Nicotiana tabacum*) and the herbs basil (*Ocimum basilicum*), oregano (*Origanum vulgare*) and thyme (*Thymus vulgaris*) all have glandular trichomes. In contract, most angiosperms, some gymnosperms and bryophytes have only non-glandular trichomes (Wagner et al., 2004). The widely studied model plant, *Arabidopsis thaliana*, has only non-glandular trichomes on its leaves.

1.4.2 Chemicals in glandular trichomes

One prominent feature of glandular trichomes is their capacity to synthesize, store and secrete large quantities of specialized metabolites. The characteristic odor of the above mentioned spicy-producing plants is due to various specialized metabolites produced by their glandular trichomes. The major classes of metabolites synthesized by

glandular trichomes are terpenoids (Gershenzon and Dudareva, 2007), phenypropenoids (Gang et al., 2001), flavonoids (Treutter, 2005), methylketones (Fridman et al., 2005), and acyl sugars (Kroumova and Wagner, 2003). However, one certain plant species may only produce a subset of those compounds, and sometimes certain specialized metabolites were only found in limited plant species. For instance, glandular trichomes of cultivated tomato species, *S. lycopersicum,* synthesize large amounts of terpenes and acyl sugars, but very little methylketones or flavonoids, while the trichomes of wild tomato, *S. habrochaites* PI126449, synthesize large amounts of methylketones but very little terpenes.

1.4.3 Biological functions of trichomes

As an appendage to the epidermis of plants, trichomes represent the first line of defense against pathogens and herbivores. Non-glandular trichomes of some plant species, such as Arabidopsis, have a needle-like shape, which can physically obstruct the movement of arthropods on the leaf or reaching for the leaf with their mouths. On the other hand, glandular trichomes use the synthesized metabolites to fend off or directly kill the pathogens or herbivores.

Trichomes also help to absorb excessive UV light due to the presence of flavonoids and other UV-absorbing metabolites in both the non-glandular and glandular trichomes. In addition, trichomes reduce the air-flow on the surface of leaves to retain a humid layer, which helps to reduce the water evaporation when the stomata are open (Ehleringer, 1984).

Since trichomes are epidermal appendages, they can easily be isolated from plants to get DNA, RNA, proteins and metabolites for further studies. This feature of trichomes has facilitated identification of metabolic enzymes from various pathways (Schilmiller et al., 2008). Thus, this dissertation research project makes use of these advantages of the trichome system by utilizing genomic, metabolic and biochemical approaches to further study the methylketone biosynthetic pathway in the tomato.

Chapter 2 Multiple biochemical and morphological factors underlie the production of methylketones in tomato trichomes[1]

2.1 ABSTRACT

Genetic analysis of interspecific populations derived from crosses between the wild tomato species *Solanum habrochaites* f. sp. *glabratum*, which synthesizes and accumulates insecticidal methylketones (MK), mostly 2-undecanone and 2-tridecanone, in glandular trichomes, and cultivated tomato (*Solanum lycopersicum*), which does not, demonstrated that several genetic loci contribute to MK metabolism in the wild species. A strong correlation was found between the shape of the glandular trichomes and their MK content, and significant associations were seen between allelic states of three genes and the amount of MK produced by the plant. Two genes belong to the fatty acid biosynthetic pathway, and the third is the previously identified Methylketone Synthase1 (MKS1) that mediates conversion to MK of β-ketoacyl intermediates. Comparative transcriptome analysis of the glandular trichomes of F2 progeny grouped into low- and high-MK-containing plants identified several additional genes whose transcripts were either more or less abundant in the high-MK bulk. In particular, a wild species-specific transcript for a gene that we named MKS2, encoding a protein with some similarity to a

[1] This work has been published as Imri Ben-Israel, Geng Yu, Michael B. Austin, Nazmul Bhuiyan, Michele Auldridge, Thuong Nguyen, Ines Schauvinhold, Joseph P. Noel, Eran Pichersky, and Eyal Fridman (2009) Multiple Biochemical and Morphological Factors Underlie the Production of Methylketones in Tomato Trichomes. Plant Physiol. 151, 1952-1964. My contribution was to express ShMKS2 in *E. coli* and analyze the volatile from the *E. coli* cultures.

well-characterized bacterial thioesterase, was approximately 300-fold more highly expressed in F2 plants with high MK content than in those with low MK content. Genetic analysis in the segregating population showed that MKS2's significant contribution to MK accumulation is mediated by an epistatic relationship with MKS1. Furthermore, heterologous expression of MKS2 in *Escherichia coli* resulted in the production of methylketones in this host.

2.2 INTRODUCTION

Plants exhibit a large range of chemical and morphological variation, reflecting different adaptations to mediating their interactions with the biotic and abiotic environment throughout their life cycle (Ehrlich and Raven, 1964). Some plant chemicals are lipophilic (oily) compounds that have high vapor pressure and therefore volatilize easily when exposed to air. Such volatiles can serve as signal molecules that either attract or repel animals. Many such compounds are also toxic and can damage a predatory organism through external or internal contact and are often synthesized in dedicated cells that also serve to store them (Wagner et al., 2004). In particular, such compounds may be synthesized and accumulated in small epidermal cell extensions on the surface of leaves, stems, and reproductive tissues called glandular trichomes (Schilmiller et al., 2008). Since the initial work on glandular trichomes in mint *Mentha spp*. (Gershenzon et al., 1992), various studies involving transcriptomics, proteomics, and metabolomics have indicated that entire metabolic pathways responsible for the production of such compounds operate within the trichomes and that these unique cells require the import of only the basic building blocks to make these chemicals (Gang et al., 2001; Fridman et al., 2005; Nagel et al., 2008; Schilmiller et al., 2008; Xie et al., 2008).

The cultivated tomato *Solanum lycopersicum* and its wild relative *Solanum habrochaites* f. sp. *glabratum* represent two of the 12 main taxa found within the *Solanum* section *Lycopersicon* (Zuriaga et al., 2009). While only limited genetic diversity is found among the cultivated *S. lycopersicum* accessions, a wide range of variance is found in the wild relatives. This richness of genetic polymorphism is well reflected by the wide repertoire and quantity of specialized compounds accumulated in their trichomes, including monoterpenes and sesquiterpenes (van der Hoeven et al., 2000; Fridmanab et al., 2006; van Schie et al., 2007), acyl sugars (Ghangas and Steffens, 1993), and methylketones (Williams et al., 1980; Fridman et al., 2005). Up to seven types of trichomes have been reported in the various *Solanum* species. One of these trichome types that has been investigated in some detail is the type VI glandular trichome, which is composed of a stalk cell with four cells at the top that form a mushroom-like shape; a cuticular sac wrapped around these cells allows accumulation of secreted compounds similar to an inflating balloon (Snyder and Carter, 1985; Werker, 2000). We have recently shown that in the wild species *S. habrochaites* f. sp. *glabratum*, the type VI glandular trichomes, which are present at high density on both the leaf surfaces and stems, contain two main MK compounds, 2-tridecanone (2TD; containing a 13-C backbone) and 2-undecanone (2UD; containing an11-C backbone), as well as some 2-pentadecanone (containing a 15-C backbone) and a few other unidentified MK. These MK are synthesized and accumulated to very high levels in these trichomes, up to 5,500 mg g^{-1} leaf fresh weight (Antonious, 2001; Fridman et al., 2005).

Analysis of a type VI-specific EST database from a MK-producing *S. habrochaites* f. sp. *glabratum* (accession PI126449 [hereafter referred to as PI]) showed

that transcripts of genes encoding plastidic enzymes of fatty acid biosynthesis are highly represented, in contrast to their relatively low representation in another tomato wild species that does not make MK accession LA1777 (Fridman et al., 2005). The comparative analysis of the two EST databases also led to the isolation and characterization of a novel gene encoding a protein belonging to the *α/β*-hydrolase family, which was specifically and exclusively expressed in type VI trichomes of MK-producing plants but not in nonproducers. Although the protein did not appear to have a transit peptide, the results of plastid import experiments indicated that it could be imported into the plastids.

Since 3-ketoacids are inherently unstable and undergo spontaneous decarboxylation (Kornberg et al., 1948), albeit at a low rate at ambient temperature, the evidence of elevated levels of fatty acid biosynthesis in these trichomes suggested that the observed straight-chain MK such as 2TD and 2UD could be derived from enzymatic or nonenzymatic decarboxylation of the respective C_{n+1} 3-ketoacids. In plants, 3-ketoacyls of fatty acids mostly occur in plastids (as 3-ketoacyl-acylcarrier proteins [ACPs]) as intermediates in the fatty acid biosynthetic pathway and in peroxisomes (as 3-ketoacyl-CoA) as intermediates in the fatty acid degradation pathway (http://lipids.plantbiology.msu.edu/?q=lipids/genesurvey/). The identification of a plastid-localized putative hydrolase led us to carry out *in vitro* assays with this enzyme, subsequently designated as METHYLKETONE SYNTHASE1 (MKS1), with the C12, C14, and C16 3-ketoacyl-ACPs as substrates (Fridman et al., 2005). In these assays, the respective C11, C13, and C15 MK were produced, suggesting that MKS1 is capable of both hydrolyzing the thioester bond and decarboxylating the resulting 3-ketoacid

intermediate. However, it was noted that the turnover rate of the enzyme was unusually low (Fridman et al., 2005).

Crosses between MK-producing and nonproducing lines followed by segregation analysis have indicated that the ability to produce MK requires multiple quantitative trait loci in addition to MKS1 (Zamir et al., 1984). Consequently, it has not been possible to breed cultivated tomato lines that produce high levels of MK in their glands. It is likely that the trait of MK production in *S. habrochaites* evolved through multiple morphological and biochemical changes that took place gradually during evolution. To uncover the additional factors influencing MK production, we took a quantitative genetic approach to identify quantitative trait loci that might affect MK production, including genes encoding biosynthetic enzymes, and tested the possible relationship between trichome characteristics and chemical content. In addition, comparative transcriptomic analysis was used to identify new genes whose differential expression is correlated with MK production in interspecific populations.

2.3 RESULT

2.3.1 Morphological and Chemical Analyses of Interspecific Populations Derived from Crosses between the Cultivated Tomato and *S. habrochaites* f. sp. *Glabratum*

The chemical profiles of leaves of the cultivated tomato *S. lycopersicum* (var M82) and the wild species *S. habrochaites* f. sp. *glabratum* (PI) differ in their shape and chemical content. In particular, leaves of the cultivated tomato contain little or no MK, while leaves of the wild species contain high levels of 2UD and 2TD, which are

synthesized and stored in the type VI glandular trichomes on the leaf surface (Fridman et al., 2005). A series of crosses were conducted between these accessions to genetically dissect the contribution of candidate genes to MK content. Tomato plants of different genetic backgrounds were then evaluated, including the two parental lines: *S. habrochaites* f. sp. *glabratum* (PI) and *S. lycopersicum* var M82 (14 plants of each), F1 hybrids of these parents (14 plants), an F2 segregating population derived from self-pollinated F1 (245 plants), progeny derived from the first and second backcrossing of F1 with M82 (82 and 72 plants, respectively), and progeny derived from the first backcrossing of F1 with PI (22 plants). All plants were randomly planted, and from each, six young leaflets were removed for chemical characterization and 2TD level determination, since 2TD is the major MK produced in the parental wild species (Fridman et al., 2005). Overall, the 2TD levels of most F2 progeny were more similar to the cultivated tomato parent (Figure 2-1). This, combined with the observation of very low values in the backcrossed generations, indicated polygenic inheritance of this trait and suggested the recessive characteristic of the wild species alleles that participate in this pathway.

Figure 2-1. Distribution plot of 2TD levels in the cultivated tomato *S. lycopersicum* (var M82), the wild species *S. habrochaites* f. sp. *glabratum* (PI), the F1 hybrid of these parents, the F2 segregating population derived from self-pollinated F1, and progeny derived from the first and second backcrossing of the F1 with M82 (BC1-M82 and BC2-M82) and the first backcrossing with PI (BC1-PI)

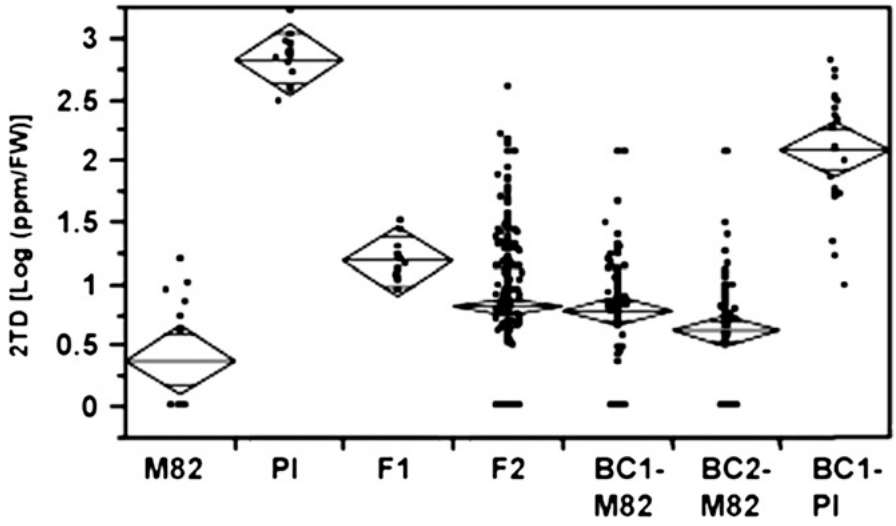

Figure 2-1. Distribution plot of 2TD levels in the cultivated tomato *S. lycopersicum* var M82), the wild species *S. habrochaites* f. sp. *glabratum* (PI), the F1 hybrid of these parents, the F2 segregating population derived from self-pollinated F1, and progeny derived from the first and second backcrossing of the F1 with M82 (BC1-M82 and BC2-M82) and the first backcrossing with PI (BC1-PI). 2TD levels were log transformed. The line across each diamond represents the group mean. The vertical span of each diamond represents the 95% confidence interval for each group. FW, Fresh weight.

Digital images of leaflet surfaces were taken to determine trichome density and its association with MK accumulation. While analyzing these images, we noticed that the F2 population segregates not only for trichome number but also for trichome shape. This observation is in agreement with previously described distinctions in trichome shape between cultivated and wild species of tomato (Snyder and Carter, 1985; Antonious, 2001). While none of the F2 plants showed clear separation of the cells at the tip of the trichomes (as the trichomes of M82 show), 31% of the population had type VI trichomes with partial separation of these cells (M82-like; Figure 2-2, A and B), 18% of the F2

progeny had round type VI trichomes, basically identical in shape to those of the wild species (PI shape; Figure 2-2, A and B), and in 51% of the plants the cells of the type VI trichomes were not separated similar to the M82 parent but the trichome appeared more square than round. The latter morphology was designated as intermediate (Figure 2-2, A and B). Interestingly, on average, plants with PI-shaped trichomes accumulated the highest levels of MK, plants with the intermediate trichomes accumulated intermediate levels of MK, and plants with M82-like trichomes accumulated the lowest levels of MK (Figure 2-2C). The mean MK values of these three groups differed significantly from each other (Tukey's honestly significant difference; $\alpha = 0.005$).

Figure 2-2. Variation and distribution of the gland shape and 2TD content in an interspecific F2 population originated from a cross between the cultivated and wild species

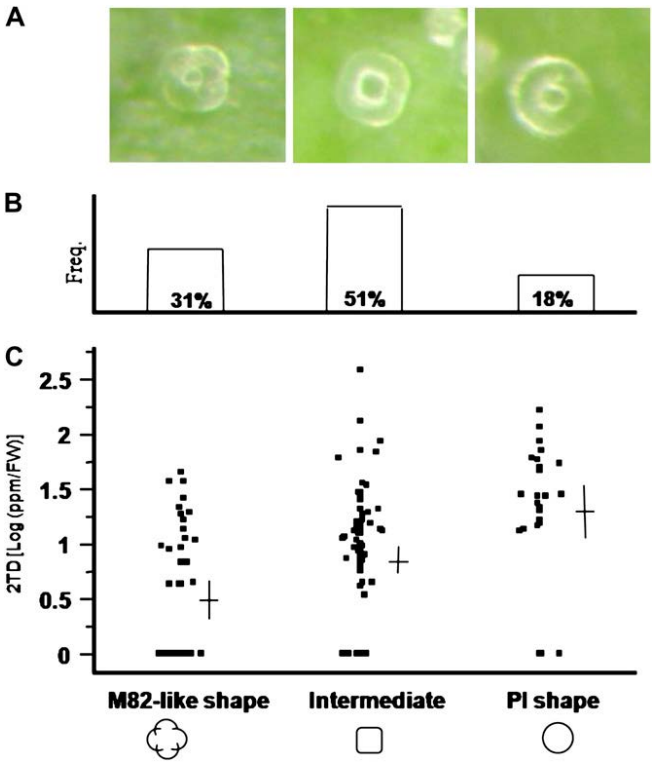

Figure 2-2. Variation and distribution of the gland shape and 2TD content in an interspecific F2 population originated from a cross between the cultivated and wild species. A, A representative binocular image of a type VI glandular trichome on the abaxial surface of a young leaflet from segregating progeny. Plants were categorized as having three types of trichomes based on six independent photographs that were taken from different leaves: M82-like shape, in which the top cells of the trichome are partially separated (left); intermediate shape, in which the cells are merged into a square-like shape (middle); and PI shape, in which the cells are merged into a globular shape (right). B, Distribution of the F2 plants among the three trichome categories. C, Distribution of the 2TD content in the three trichome categories. Horizontal and vertical lines (black crosses) represent the average and 95% confidence limits, respectively. FW, Fresh weight.

Since MKS1 has been previously identified as involved in the MK pathway (Fridman et al., 2005), we looked for a possible relationship between the MKS1 genotype and the shape of the trichomes in the interspecific segregating F2 population. There were significant differences (Pearson test, $P < 0.003$) in the frequencies of the three groups of plants with different type VI trichome shapes among the three genotypes of MKS1. In particular, no F2 progeny exhibited PI-shaped trichomes among homozygotes for the cultivated allele of MKS1 (C/C), and the trichomes of most of the plants in this group bore an M82-like shape (Figure 2-3).

Figure 2-3. Mosaic plot of trichome shape in each of the genotypic classes of MKS1

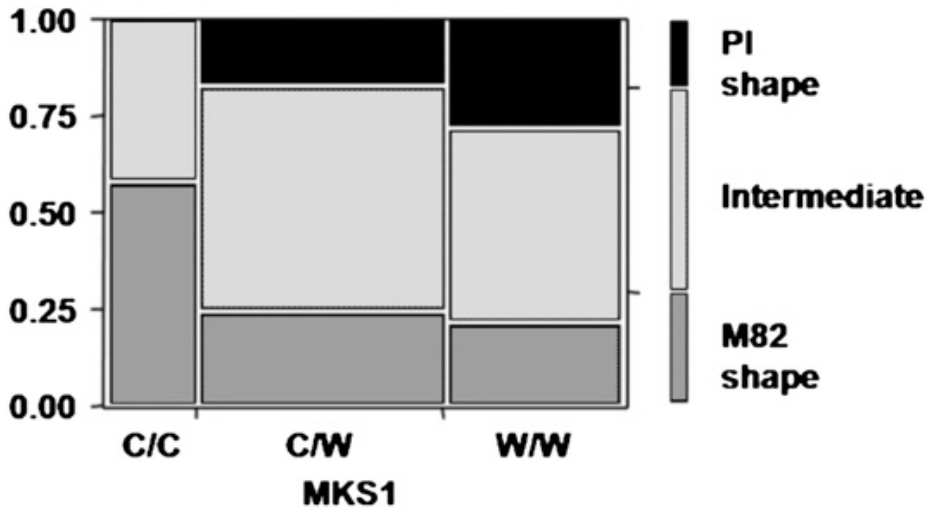

Figure 2-3. Mosaic plot of trichome shape in each of the genotypic classes of MKS1. C, Cultivated (M82) allele; W, wild species (PI) allele.

2.3.2 Association between Candidate Genes, Trichome Characteristics, and MK Content

The association between variation in candidate structural genes and 2TD content was examined in genetic mapping experiments employing these genes as simple PCR markers, cleaved amplified polymorphism sequences, or single-nucleotide polymorphism (SNP) markers, after aligning the open reading frame (ORF) of the alleles from both species and designing amplicons flanking the insertion/deletion, SNP, or restriction site (see 2.6 MATERIALS AND METHODS). The latter approach used high-resolution melt technology (HRM Assay Design and Analysis, CorProtocol 6000, 2006).

Since MK are derived from fatty acids, we examined the following genes from the fatty acid biosynthetic pathway (http://lipids.plantbiology.msu.edu/?q=lipids/genesurvey/)

as genetic markers: acetyl-CoA carboxylase (ACC), malonyl-CoA:ACP transacylase (MaCoA-ACP trans), 3-ketoacyl-ACP synthase III (KASIII), 2,3-trans-enoyl-ACP reductase, 3-ketoacyl-ACP synthase I (KASI), ACP1, and ACP2. The MKS1 locus was also included in the genetic screening. Association test was conducted by multiple regression analysis using the allelic state of the different genes in each F2 progeny and trichome characters as a predictor of 2TD level. This analysis (power = 0.999) showed that segregating progeny carrying the wild allele in two loci (MKS1 and ACC) contain significantly (P , 0.0003 and P , 0.003, respectively) higher amounts of 2TD (Figure 2-4A), while for the MaCoA-ACP trans locus, the opposite trend was observed: plants carrying the wild species allele had, on average, significantly (P , 0.039) less 2TD content. In addition, a significant positive correlation between the density and shape of the trichomes and the amount of MK in the leaves was found (Figure 2-4A). This multiple regression reinforced the previous results, indicating an association between trichome morphology and 2TD levels (Figure 2-2), and overall this model explained approximately one-third of the total 2TD phenotypic variation in the F2 population ($r^2 = 0.333$; Figure 2-4B). To test whether the three candidate genes that are significantly associated with 2TD levels in the segregating population (MKS1, ACC, and MaCoA-ACP trans) exhibit differential expression between the wild and cultivated species, a quantitative reverse transcription (qRT)-PCR approach was taken. Primer pairs that fully matched both alleles were designed for each gene, and qRT-PCR was conducted using RNA from trichomes of both accessions (see 2.6 MATERIALS AND METHODS). MKS1, ACC, and MaCoA-ACP trans showed 355-, 2.7-, and 7.7-fold higher expression, respectively, in the trichomes of the PI parent versus those of the M82 parent (Figure 2-4C).

Figure 2-4. Association study of selected genes with 2-TD content

A

Factor	Effect	Prob>|t|
MKS1	0.51	0.0003
ACC	0.37	0.003
MaCoA:ACP trans	-0.25	0.039
Shape	0.23	0.003
Density	0.04	0.013

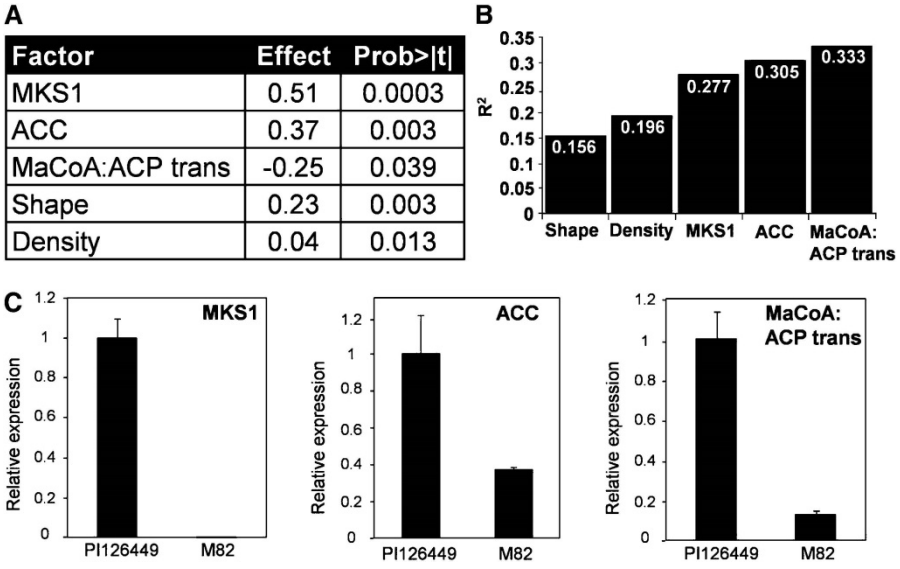

Figure 2-4. Association study of selected genes with 2-TD content. A, Association analysis of candidate genes and trichome characteristics with leaf 2TD content. The multiple regression model includes all of the tested factors for which a significant effect was found. The effect is defined as the contribution of one level/allele of the tested factor to the 2TD content. B, Accumulated variation explained by the model (r^2) with each additional factor. C, Expression of MKS1, ACC, and MaCoA-ACP trans in trichomes of the wild species *S. habrochaites* f. sp. *glabratum* (PI) and the cultivated tomato *S. lycopersicum* (var M82). mRNA abundance in isolated trichomes was determined by qRT-PCR. Expression levels for the various samples were normalized to the expression of *Actin*. Data are averages of three biological replicates, and error bars represent SE.

Mapping these three genes on the tomato genome using the *Solanum pennellii* introgression line population (Eshed and Zamir, 1995; Liu and Zamir, 1999) identified genes ACC and MaCoA-ACP trans on chromosome 1, bin 1-B. Based on the F2 mapping population, the two loci are 8.8 centimorgan (cM) apart (39 recombination events in 220 F2 progeny successfully scored for both loci). *MKS1* was localized to bin 1-I, more than 50 cM away. There were no significant linkages with any of the other loci tested.

2.3.3 Transcriptome Analysis of Trichomes from Bulked Segregants

Bulked segregant analysis (Michelmore et al., 1991) was used to compare transcriptomes of plants with low and high MK contents. Five plants with high levels of 2TD and five plants with no detectable 2TD were selected from the segregating F2 population (total of 245 plants) and propagated for this analysis. RNA from the trichomes of each of these two groups of plants (with high and low MK contents) was extracted, reverse transcribed, labeled, and hybridized to a custom-made microarray containing tomato genes (see 2.6 MATERIALS AND METHODS). A comparison of the hybridization results revealed a number of genes whose transcripts were present at either higher or lower levels in the high-MK-containing plants relative to their low-MK counterparts (Table 2-1). In particular, one wild species-specific transcript of a gene that we subsequently designated MKS2 (see below) was 337-fold more highly expressed in F2 plants with high versus low MK content, while a similar transcript, derived from the cultivated species, was 7.5-fold more highly expressed in the F2 plants with low versus high MK content (Table 2-1).

Table 2-1 Microarray analysis of genes differentially expressed in high- and low-MK bulks

Gene code [b]	Annotation	Ratio [c]
DN167657	Thioesterase family protein (Sh allele)	+336.6
AI779239	rRNA-16S ribosomal RNA	+62.6
AF230371	Allene oxide synthase	+46.7
BI925004	Plasma membrane intrinsic protein	+39
BI931228	Unknown	+27.3
AW616884	Dehydrodolichyl diphosphate synthase	+24.1
DN169296	unknown	+18.3
DB719610	Calcium-binding EF hand family protein	+15.5
AP007224	RuBisCO small subunit 1A	+14.7
DN169129	Major latex protein-related (Sh allele)	+12.9
AW039905	4-coumarate-CoA ligase	+11.8
AI777019	Unknown	+11.2
AW615872	Glycosyltransferase family 14 protein	+11.1
BF097749	mitochondrial 26S ribosomal RNA protein	+9.2
BW688217	Unknown protein	+9.1
BM412813	Methyltransferase family 2 protein	+8.8
DN170232	invertase/pectin methylesterase inhibitor family protein	+8.4
DN171038	ARF GTPase-activating domain-containing protein	+7.2
AY136419	X-Pro dipeptidase	-7.1
BG131749	Thioesterase family protein	-7.5
AI772024	Unknown???	-7.6
DB722221	Unknown	-7.7
	(same as ES893822)	
ES893822	Cell wall protein precursor	-7.9
BG643000	Phospholipase A2 beta	-8.0
BW690350	FH interacting protein 1 (FIP1)	-8.3
AW034502	Cytochrome P450, putative	-8.9
TZEF819	tomato transposon	-9.5
BI928231	NAD-dependent epimerase/dehydratase family protein	-10.1
BI932160	UDP-glucoronosyl/UDP-glucosyl transferase family protein	-10.6
BG128416	Unknown	-10.9
AW624755	Major latex protein related	-11.9
ES896328	kDa chaperonin, chloroplast	-12.5
BF112895	Unknown	-15.2

[a] cDNA sequences from the customized microarry that were upregulated or downregulated by more than 7-fold are shown

[b] Corresponding GenBank accession numbers with the highest similarity are shown

[c] Ratio depicts the difference in average ratios of High-MK over Low-MK bulks in four hybridizations.

2.3.4 MKS2 Shares Sequence Identity with Hotdog-Fold Thioesterases

The MKS2 protein is 52% to 70% identical to several plant proteins with no proven functions encoded by genes in the Arabidopsis (*Arabidopsis thaliana*) and rice (*Oryza sativa*) genomes, by many ESTs from various plant species of the angiosperm family, as well as from white spruce (*Picea glauca*; Figure 2-1). Sequence similarity established that the 149-residue protein encoded by MKS2 is a member of the 4-hydroxybenzoyl-CoA thioesterase (4HBT) subfamily of hotdog-fold enzymes (Dillon and Bateman, 2004). Although only recently discovered, hotdog domains constitute a broad superfamily that is evolutionarily unrelated to the vast *α/β*-hydrolase-fold superfamily of (thio)esterases, of which MKS1 is a member (Hotelier et al., 2004). Notably, several hotdog-fold subfamilies required for typical fatty acid metabolism, including the FabA/FabZ 3-hydroxy-acyl-ACP dehydratases and the FatA/FatB saturated acyl-ACP thioesterases, occur instead as longer sequences that represent a tandem duplication of the hotdog fold (Dillon and Bateman, 2004; Mayer and Shanklin, 2005). BLAST search revealed the MKS2 amino acid sequence to also be 34% identical (and 48% similar) over 131 aligned residues to a structurally characterized putative thioesterase from *Thermus thermophilus* (Protein Data Bank [PDB] code 1Z54), which in turn shares 26% sequence identity and 51% similarity with 4HBTover 91 aligned residues. The crystallized 1Z54 protein backbone nearly perfectly overlays with known 4HBT structure (PDB code 1L09). This “bridging” 1Z54 sequence and crystal structure firmly establish the homology of MKS2 to the well-characterized 4HBT, and 1Z54 facilitates a structural homology based examination of the MKS2 sequence. In addition, MKS2 conserves the catalytic Asp-17 of

4HBT, although our model predicts extensive substitution of juxtaposed residues in the MKS2 active-site cavity relative to any characterized 4HBT subfamily member.

2.3.5 MKS2 Is Associated with MK Content and Reveals an Epistatic Interaction with MKS1

Nucleotide differences were used to employ the *MKS2* gene as a DNA HRM marker (Figure 2-6A) and to investigate the association between the allelic state in this locus and the 2TD content variation in the segregating population. The allelic variation in *MKS2* was significantly associated with 2TD content (P, 0.0001) and ranked as the second-most contributing factor (after *MKS1*) among the loci thus far identified in this quantitative analysis (Figure 2-6B). Moreover, inclusion of this locus in the multiple regression analysis increased the r^2 of the model from 0.333 to 0.485 (Figure 2-6C). Expression analysis by qRT-PCR showed that *MKS2* is 980-fold more highly expressed in the trichomes of the high-MK accumulator PI parent than in those of the M82 parent (Figure 2-6D), similar to MKS1, ACC, and MaCoA-ACP trans (Figure 2-4C).

Figure 2-5. Amino acid sequence alignment of SlMKS2, ShMKS2, and related proteins

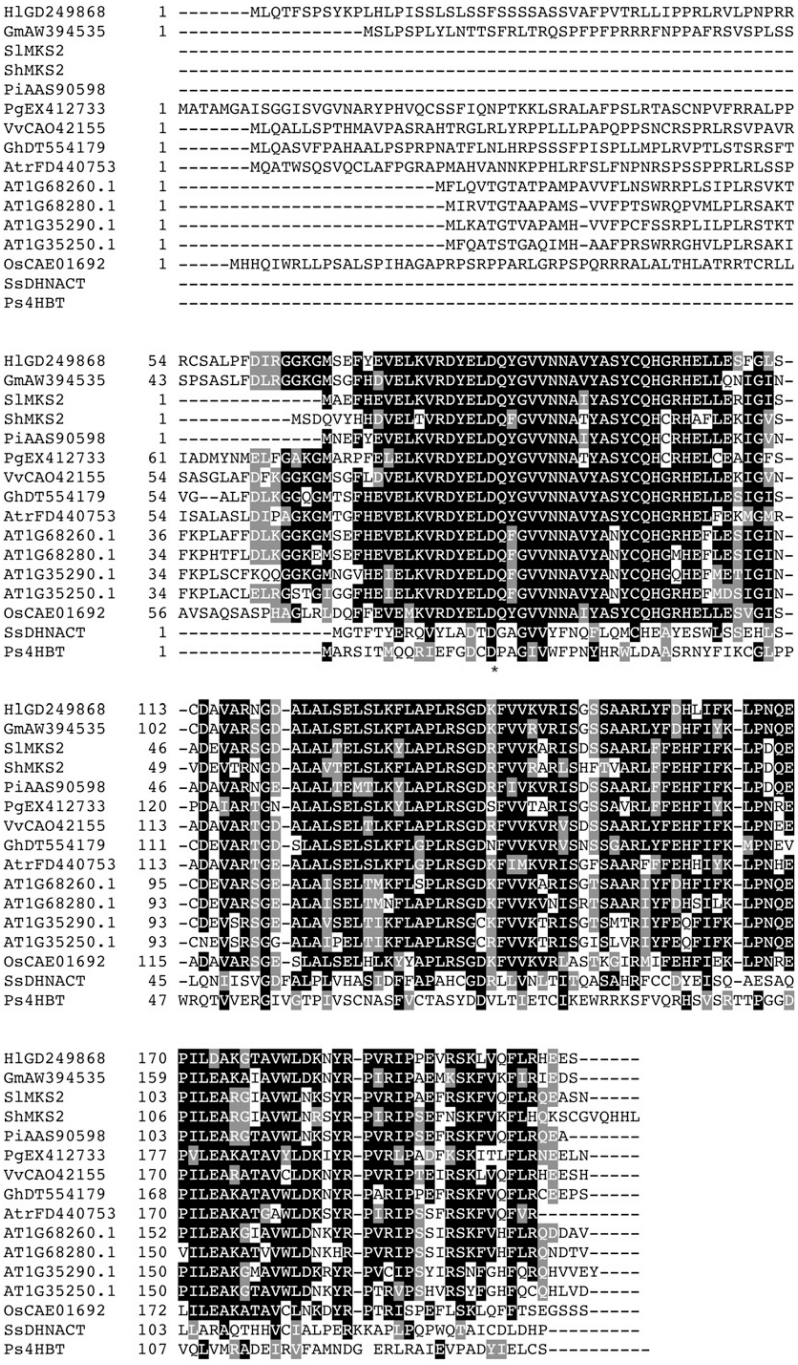

Figure 2-5. Amino acid sequence alignment of SlMKS2, ShMKS2, and related proteins. White letters on a black background indicate identical amino acids in the majority (eight or more) of the sequences. White letters on a gray background indicate conserved amino acid substitutions. The asterisk indicates the catalytic Asp residue identified in the enzyme 4HBT. The complete cDNA sequences of *SlMKS2* and *ShMKS2* indicate that the ORFs begin as indicated in this figure. At, *Arabidopsis thaliana*; Atr, *Amborella trichopoda*; Gh, *Gossypium hirsutum*; Gm, *Glycine max*; Hl, *Humulus lupulus* (cv Phoenix); Os, *Oryza sativa*; Pa, *Prunus armeniaca*; Pi, *Petunia integrifolia* subsp. *inflata*; Pg, *Picea glauca*; Ps, *Pseudomonas sp*. (strain CBS-3); Sh, *Solanum habrochaites*; Sl, *Solanum lycopersicum*; SsDHNACT, *Synechocystis sp.* PCC6803 1,4-dihydroxy-2-naphthoyl-CoA thioesterase; Vv, *Vitis vinifera*. Accession numbers are as follows: Atr, FD440753; Gh, DT554179; Gm, AW394535; Hl, EX521228; Os, CAE01692; Pi,AAS90598; Pg, EX412733; SsDHNACT, NP442358; Vv, CAO42155.

Figure 2-6. Association analysis of *MKS2* and other MK-modulating loci with 2TD levels

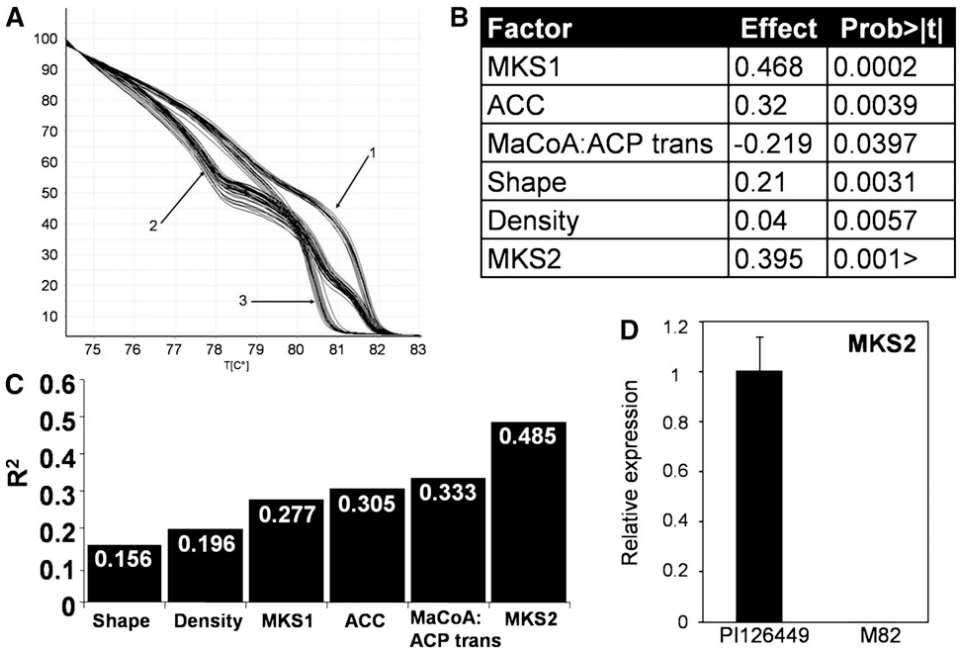

Figure 2-6. Association analysis of *MKS2* and other MK-modulating loci with 2TD levels. A, Allelic distribution at the *MKS2* locus in the F2 population using HRM marker: 1, homozygous for the M82 allele; 2, heterozygous; 3, homozygous for the PI allele. B, Multiple regression analysis for testing the association of candidate genes and trichome characteristics to the 2TD content in the leaves. The model includes all of the tested factors for which a significant effect was found. The effect is defined as the contribution of one level/allele of the tested factor to the 2TD content. C, Accumulated variation explained by the model (r^2) with each additional factor. D, Expression of *MKS2* in trichomes of the wild species *S. habrochaites* f. sp. *glabratum* (PI) and the cultivated tomato *S. lycopersicum* (var M82). mRNA abundance in isolated trichomes was determined by qRT-PCR. Expression levels were normalized to the expression of *Actin*. Data are averages of three biological replicates, and error bars represent SE.

In an attempt to define possible epistatic interactions between the different genetic components of the MK network, the genetic factors that significantly contribute to MK variation in the test population were evaluated for possible two-way interactions. This analysis identified a single significant interaction between the *MKS2* and *MKS1* loci (Figure 2-7A). The data showed that to achieve high levels of these compounds, the plant has to carry at least one wild species allele in each of these two interacting loci. While *MKS1* shows a dominant mode of inheritance, that of *MKS2* is only partially so: all three genotypic classes differ significantly. Indeed, incorporation of the interaction between the two loci into a new regression model by grouping the plants according to the two-locus haplotypes (Figure 2-7B) increased the r^2 of the model from 0.485 to 0.545 (Figure 2-7C).

Figure 2-7. Genetic interaction of the *MKS1* and *MKS2* loci

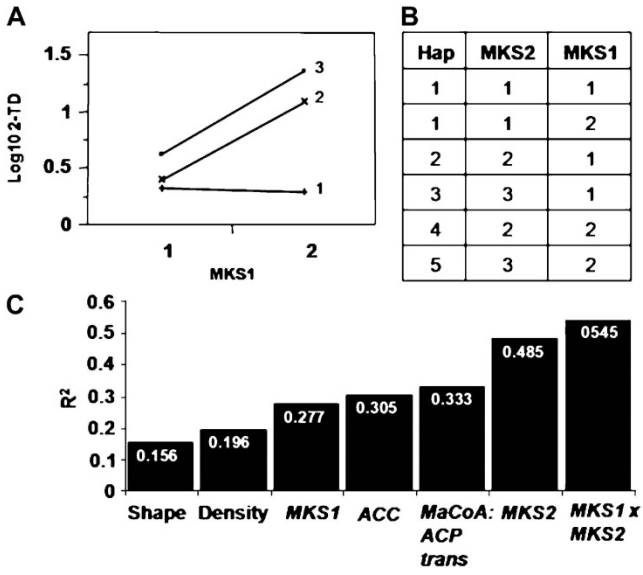

Figure 2-7. Genetic interaction of the *MKS1* and *MKS2* loci. A, 2TD least-square means lot. The x axis represents the genotypes of the *MKS1* locus: 1, homozygous for the M82 allele; 2, heterozygous or homozygous for the PI allele. The lines represent different genotypes of the *MKS2* locus: 1, homozygous for the M82 allele; 2, heterozygous; 3, homozygous for the PI allele. B, Levels of the haplotype (Hap) factor. *MKS1*: 1, homozygous for the M82 allele; 2, heterozygous or homozygous for the PI allele. *MKS2*: 1, homozygous for the cultivated allele; 2, heterozygous; 3, homozygous for the PI allele. C, Accumulated variation explained by the model (r^2) with each additional factor.

2.3.6 Heterologous Expression of *MKS2* in *Escherichia coli*

To investigate the biochemical activity of MKS2, the full ORFs of the wild species allele (*ShMKS2*) and the cultivated allele (*SlMKS2*) were amplified and ligated into an *E. coli* expression vector (see 2.6 MATERIALS AND METHODS). These vectors were introduced into *E. coli* BL21 cells, and *ShMKS2* or *SlMKS2* expression was induced by the addition of isopropylthio-*β*-galactoside (IPTG; see 2.6 MATERIALS AND METHODS). After induction and overnight growth, the culture was analyzed by solid-phase microextraction of its headspace followed by gas chromatography-mass spectrometry (GC-MS; see 2.6 MATERIALS AND METHODS). The major compound

in the headspace of the *E. coli* cells expressing ShMKS2 was identified as 2TD (Figure 2-8A). Lower amounts of 2UD and 2-pentadecanone were also detected, as well as the reduced alcohol forms of 2UD and 2TD (i.e. 2-undecanol and 2-tridecanol). The headspace of the *E. coli* cells expressing *SlMKS2* contained 2UD as well as 2-nonanone as the two main MK and only trace amounts of 2TD. The headspace also contained 2-nonanol and 2-undecanol (Figure 2-8B). However, the major headspace compound produced by *SlMKS2* expressing cells eluted slightly later than 2TD (peak 1 in Figure 2-8B and also present at lower levels in the chromatograph in Figure 2-8A). MS analysis suggested that it is a 2-tridecenone, but the position of the double bond has not yet been determined.

Figure 2-8. GC-MS analysis of volatile compounds produced in *E. coli* when *ShMKS2* (A) or *SlMKS2* (B) is expressed

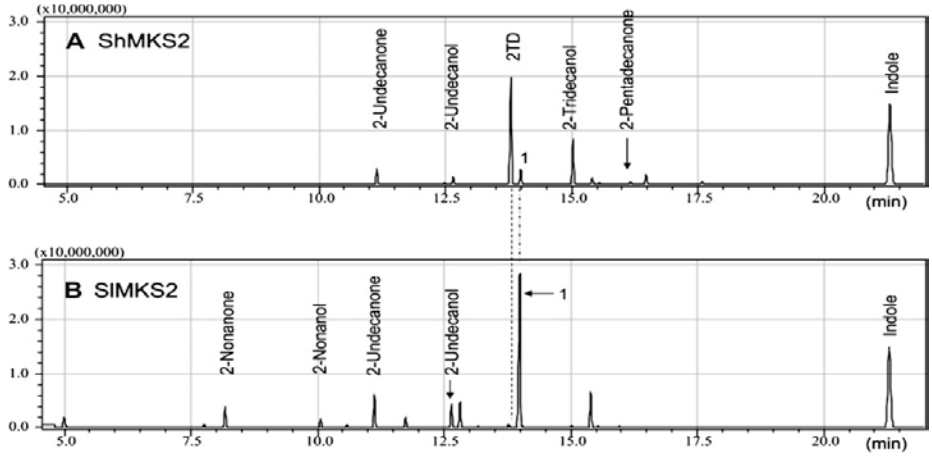

Figure 2-8. GC-MS analysis of volatile compounds produced in *E. coli* when *ShMKS2* (A) or *SlMKS2* (B) is expressed. Peak 1 is 2-tridecenone, but the position of the double bond has not yet been determined.

2.4 DISCUSSION

2.4.1 Developmental and Biochemical Connection in MK Synthesis

One of the most surprising findings of this study was the tight relationship between the shape of the trichomes and MK content (Figure 2-2). The round and globular trichome shape of the wild species and its progeny was significantly associated with higher MK content. While this observation suggests that morphology constitutes a general barrier to accumulation of volatile compounds, analysis of other volatile compounds in the F2 population did not support this. For example, the distribution of one of the other major volatiles in the glandular trichomes of PI, β-caryophellene, was not correlated with trichome shape. Another possible explanation is that since cuticular waxes are complex mixtures of C20-C34 straight-chain aliphatics derived from very-long-chain fatty acids (Jetter and Kunst, 2008), the diversion of fatty acid pool toward MK comes at the expense of cuticle biosynthesis. This is also supported by the three-way relationship between MK content, trichome shape, and the genotype of *MKS1* in the segregating population (Figure 2-2 and 2-3). However, we cannot reject the possibility that the connection between *MKS1* variation and trichome shape might be due to genetic linkage with a gene(s) that modulates the development of this specialized organ. The globular shape of the wild species trichome may be comparable to the “fused” organ morphology seen in mutants with defective cuticles (Sinha and Lynch, 1998). These fusion phenotypes have been associated with defects in several genes that modulate the biosynthesis and deposition of very-long-chain fatty acids, including enzymes (Efremova et al., 2004) and transporters (Bird et al., 2007). Together, these observations suggest a model in which enhanced activity of MK biosynthesis in the wild species may underlie

the diversion of the fatty acids to MK at the expense of the synthesis of very-long-chain fatty acids, hence changing the morphology of the trichomes.

2.4.2 The Genetic Basis for MK Biosynthesis in *S. habrochaites* f. sp. *glabratum*

Although MK have been found in several plant lineages, their occurrence in *S. habrochaites* f. sp. *glabratum* is unique in the *Solanum* genus, suggesting a monophyletic evolution of the specialized metabolic pathway in this subspecies. This study was aimed at identifying the genetic network required for the operation of this pathway within a single cell type, the glandular trichome. The quantitative mode of inheritance of MK in the F2 population (Figure 2-1) clearly indicates that several genes are involved in the biosynthesis of these compounds and that most of the wild species alleles are recessive, as reflected in the backcrossed population (Figure 2-1).

The multiple regression analysis showed that the *S. habrochaites* f. sp. *glabratum* alleles of genes encoding the first enzyme in the fatty acid biosynthetic pathway, ACC, as well as the enzyme MKS1 are both positively associated with MK biosynthesis. In contrast, other genes encoding enzymes that catalyze intermediate steps did not show this positive correlation. The analysis used in this study for associating candidate genes with MK variation has a few shortcomings that stem from the population structure and the lack of additional genetic markers. The availability of additional markers would have strengthened the association of the variation of MK with the candidate genes and reduced the possibility that such associations are the result of linkage disequilibrium with other causative linked and nonlinked loci. The facts that all of the genes that were included in the multiple regression analysis were also differentially expressed in the two species (Figure 2-4C and 2-6D) and that the F2 population genotyped and phenotyped in this

study is relatively large support the conclusion that a major portion of the MK variation observed in this interspecific population can indeed be attributed to diversity in these genes rather than to other genes that may be in linkage disequilibrium with these candidates. Overall, it appears that the flux in the MK pathway is controlled at the gene expression level and that the alleles from both species encode almost identical proteins that are therefore likely to be equally active.

Interestingly, the wild-type allele of the gene encoding MaCoA-ACP trans, the enzyme that acts immediately after ACC, was inversely associated with MK content. The relative contribution of this locus to the chemical variation was very low (Figure 2-6), and our genetic analysis indicated that the genes encoding these two enzymes are tightly linked (8.8 cM) on chromosome 1. Since these loci act in repulsion (i.e. in the first locus the wild species allele increases MK content and in the other locus the wild species allele reduces it), the results depicted in Figure 2-6 are somewhat biased. The magnitude of the positive and negative additive effects of *ACC* and *MaCoA-ACP* trans loci on MK content is likely to be higher due to linkage drag, which is not included in a single-point analysis such as that conducted in this study.

The combination of a classical genetic approach, bulked segregant analysis, (Michelmore et al., 1991); and transcriptome analysis of the glandular trichomes led to the discovery of a new participant in MK accumulation in these specialized cells (MKS2). Interestingly, the microarray transcriptome analysis did not detect differences in *MKS1* expression levels between bulked high- and low-MK-containing F2 plants. Genotyping the individual members of the two groups of five plants explains this unexpected result: four plants from the low-MK bulk carried the wild allele at this locus (*ShMKS1*) and

three of them were homozygous, giving a total dosage of seven *ShMKS1* alleles. Similar dosages of *ShMKS1* alleles were found in the high-MK bulk in two heterozygous and three homozygous plants, giving a total of eight alleles and leading to equivalent transcripts in both bulks. Similarly, plants that accumulate no MK showed high levels of MKS1 protein in immunoblot tests of an F2 population segregating for MK content (Fridman et al., 2005). This indicates that the regulation of MKS1 is in *cis* rather than in *trans* (expression controlled by the locus itself and not by other unlinked factors). This conclusion is also supported by the fact that the wild species' promoter can drive high levels of GFP and GUS expression in glands of the cultivated tomato (E. Fridman and R.C. Schuurink, unpublished data). In addition, these plants provided a specific and strong demonstration of the epistatic relationship occurring between *MKS1* and *MKS2* (Figure 2-7). Four out of five low-MK plants that carried the *ShMKS1* allele were found to be homozygous for the cultivated allele at the *MKS2* locus (*SlMKS2*), thus lacking the wild species allele ShMKS2. The fifth plant, on the other hand, presented the opposite pattern (i.e. heterozygous at the *MKS2* locus but homozygous for the cultivated allele at the *MKS1* locus, *SlMKS1*), thus lacking the wild species allele *ShMKS1*. Conversely, analysis of the two-locus haplotype in the high-MK bulk plants showed that they carried at least one wild allele at both the *MKS1*and *MKS2* loci (*ShMKS1* and *ShMKS2*).

2.4.3 The Role of *MKS1* and *MKS2* in MK Biosynthesis

The protein with the highest sequence similarity to MKS2 that has established enzymatic activity is 4-hydroxybenzoyl-CoA thioesterase from *Pseudomonas sp.* strain CBS3. Indeed, the 1Z54 and 4HBT crystal structures reveal similar homotetrameric assemblies, which our MKS2 model also reflects. The conservation of the 4HBT catalytic

Asp-17 by ShMKS2 and SlMKS2 suggests that they are likely also thioesterases. Moreover, the production of MK in *E. coli* cells expressing either allele of *MKS2* is a strong indication that the heterologous MKS2 enzyme may be capable of hydrolyzing (Figure 2-9, step I) and perhaps also decarboxylating (Figure 2-9, step II) 3-ketoacyl intermediates, analogous to the reaction catalyzed by MKS1. However, the production of MK in *E. coli* expressing *MKS2* is not informative with regard to the specific substrates 3-ketoacyl-ACPs or 3-ketoacyl-CoA, because *E. coli* cells produce both types of substrates.

Figure 2-9. Illustration of the hydrolysis (I) and decarboxylation (II) steps that mediate MK biosynthesis from 3-ketoacyl intermediates

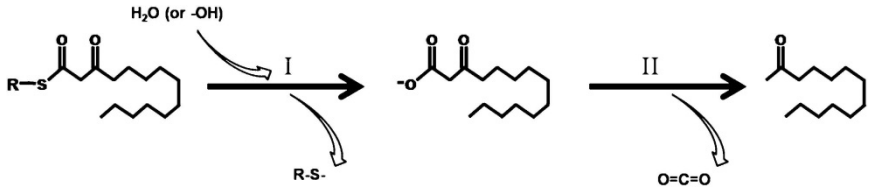

Figure 2-9. Illustration of the hydrolysis (I) and decarboxylation (II) steps that mediate MK biosynthesis from 3-ketoacyl intermediates. R represents either ACP or CoA.

Proteins with high levels of identity to tomato MKS2 are found throughout the plant kingdom, but interestingly, all such sequences outside Solanaceae contain an N-terminal extension predicted to be a plastid or mitochondrial transit sequence (Fig. 5). SlMKS2 and ShMKS2 (and also a petunia [*Petunia integrifolia*] MKS2 homolog) lack such a transit peptide, raising the possibility that these Solanaceae proteins are not localized in the plastids and that their substrates may therefore not be 3-ketoacyl-ACPs but rather 3-ketoacyl-CoAs. The MKS2 proteins, however, do not contain any other obvious subcellular targeting sequences (e.g. no obvious PTS1 or PTS2 sequences that would target the protein to the peroxisomes).

The presence of two distinct enzymes that contribute to the production of the same compound in the same organ, and even in the same cell, is not unprecedented, and such functional redundancy was recently reported for eugenol biosynthesis in Clarkia (Koeduka et al., 2009). In the case of MKS1 and MKS2, however, genetic evidence for epistatic interactions between the two loci suggests that they do not act independently of each other and therefore raises some questions. Do MKS1 and MKS2 form a complex? If such a complex is formed, does each type of subunit carry out both reactions of thioester bond hydrolysis and decarboxylation (Figure 2-9, steps I and II), or does each subunit catalyze only one of these reactions?

Alternatively, epistatic interactions may indicate not a physical interaction but that they act sequentially in the pathway from 3-ketoacyl intermediates to MK. A closer analysis of the genetic data reveals that although the two wild species alleles in *MKS1* and *MKS2* loci are required for the accumulation of high 2TD levels in tomato, some levels are nevertheless found in plants that carry only the *MKS2* wild allele (*ShMKS2*) but not vice versa (Figure 2-7). These results suggest a model for MK biosynthesis in the trichomes in which MKS2 works upstream of MKS1. By this model, MKS2 hydrolyzes the 3-ketoacyl intermediates (Figure 2-9, step I) and a low level of spontaneous decarboxylation (Figure 2-9, step II) can occur to produce MK (Kornberg et al., 1948), a step that is sped up by MKS1 when present.

A resolution between these competing hypotheses will require a determination of the subcellular localization of the MKS2 protein, whether it physically interacts with MKS1 and the substrate it acts on, either independently or in complex with MKS1. The

results of these experiments may in turn require a reassessment of the subcellular localization of MKS1 as well as its substrate specificity.

2.5 CONCLUSION

The above results present the complex monophyletic evolution of a specialized pathway and highlight the power of incorporating morphological and chemical data for a detailed understanding of pathways that appear to be isolated in specialized cells. The combined data provide a framework for determining the molecular and biochemical bases for the unexpected relationships between shape and content of the glandular trichomes. Moreover, the genetic and biochemical relationship between *MKS1* and newly identified *MKS2* loci highlights the major role of epistasis interactions in determining phenotypic variation among populations and emphasizes the importance of taking it into account when dissecting the genetic basis of complex phenotypes.

2.6 MATERIALS AND METHODS

2.6.1 Plant Material, and Interspecific F2 and Backcross Populations

Solanum lycopersicum var M82 indeterminate and *Solanum habrochaites* f. sp. *glabratum* (PI) were obtained from the Tomato Seed Stock Center at the University of California and from the U.S. Department of Agriculture Agricultural Research Service. A single PI plant served as the male to fertilize *S. lycopersicum*. The hybrids (1) were selfed to obtain the F2 population, (2) served as the male for fertilizing M82 to obtain BC1-M82, or (3) served as the female for a PI male to obtain BC1-PI. Seeds were sprouted in trays for 2 d in a closed room at 25 ℃ and 95% humidity and were then grown in an open greenhouse for 3 weeks. Seedlings were transplanted to the greenhouse, trellised with

ropes, and grown in red loam soil with 1 m3 of water and 50 mL of fertilizer (Shefer; ICL Fertilizer) per day. For bulk analysis, F2 plants were propagated by cuttings using rooting powder with 0.3% indole-3-butyric acid. Cuttings were rooted in germination trays held under spraying water for 0.5 h twice a day.

2.6.2 Volatile Analysis

Six young leaflets (the first, second, and third from the first or second leaves) were sampled into scintillation vials on ice, and volatiles were extracted and analyzed as described by Fridman et al. (2005).

2.6.3 Morphology Indexes

Six young leaflets (opposite those taken for volatile analysis) were sampled into scintillation vials, and a digital photograph of the central upper surface was taken. Mean trichome number per square millimeter was calculated, and trichome shape was classified as follows: wild shape (PI shape), intermediate shape (intermediate), and cultivated-like shape (M82-like shape).

2.6.4 Genotyping

DNA samples were extracted from approximately 100 mg of fresh young tomato leaves and buds following the protocol described by Murray and Thompson (1980). See Supplemental Table S1 for PCR conditions and primers. KASI PCR products (15 mL) were digested with 1 mL of Taq restriction enzyme (New England Biolabs) for 1 h at 65 ^{o}C in a reaction that included 2 mL of 103 buffer and 10 mg of bovine serum albumin (20 mL total).

2.6.5 HRM Genotyping

Sequences were aligned using the Align function in the Vector NTI software package (Invitrogen) to identify SNPs between the sequences of the *S. lycopersicum* and *S. habrochaites* alleles. The *S. habrochaites* sequences were taken from an EST library produced from the glandular trichomes of PI, and the *S. lycopersicum* alleles were retrieved from the total tomato EST repository (SOL database; http://www.sgn.cornell.edu/index.pl). Three different pairs of primers flanking the identified SNPs (amplicon size varied from 60 to 100 bp per SNP) were selected for each gene using the Primer3 software (http://primer3.sourceforge.net). First, PCR was conducted with a test panel that included the parental lines M82 and PI and their hybrid (F1). PCR products were analyzed on an agarose gel (3%), and reactions that produced a single product with no primer dimers were selected for HRM analysis on a Rotor-Gene 6000 (Corbett Research). Primers that showed the best allelic discrimination by HRM examination were selected to score the genotype of the F2 population. HRM was performed immediately after the PCR cycles as a single run following themanufacturer's default parameters.

2.6.6 Transcriptome Analysis

Trichome isolation was performed as described by Fridman et al. (2005). The tomato microarray design, cDNA synthesis, hybridization, and analysis were performed by Genotypic Technology. The microarray was a complex of 44,000 probes of 25 bp each representing all of the tomato ESTs (Tomato Gene Index; http://compbio.dfci.harvard.edu/tgi/cgi-bin/tgi/gimain.pl?gudb=tomato). Total trichome RNA (5 mg) was labeled with Cy3 and Cy5, and hybridization was repeated four times

(two repeats for each dye swap between RNA samples) in a 4 3 44 format following the version 5.5.X protocol for a two-color array. Results were analyzed following the GE-v5_95_Feb07 protocol (Agilent). More details can be found at the National Center for Biotechnology Information Gene Expression Omnibus under GSE16431.

2.6.7 qRT-PCR

Total RNA was isolated from isolated glandular trichomes as described previously (Fridman et al., 2005). The RNAwas subjected to DNase treatment using a DNA-free kit (Ambion), and first-strand cDNA was synthesized by SuperScript II reverse transcriptase (Invitrogen) with poly-T primers in parallel with a negative control reaction in which no SuperScript II reverse transcriptase was added. qPCR utilizing power SYBR-Green PCR master mix (Applied Biosystems), gene-specific primers, and a dilution series of each cDNAwas performed as described previously (Varbanova et al., 2007). qPCR was performed using the StepOnePlus Real-Time PCR System (Applied Biosystems), and the conditions were as follows: 95 ℃ for 3 min, 50 cycles of 95 ℃ for 15 s, 60 ℃ for 30 s, and 72 ℃ for 30 s, followed by a melting cycle of 55 ℃ to 95 ℃ with an increasing gradient of 0.5 ℃ and a 10-s pause at each temperature. All reactions were performed in triplicate, and each experiment was repeated twice. ShMKS2 and SlMKS2 allele-specific primers were designed as follows: *ShMKS2* forward, 5'-GCCTATATTGGAGGCAAGAGGA-3', and *ShMKS2* reverse, 5'-TGTACACCGCAACTCTTCTGGT-3'; *SlMKS2* forward, 5'-ATGCAAGTTATTGCCAACATGG-3', and *SlMKS2* reverse, 5'-GAAAAACAAACGAGCAGCTGAA-3'; *ACC* forward, 5'-CTGCTAGGAAAGCTCATCGTATGG-3', and *ACC* reverse, 5'-

GTGGTAGGAACTCCAGTGATAACG-3'; *MaCoA-ACP* trans forward, 5'-GAATGACGGTACGTCTAGCTGTTG-3', and *MaCoA-ACP* trans reverse, 5'-GGTGAAGTCACCTGGCTAGCTAAT-3'. *Actin* transcript amplification was used as an internal control, with the forward primer 5'-AACACCCTGTTCTCCTGACTGA-3'and the reverse primer 5'-AACACCATCACCAGAGTCCAAC-3'.

2.6.8 Sequence Analysis

Alignment of multiple protein sequences was performed using the ClustalW program (Thompson et al., 1997).

2.6.9 Statistical Analysis

Statistical analyses were conducted with JMP software (SAS Institute). Since phenotypic data of the segregating F2 and backcrossed plants did not fit the normal distribution, they were log transformed. A nonparametric test (Wilcoxon) was used to test the MKS1 genotype effect on 2TD levels of 221 plants under the Fit Y by X function (because of unequal variances). The association between trichome shape and the 2TD levels was tested for 164 plants by ANOVA under the Fit Y by X function. For these two tests, the tested factor was set as a character and the 2TD levels were continuous. Association between the MKS1 locus genotype and trichome shape was tested in 134 plants by Pearson test for category parameters under the Fit Y by X function. Data of 122 individuals were used for multiple regression analysis that performed by choosing the "stepwise" option in the Fit model function, and the "forward" direction was used for building the final regression model. All factors included in the analysis were set as continuous. Power was calculated with G*Power 3.1.0 software (Faul et al., 2007). Regression analysis, including the MKS1 and MKS2 interaction, was performed by

replacing these two singular factors with a new factor representing the haplotype at those loci. Interactions between genes were tested by two-way ANOVA under the Fit model function.

2.6.10 Isolation of Full-Length *ShMKS2* and *SlMKS2* cDNAs and Expression in *E. coli*

The following primers were used to amplify the full ORF of MKS2 from PI (*ShMKS2*) orM82 (*SlMKS2*) leaf cDNA into the TA cloning vector (pCRT7/CT TOPO-TA; Invitrogen): forward, 5'-ATGAGTGATCAGGTCTATCACC-3'; reverse, 5'-CTCTTGATCTGGAAGCTTGA-3'. The sequences of these cDNAs were verified and transferred into the *E. coli* expression vector pET28 (Novagene). The pET28 vectors carrying ShMKS2 or SlMKS2 were mobilized into *E. coli* BL21(DES) cells, and gene expression was induced by the addition of 2 mM IPTG after the culture optical density at 595 nm value had reached 0.6 to 0.7. After IPTG addition, *ShMKS2*- and *SlMKS2*-expressing bacterial cells were grown at 30 ℃ or 18 ℃ overnight.

2.6.11 Headspace Analysis of Spent Media of *E. coli* Cultures Expressing *ShMKS2* and *SlMKS2*

After induction with IPTG and growth overnight, 1 mL of culture was placed in a glass vial at 42 ℃. The vial was capped with a screw cap in which a small hole had been bored. The needle of a solid-phase microextraction device was inserted into the vial through the hole in the cap and the fiber was extended for 30 min for volatile collection, after which the fiber was withdrawn and then injected into the GC device. GC-MS analysis was performed as described previously (Fridman et al., 2005). Labeled peaks in

Figure 2-8 were identified by comparison of retention time and MS of authentic standards (MK) or MS and Kovac indices (alcohols).

2.6.12 Homology Modeling

The MKS2 homology model was constructed using MODELLER (Sali and Blundell, 1993), and the illustration was prepared using MOLSCRIPT (Kraulis, 1991), with final rendering by POV-Ray (Persistence of Vision Ray tracer; available at www.povray.org).

Sequence data from this article have been deposited with the GenBank data library under the following accession numbers: *ShMKS2* from *S. habrochaites* f. sp. *glabratum* (accession PI126449), EU883793; *ShMKS2* from *S. lycopersicum* (var M82), EU908050. The Gene Expression Omnibus accession number for raw microarray data and platform description is GSE16431.

Chapter 3 Enzymatic functions of wild tomato methylketone synthases 1 and 2[2]

3.1 ABSTRACT

The trichomes of the wild tomato species *Solanum habrochaites* subsp. *glabratum* synthesize and store high levels of methylketones, primarily 2-tridecanone and 2-undecanone, that protect the plants against various herbivorous insects. Previously, we identified cDNAs encoding two proteins necessary for methylketone biosynthesis, designated methylketone synthase 1 (ShMKS1) and ShMKS2. Here we report isolation of genomic sequences encoding ShMKS1 and ShMKS2 as well as the homologous genes from the cultivated tomato, *S. lycopersicum*. We show that a full-length transcript of *ShMKS2* encodes a protein that is localized in the plastids. By expressing *ShMKS1* and *ShMKS2* in *E. coli* and analyzing the products formed, as well as by performing *in vitro* assays with both ShMKS1and ShMKS2, we conclude that the ShMKS2 protein acts as a thioesterase hydrolyzing 3-ketoacyl-ACPs (plastid localized intermediates of fatty acid biosynthesis) to release 3-ketoacids, and that ShMKS1 subsequently catalyzes the decarboxylation of these liberated 3-ketoacids, forming the methylketone products. Genes encoding proteins with high similarity to ShMKS2, a member of the “hot-dog fold”

[2] This work has been published as Geng Yu, Thuong T.H. Nguyen, Yongxia Guo, Ines Schauvinhold, Michele E. Auldridge, Nazmul Bhuiyan, Imri Ben-Israel, Yoko Iijima, Eyal Fridman, Joseph P. Noel, and Eran Pichersky (2010) Enzymatic Functions of Wild Tomato *Solanum habrochaites glabratum* Methylketone Synthases 1 and 2. Plant physiol. 154, 67-77. My contribution was to get the genomic information the homologs of *ShMKS1* and *ShMKS2* in cultivated tomato, verify their gene structures, transcription start site of *ShMKS1d* and subcellular localization of ShMKS2 as well as study the methylketone production in *E. coli* expressing *ShMSK2*.

protein family that is known to encompass other thioesterases in non-plant organisms, are present in plant species outside the genus *Solanum*. We show that a related enzyme, from *Arabidopsis thaliana*, also produces 3-ketoacids when recombinantly expressed in *E. coli*. Thus, the thioesterase activity of proteins in this family appears to be ancient. In contrast, the 3-ketoacid decarboxylase activity of ShMKS1, which belongs to the *α/β*-hydrolase fold superfamily, appears to have emerged more recently, possibly within the genus *Solanum*.

3.2 INTRODUCTION

Many plants develop glandular trichomes, or appendages, on their aerial parts that synthesize and store specialized (secondary) metabolites involved in plant defense (Schilmiller et al., 2008). Plants in the Solanaceae family exhibit a particularly wide range of different types of glandular trichomes (Luckwill, 1943), each with its own repertoire of specialized compounds that also varies across species, and this chemodiversity is particularly pronounced in the genus *Solanum* (Schilmiller et al., 2010a). For example, Type VI glands in *Solanum lycopersicum* (cultivated tomato) produce mostly terpenes (Schilmiller et al., 2009), while the Type VI glands of *S. habrochaites* subspecies *glabratum* produce high levels of methylketones (up to 8 mg/g leaf fresh weight) consisting mostly of 2-tridecanone and 2-undecanone (Williams et al., 1980; Antonious, 2001; Fridman et al., 2005).

The biosynthetic pathway to methylketones has only recently begun to be investigated (Fridman et al., 2005; Ben-Israel et al., 2009). It is well established that 3-ketoacids are somewhat unstable and can readily undergo decarboxylation when subjected to high temperature and/or non-physiological pH values; a low-level

spontaneous decarboxylation occurs under milder conditions (Kornberg et al., 1948). Decarboxylation of 3-keto fatty acids could thus give rise to straight-chain methylketones such as those found in the *S. habrochaites* glands (Figure 3-1). In plants, 3-keto fatty acids could themselves be derived from the hydrolysis of either 3-ketoacyl-ACPs, which are intermediates in the fatty acid biosynthetic pathway of chloroplasts, or could be derived from 3-ketoacyl-CoAs, which are intermediates in the degradation of fatty acids in the peroxisomes see Figure 3-1 (Buchanan et al., 2000).

Figure 3-1. A schematic reaction sequence for the synthesis of straight-chain methylketones

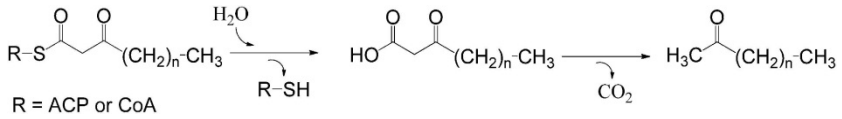

Figure 3-1. A schematic reaction sequence for the synthesis of straight-chain methylketones. 3-Ketoacyl-ACP or 3-ketoacyl-CoA intermediates of fatty acid synthesis and degradation, respectively, are first hydrolyzed, and the resulting 3-ketoacids are then decarboxylated to give the corresponding 2-methylketone.

Initial analysis of a Type VI-specific EST database from a methylketone-producing line of *S. habrochaites glabratum* (accession PI126449) for highly expressed genes, followed by comparative gene expression analysis (using *S. habrochaites* accessions with varying amounts of methylketones), identified the gene methylketone synthase 1 (*ShMKS1*) whose expression level positively correlated with high levels of methylketone formation (Fridman et al., 2005). The 265-residue long protein encoded by *ShMKS1* belongs to the *α/β*-hydrolase superfamily of proteins (Hotelier et al., 2004; Forouhar et al., 2005; Yang et al., 2008). Although ShMKS1 does not have a cleavable N-terminal transit peptide, chloroplast import experiments indicated that it could be

transported into this organelle (Fridman et al., 2005). Initially, using *in vitro* biochemical assays, recombinant ShMKS1 appeared to catalyze the conversion of 3-ketomyristoyl-ACP, an intermediate in fatty acid biosynthesis in the chloroplasts, to 2-tridecanone, suggesting that ShMKS1 possesses both thioesterase and decarboxylase activities that sequentially remove the ACP moiety and decarboxylate the 3-ketomyristic acid intermediate (Fridman et al., 2005). However, it was noted that the *in vitro* rate of production of 2-tridecanone from 3-ketomyristoyl-ACP using ShMKS1 was extremely slow (Fridman et al., 2005).

More recently, extensive genetic and genomic analyses have identified additional genes responsible for the high-level production of methylketones in *Solanum habrochaites* (Ben-Israel et al., 2009). These results validated earlier genetic analysis that concluded that methylketone production had a multilocus basis, which explains why it has proven difficult to breed cultivated tomato lines that produce high levels of methylketones in their trichomes (Zamir et al., 1984). Some of the loci identified encode fatty acid biosynthetic enzymes, a result that is consistent with the need to increase the flux in fatty acid anabolism that provides, directly or indirectly, the substrates for methylketone biosynthesis. Another locus, designated *MKS2*, identified in that study encodes a protein with homology (but <15% identity) to a 4-hydroxybenzoyl-CoA thioesterase (4HBT), a protein belonging to the “hot-dog fold” family, from a *Pseudomonas* bacterium (Benning et al., 1998). Our analysis indicated that high level expression of the *S. habrochaites glabratum* gene, *ShMKS2*, in the glandular trichomes was required for high level production of methylketones (Ben-Israel et al., 2009).

Evolutionarily related proteins are encoded in the genomes of various plants, but no functions have yet been assigned to any such plant proteins (Ben-Israel et al., 2009).

Extensive genetic analysis of an interspecific F2 population between the cultivated and wild species displayed significant epistatic interaction between the MKS1 and MKS2 loci. Plants lacking the *ShMKS2* allele failed to accumulate any methylketones regardless of the allelic state of the MKS1 locus, while absence of the *ShMKS1* allele resulted in significantly reduced levels of methylketones. This raised the possibility that the MKS2 protein acts upstream of MKS1 in the pathway for methylketone biosynthesis (Ben-Israel et al., 2009). Furthermore, expression of *ShMKS2* cDNA in *E. coli* cells resulted in the production of 2-tridecanone, 2-undecanone, and several other methylketones (Ben-Israel et al., 2009). These genetic and biochemical observations raised the question as to what specific catalytic role ShMKS2 plays in the biosynthesis of methylketones in wild tomato trichomes, and whether it works in parallel or in tandem with ShMKS1. Here we show that ShMKS2 catalyzes the hydrolysis of the 3-keto acyl-ACP thioester bond, and ShMKS1 catalyzes the subsequent decarboxylation of the released 3-keto fatty acid, during methylketone biosynthesis.

3.3 RESULTS

3.3.1 Genes encoding MKS1 in *Solanum lycopersicum* and *S. habrochaites glabratum*

Data mining of the genomic "scaffolds" of *S. lycopersicum* (http://solgenomics.net/) indicated that its genome includes at least four genes on two scaffolds, 05390 and 05477, encoding proteins of 264-283 amino acids in length that are

>75% identical to ShMKS1. We designated these genes *SlMKS1a*, *SlMKS1b, SlMKS1d* and *SlMKS1e;* a gene designated as *SlMKS1c* on scaffold 05477 appears to be a non-functional gene because it contains a premature stop codon. *SlMKS1a* is the most similar gene to *ShMKS1*, encoding a protein with 95% identity to ShMKS1 (Figure 3-2). Proteins with similar size that are app. 54% identical to ShMKS1 have recently been found in the genome of poplar and grape (Figure 3-2), although their functions are presently unknown. However, the most similar protein encoded by a gene in the Arabidopsis genome, AtMES3 (a protein capable of hydrolyzing methyl IAA and methyljasmonate (Yang et al., 2008), is only 40% identical to ShMKS1 (Figure 3-2). As reported previously for ShMKS1 (Fridman et al., 2005), the N-terminal region of all these newly identified *S. lycopersicum* MKS1 proteins as well as the analogous region within the closely related homologs from other species do not appear to constitute N-terminal extensions that could function as cleavable transit peptides.

Interestingly, while SlMKS1a is the most similar gene to ShMKS1, only one cDNA for it was found in the NCBI database, consistent with its low expression level (Fridman et al., 2005; Ben-Israel et al., 2009). Moreover, no cDNAs/ESTs were found for SlMKS1b while a small number of ESTs for SlMKS1d and SlMKS1e were observed, the majority of which were obtained from trichomes.

We used oligonucleotide primers encoding the beginning and end of the coding region of ShMKS1 in PCR experiments with genomic DNA to isolate the DNA fragment containing all exons and introns of this gene. The number and positions of introns in ShMKS1 were found to be the same as those found in the *S. lycopersicum* MKS1 genes.

Figure 3-2. Comparison of the protein sequence of *S. habrochaites glabratum* ShMKS1 with homologous (MKS1-Like, or MKS1L) sequences from *S. lycopersicum*, grape (Vv), poplar (Pt), and Arabidopsis (At)

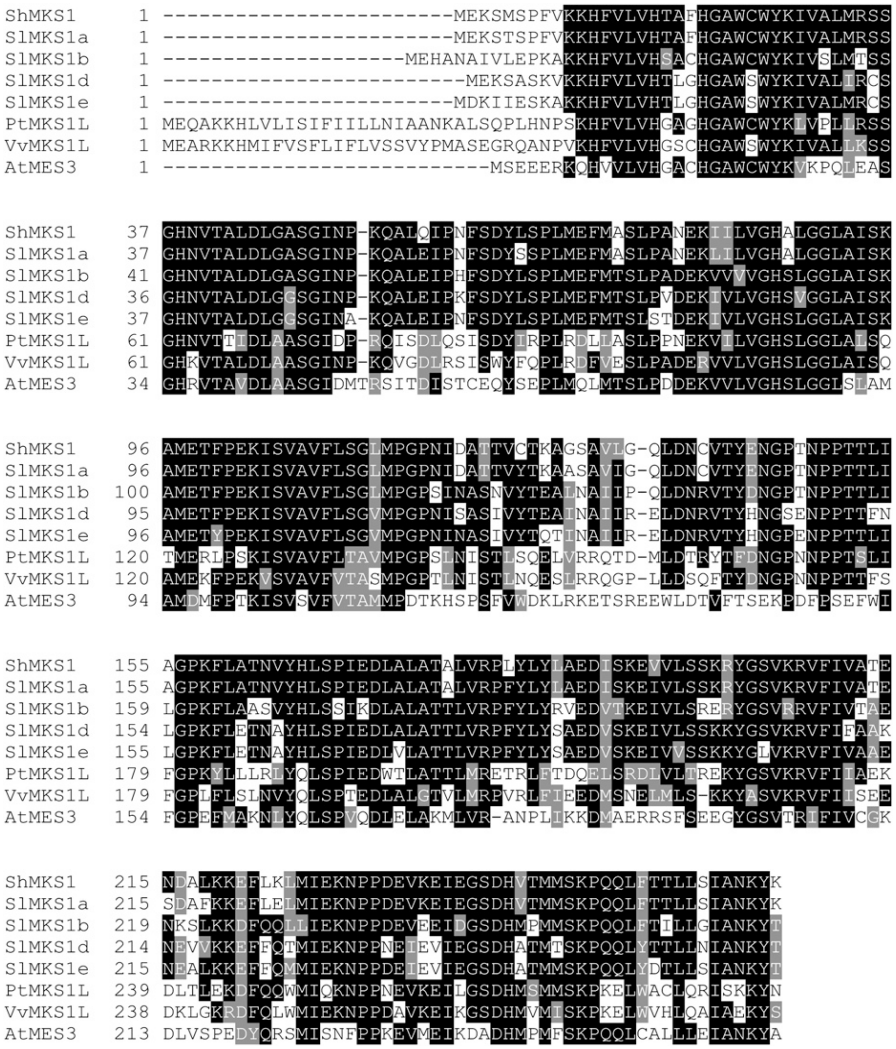

Figure 3-2. Comparison of the protein sequence of *S. habrochaites glabratum* ShMKS1 with homologous (MKS1-Like, or MKS1L) sequences from *S. lycopersicum*, grape (Vv), poplar (Pt), and Arabidopsis (At). Accession numbers are as follows: ShMKS1, GU987105; SlMKS1a, GU987107; SlMKS1b, GU987108; SlMKS1d, GU987110; SlMKS1e, GU987111; PtMKS1L, XM_002313048; VvMKS1L, XM_002284871; AtMES3, At2g23610.

3.3.2 Genes encoding MKS2 in *S. lycopersicum* and *S. habrochaites glabratum*

We previously reported that the longest available *ShMKS2* cDNA contained an open reading frame, starting with a Met codon (ATG), of 149 codons. In addition, we showed that the protein encoded by this cDNA was highly similar (>70% identity across the equivalent region) to putative, functionally uncharacterized proteins from numerous plant species, including four from Arabidopsis, as well as showing limited similarity (<15% identity) to 4HBT from *Pseudomonas* sp. (Ben-Israel et al., 2009). Based on this *ShMKS2* cDNA sequence and analysis of homologous ESTs from *S. lycopersicum* available at the time, the orthologous *S. lycopersicum* gene was deemed to encode a protein similar in size to that of ShMKS2, and consequently a cDNA was isolated by RT-PCR from *S. lycopersicum* and named *SlMKS2* (Ben-Israel et al., 2009). Protein sequence comparisons indicated that the homologous proteins from all other plant species, with the exception of the Solanaceae proteins, have an N-terminal extension that was predicted to function as a transit sequence to direct the protein into the plastids (Ben-Israel et al., 2009).

Mining the *S. lycopersicum* genome resulted in the identification of three genes on the same "scaffold" (Scaffold 04161) that encode proteins with >90% identity (within the equivalent region) to previously reported ShMKS2 sequences; we named these genes *SlMKS2a*, *SlMKS2b*, and *SlMKS2c* (Figure 3-3). The EST databases contain ESTs for *SlMKS2a* and *SlMKS2b*, but not for *SlMKS2c*. Consistent with this observation, our previously reported *SlMKS2* cDNA is derived from *SlMKS2a*, although *SlMKS2c* encodes a protein with a higher identity to ShMKS2 (95%). All three of these *S. lycoperiscum MKS2* genes have 5 exons and 4 introns (whose positions are conserved in

comparison to the intron positions in the homologous Arabidopsis genes). By comparing the sequence of the previously reported *SlMKS2* cDNA with the genomic sequence of *SlMKS2a*, from which it is derived, and the sequence of *ShMKS2* cDNA with the genomic sequence of *SlMKS2c*, to which it is most similar, we noted that the first ATG codon of the open reading frame in each of these cDNAs were equivalent to the ATG codon that occurs in positions 2-4 of exon 2 in the *SlMKS2a* and *SlMKS2c* genes. This suggested that these previously reported *MKS2* cDNAs from both species were incomplete. Indeed, although no *SlMKS2a* EST that contains the entire coding region of exon 1 is available, the sequence of one *SlMKS2b* EST that includes the entire coding region of exon 1 is now in the EST database of NCBI (accession number DB688740).

To determine the beginning of the transcript of *ShMKS2*, two independent 5'RACE experiments were performed using two specific primers complementary to the 3' end and middle of the coding region, respectively. Analysis of the DNA fragments produced in these experiments by agarose gel electrophoresis gave a single sharp band in both cases. The sequence of the resulting fragments from both experiments was determined, and in both cases the sequence obtained indicated that *ShMKS2* transcripts are considerably longer at their 5' ends than was previously seen in the cDNA. This newly uncovered 5' end sequence, identical in both 5'RACE experiments, included the region homologous to exon 1 in the *SlMKS2* genes, which encodes a putative transit peptide, as well as 63 nucleotides of 5' UTR. To determine the complete genomic structure of the *ShMKS2* gene, we used a forward oligonucleotide primer based on the sequence at the beginning of the coding region in exon 1 of *ShMKS2* (as determined by the 5' RACE experiment) and a reverse primer based on the sequence at the end of the

coding region in a PCR experiment with *S. habrochaites glabratum* genomic DNA, and isolated and characterized the genomic fragment containing *ShMKS2*. Using a homology-based PCR approach, we also isolated a 1.5-kb fragment upstream of exon 1 of *ShMKS2* with a forward oligonucleotide primer whose sequence was based on the sequence of the promoter of *SlMKS2c* and a reverse primer derived from the beginning of the *ShMKS2* coding region. Analysis of the complete sequence of the *ShMKS2* gene indicates that its structure, with 5 exons and 4 introns and encoding a protein of 208 amino acid residues with a predicted plastidic transit peptide, is very similar to that of the *S. lycopersicum MKS2* genes (Figure 3-3).

3.3.3 Subcellular localization of ShMKS2

To determine the subcellular localization of ShMKS2 proteins, we injected *Nicotiana benthamiana* leaves with a solution of *Agrobacterium tumefaciens* cells carrying various constructs in which the *ShMKS2* had been fused to the enhanced green fluorescent protein (eGFP) under the control of the cauliflower mosaic virus 35S promoter, and visualized targeting by confocal microscopy. No green fluorescence was detected in tobacco leaf cells transformed with an empty binary vector (Figure 3-4A-C). In the tobacco leaf cells transformed with the full-length *ShMKS2-eGFP* construct, GFP-labeled signals, seen as a punctate pattern, were observed from the same area from which red auto-fluorescence from chlorophyll in the chloroplasts is observed and therefore identified as the chloroplasts (Figure 3-4D-F). In tobacco leaf cells transformed with a *ShMKS2-eGFP* construct that lacked the putative ShMKS2 transit peptide, the green fluorescence dots no longer coincided with chloroplast red fluorescence (Figure 3-4G-I).

Figure 3-3. Comparison of the protein sequence of *S. habrochaites glabratum* ShMKS2 with homologous sequences from *S. lycopersicum*, Arabidopsis, and *Pseudomonas species* (Ps)

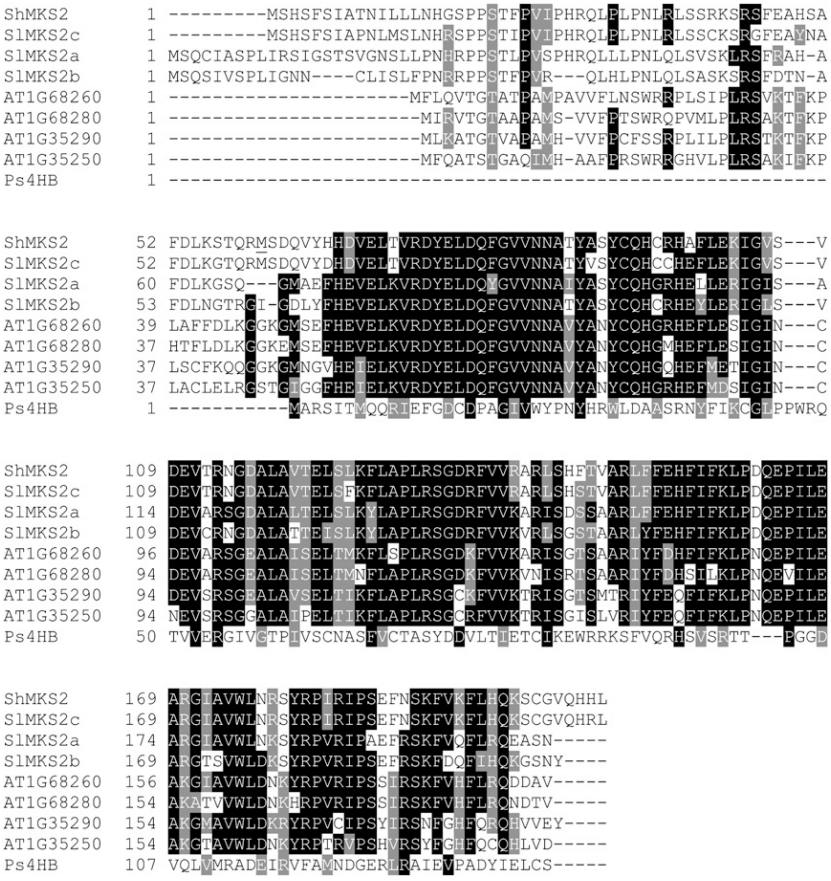

Figure 3-3. Comparison of the protein sequence of *S. habrochaites glabratum* ShMKS2 with homologous sequences from *S. lycopersicum*, Arabidopsis, and *Pseudomonas species* (Ps). Accession numbers are as follows: ShMKS2, GU987106; SlMKS2a, GU987112; SlMKS2b, GU9877113; SlMKS2c, GU987114; Ps4HB, EF569604. The initiating Met codon used to produce ShMKS2 protein without the transit peptide is underlined.

Figure 3-4. Subcellular localization of ShMKS2-eGFP fusion proteins in *N. benthamiana* leaf cells

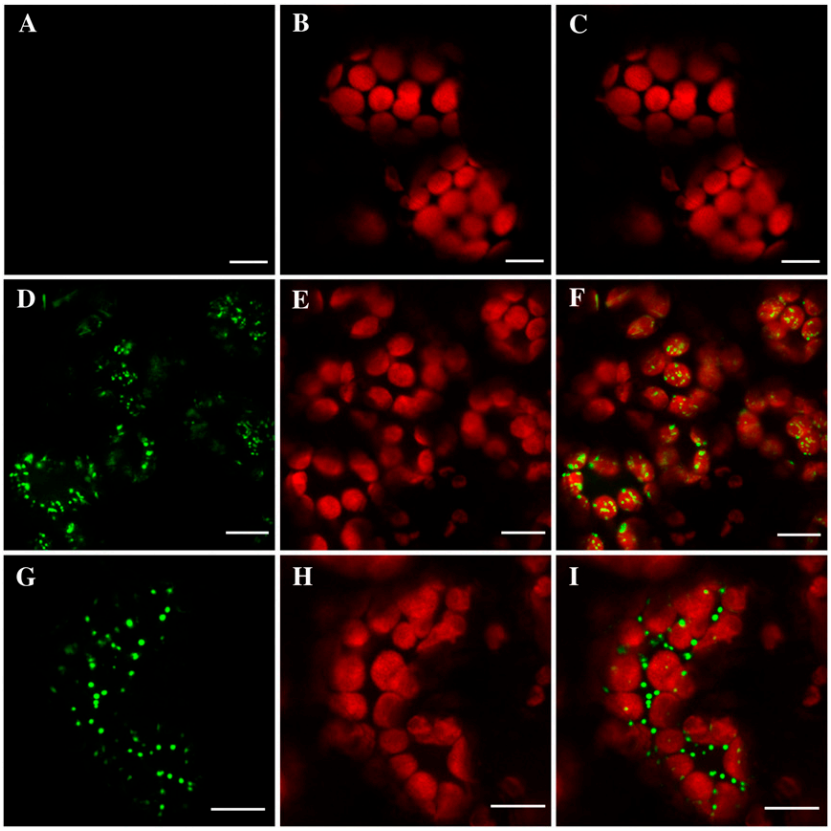

Figure 3-4. Subcellular localization of ShMKS2-eGFP fusion proteins in N. benthamiana leaf cells. The panels shown on the left exhibit green fluorescence from eGFP, the panels in the middle show red fluorescence from plastidic chlorophyll, and each panel in the right column exhibits an overlay of the two panels to its left. A to C, Tobacco cells infiltrated with an empty binary vector. D to F, Tobacco cells infiltrated with a binary vector carrying the complete opening reading frame of ShMKS2 fused to eGFP. G to I, Tobacco cells infiltrated with a binary vector carrying the ShMKS2 gene lacking the putative transit peptide and fused to eGFP. Bars = 10 μm.

3.3.4 Expression of *ShMKS1* and *ShMKS2* in *E. coli* and production of methylketones

We previously reported that analysis of the spent medium of *E. coli* cells expressing *ShMKS2* demonstrated the presence of several methylketones with 2-

tridecanone, 2-tridecenone, and 2-undecanone predominating (Ben-Israel et al., 2009). However, in those studies, no attempt was made to measure the production of 3-ketoacids, which are the putative intermediates in the synthesis of the final methylketone products (Figure 3-1). Direct measurement of 3-ketoacids is difficult since these compounds are water-soluble and unstable upon analysis by GC-MS. However, an indirect approach to measure such compounds has been developed in which a solution contains 3-ketoacids is treated with sulfuric acid and heat, leading to greatly enhanced spontaneous decarboxylation and thus the conversion of the water-soluble products to easily extractable methylketones, which can then be directly measured by GC-MS (Matiasek et al., 2001).

To test if expression of either *ShMKS1* or *ShMKS2* (without its transit peptide) in *E. coli* results in the formation of 3-ketoacids, we collected spent media of bacterial cells expressing each of them (by centrifuging out the cells at the end of the incubation period), heated the spent media at 75 ^{o}C for 30 min in the presence of 1 M sulfuric acid, followed by extraction with hexane, then performed GC-MS analyses on the extracted material. Spent media of cells expressing either *ShMKS1* or a plant gene unrelated to the methylketone pathway, as well as of cells carrying the same vector (*pEXP5-CT/TOPO*) without an introduced gene, contained no methylketones with or without the acid and heat treatment (Figure 3-5). The spent medium of *E. coli* cells expressing *ShMKS2*, on the other hand, contained 5.6±0.32 µg/µl of total methylketones, and the amount of methylketones increased 8 fold, to 40.7±2.1 µg/µl, after the spent medium was treated with acid and heat (Figure 3-5).

Figure 3-5. Total amount of methylketones found in spent medium of *E. coli* cells expressing *ShMKS1*, *ShMKS2*, and *ShMKS2*(D79A) (all missing the transit peptide-coding region) from the *pEXP-TOPO-CT* bacterial expression vector

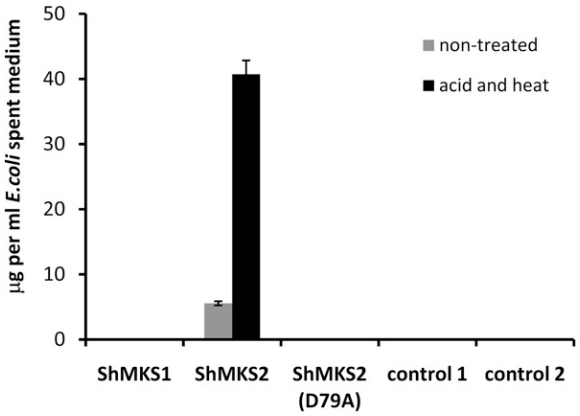

Figure 3-5. Total amount of methylketones found in spent medium of *E. coli* cells expressing *ShMKS1, ShMKS2,* and *ShMKS2* (D79A) (all missing the transit peptide-coding region) from the *pEXP-TOPO-CT* bacterial expression vector. Cells were grown and spent medium was collected and treated as described in "3.5 MATERIALS AND METHODS". Control 1 cells expressed *Clarkia breweri* Isoeugenol synthase1 on *pEXP5-CT/TOPO* (Koeduka et al., 2008). Control 2 cells contained a *pEXP5-CT/TOPO* vector with no insert. Values are averages 6 SE calculated from three experiments.

In the bacterial thioesterase 4HBT, the aspartate residue at position 17 was identified as the catalytic residue required for thioester bond hydrolysis (Benning et al., 1998). We mutated the equivalent Asp codon in *ShMKS2* (the Asp encoded by codon 79 of the complete open reading frame) to an Ala codon and expressed the mutated genes (without the transit peptide-encoding region) in *E. coli*. The spent medium of cells expressing this mutant ShMKS2 protein did not contain any methylketones, with or without prior acid and heat treatment (Figure 3-5).

A more detailed analysis of the spent medium of *E. coli* cells expressing *ShMKS2* showed that the major compounds in the untreated spent medium were 2-undecanone (0.51 µg/µl), 2-tridecanone (2.6 µg/µl spent medium), 2-tridecenone (1.1 µg/µl) and 2-

pentadecenone (1.3 µg/µl) (Figure 3-6). Lower amounts of 2-nonanone (0.05 µg/µl) were also detected (Figure 3-6). When the spent medium was heated at 75°C for 30 min in the absence of sulfuric acid, the yield of methylketones increased up to 0.60±0.05 µg/µl 2-nonanone (11.9-fold over non-treated control), 6.22±0.34 µg/µl 2-undecanone (12.2-fold), 9.61±0.27 µg/µl 2-tridecanone (3.7-fold), 9.94±0.83 µg/µl 2-tridecenone (9-fold) and 3.94±0.54 µg/µl 2-pentadecenone (3.0-fold). The yield was increased even further in the combined heat and acid treatment, reaching a maximum levels of 0.95±0.04 µg/µl 2-nonanone (19.1-fold over non-treated control), 9.33±0.61 µg/µl 2-undecanone (18.3-fold), 11.99±0.51 µg/µl 2-tridecanone (4.6-fold), 13.04±0.63 µg/µl 2-tridecenone (11.8-fold) and 5.69±0.21 µg/µl 2-pentadecenone (4.4-fold).

Figure 3-6. Methylketone production by *E. coli* cells expressing *ShMKS2*

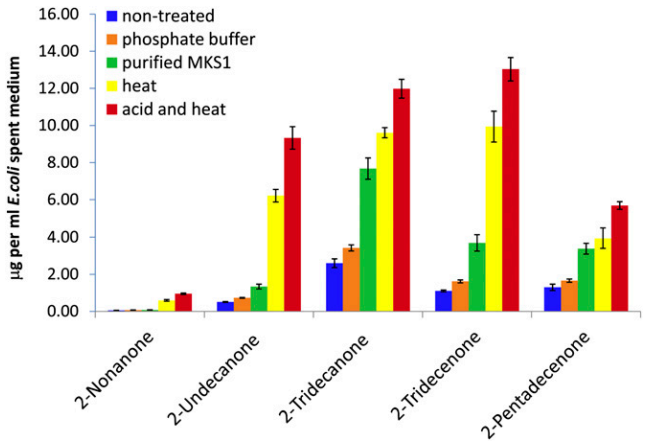

Figure 3-6. Methylketone production by *E. coli* cells expressing *ShMKS2*. Treated and nontreated spent medium of *E. coli* cells expressing ShMKS2 (without the transit peptide-coding region) was extracted with hexane, and the methylketone content was measured by GC-MS. Treatments included heat, acid and heat, purified ShMKS1 protein in phosphate buffer, and phosphate buffer alone. Values are averages 6 SE calculated from three experiments. See text for details.

When purified ShMKS1 (75 µg/ml in 12.5 mM Na^+-phosphate buffer, pH 6.8) was added to the spent medium (3 µg/ml final concentration) 2 h prior to hexane

extraction, levels of extractable methylketones were significantly higher than in spent medium treated with phosphate buffer alone, although not as high as the levels of methylketones observed after acid and heat treatments. Moreover, the treatment with purified ShMKS1 seemed to favor an increase in 2-tridecanone over other methylketones (Figure 3-6).

3.3.5 *In vitro* decarboxylase activity assays for ShMKS1 and ShMKS2

To examine the potential decarboxylating activity of ShMKS1 as well as ShMKS2 *in vitro*, we tested partially purified recombinant ShMKS1 and ShMKS2 proteins for decarboxylation activity with 3-ketomyristic acid. Notably, ShMKS1 produced 2.6 nM of 2-tridecanone/µg of protein/min, while ShMKS2 had essentially no activity (Figure 3-7). Measurements of the kinetic parameters of ShMKS1 determined a Km value of 18.4 ± 5.6 µM for 3-ketomyristic acid, with an apparent k_{cat} of 227.9 ± 24.1 min^{-1}.

Figure 3-7. Decarboxylase activity assays for ShMKS1 and ShMKS2 using 3-ketomyristic acid as the substrate

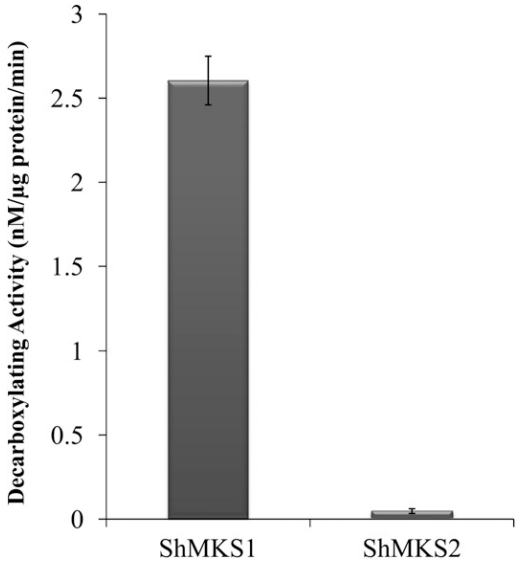

Figure 3-7. Decarboxylase activity assays for ShMKS1 and ShMKS2 using 3-ketomyristic acid as the substrate. Purified recombinant proteins were assayed as described in "3.5 MATERIALS AND METHODS", and the mean and SD values were calculated from three replicates and given as shown.

3.3.6 *In vitro* thioesterase activity assays for ShMKS1 and ShMKS2

To examine the potential thioesterase activity of ShMKS1 and ShMKS2 *in vitro*, we added 2.5 µg of each protein to a 500 µl solution of freshly prepared 3-ketomyristoyl-ACP. This substrate was prepared enzymatically by the addition of multiple chemicals and enzymes and letting the reaction proceed for 5 hours (see Methods). Due to the highly unstable nature of 3-ketomyristoyl-ACP, this compound was not further purified but instead the solution in which it was synthesized was then used and is referred to below as the "substrate solution".

Aliquots of this substrate solution were incubated with either buffer, ShMKS1, ShMKS2, or both ShMKS1 and ShMKS2 for 30 min at 23 ^{o}C. Extraction of buffer-incubated substrate solution with hexane, followed by GC-MS analysis, resulted in detection of almost no 2-tridecanone (Figure 3-8). However, when the buffer-incubated substrate solution was treated with acid and then heated at 75 ^{o}C, cooled and then extracted with hexane, 2-tridecanone was detected (Figure 3-8), indicating that the substrate solution contained free 3-ketomyristic acid in addition to 3-ketomyristoyl-ACP. When the substrate solution was incubated with ShMKS1, followed by treatment with acid and heat and then extracted with hexane, the amount of 2-tridecanone obtained was slightly higher, but not significantly so (t test, $P = 0.062$, $\alpha = 0.05$), than the amount found in the buffer-incubated sample treated with acid and heat (Figure 3-8). However, when the substrate solution was incubated with ShMKS2, then further treated with acid and heat, the amount of 2-tridecanone was approximately 3-fold higher than that found in

buffer-incubated substrate solution treated with acid and heat. Finally, when the substrate solution was co-incubated with both ShMKS1 and ShMKS2 for 30 min, then directly extracted with hexane (without acid and heat treatment), the amount of 2-tridecanone was approximately 6-fold higher than that found in buffer-incubated substrate solution treated with acid and heat (Figure 3-8).

Figure 3-8. Thioesterase activity assays for ShMKS1 and ShMKS2

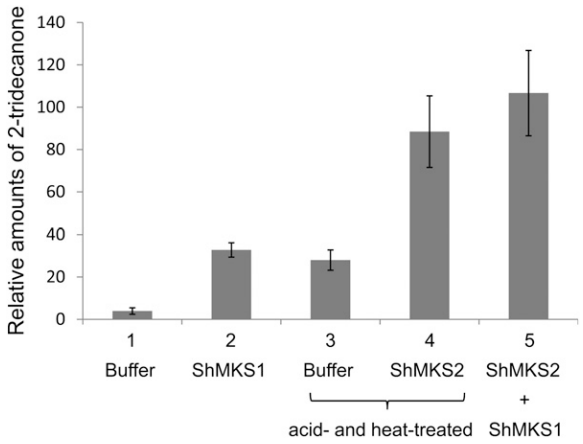

Figure 3-8. Thioesterase activity assays for ShMKS1 and ShMKS2. To a 500-mL solution of enzymatically prepared 3-ketomyristoyl-ACP (see "3.5 MATERIALS AND METHODS"), a 20-mL solution of the following was added: lane 1, enzyme buffer; lane 2, 2.5 mg of ShMKS1 in buffer; lane 3, enzyme buffer; lane 4, 2.5 mg of ShMKS2; lane 5, 2.5 mg of ShMKS1 and 2.5 mg of ShMKS2. Each reaction was incubated for 30 min at 23 ^{o}C, after which the reaction solution was either extracted directly with hexane (lanes 1, 2, and 5) or first treated with acid and heated at 75 ^{o}C for 30 min (lanes 3 and 4), then cooled down to room temperature and extracted with hexane. Hexane extracts were analyzed by GC-MS. Mean and SD values were calculated from three replicates.

3.4 DISCUSSION

3.4.1 Enzymatic activities of ShMKS1 and ShMKS2

Although we previously reported that incubation of crude preparation of 3-ketomyristoyl-ACP derived from a complex mixture of fatty acid biosynthetic components with purified ShMKS1 resulted in the appearance of 2-tridecanone, we also

noted that the yield of the *in vitro* reaction was exceedingly low (Fridman et al., 2005). Here we report that the expression of *ShMKS1* in *E. coli* does not result in the production of methylketones. On the other hand, we confirm and expand on a subsequent finding (Ben-Israel et al., 2009) that methylketones are present in the growth medium of *E. coli* expressing *ShMKS2* (Figure 3-6). Furthermore, treatments of this spent medium with acid and heat, or with purified ShMKS1 protein, greatly elevate the levels of methylketones extracted and detected by GC-MS analysis. Since it is well established that treatment with acid and heat, or even heat alone, greatly accelerates decarboxylation of 3-ketoacids to form methylketones (Matiasek et al., 2001), the increase in levels of methylketones extracted from this spent medium after the acid and heat treatment indicates that substantial amounts of 3-ketoacids were present in the spent medium prior to this approach.

Because treatment of spent medium of ShMKS2-expressing *E. coli* cells with purified ShMKS1 also increased the extractable levels of methylketones by several fold, it appears that ShMKS1 possessed decarboxylase activity. However, ShMKS1 did not seem to possess a thioesterase activity when expressed in *E. coli*, since no methylketones could be detected from the spent medium of these cells even after acid and heat treatment, indicating a lack of 3-ketoacids in the spend medium (Figure 3-5). On the other hand, the presence of substantial amounts of 3-ketoacids in the spent medium of *E. coli* cells expressing ShMKS2 indicated that ShMKS2 acted as a thioesterase, producing 3-ketoacids from either 3-ketoacyl-CoA or 3-ketoacyl-ACP precursors.

To directly test the enzymatic activities of ShMKS1 and ShMKS2, we carried out *in vitro* assays. ShMKS1 exhibited decarboxylase activity on 3-ketomyristic acid, while

ShMKS2 did not (Figure 3-7). When ShMKS1 was added to a solution containing enzymatically synthesized 3-ketomyristoyl-ACP but also some free 3-ketomyristic acid (due to the instability of 3-ketomyristoyl-ACP it could not be further purified), the amount of 2-tridecanone obtained was slightly higher than the amount of 2-tridecanone observed after incubation of the same volume of substrate solution with buffer, followed by acid and heat treatment. However, the difference was not significant, thus indicating that ShMKS1 possesses little or no thioesterase activity, consistent with the lack of 3-ketoacids in the spent medium of *ShMKS1*-expressing *E. coli*. On the other hand, incubation of the substrate solution with ShMKS2 resulted in 3-fold higher amount of 2-tridecanone than what would be expected simply from the acid- and heat-induced decarboxylation of the 3-ketomyristic acid present in the solution (Figure 3-8). It must therefore be concluded that the additional amount of 2-tridecanone resulted from the thioesterase activity of ShMKS2 on 3-ketomyrstoyl-ACP, leading to 3-ketomyristic acid which was then decarboxylated in the acid and heat treatment. Furthermore, co-incubation of substrate solution with both ShMKS1 and ShMKS2 resulted in a 2-fold larger amount of 2-tridecanone than that produced in the incubation with only ShMKS2, possibly because the decarboxylating activity of ShMKS1 relieved the ShMKS2 thioesterase enzyme of product inhibition.

Since the *in vitro* assays indicate that ShMKS2 lacks decarboxylase activity, the small amount of methylketones (relative to the corresponding 3-ketoacids) found in the spent medium of *ShMKS2*-expressing *E. coli* cells (Figure 3-6) was thus most likely due to slow non-enzymatic decarboxylation of the 3-ketoacids released by ShMKS2 during the overnight incubation period.

Taken together, these data suggest that in tomato trichomes, ShMKS2 and ShMKS1 work sequentially, ShMKS2 first liberating 3-ketoacids and ShMKS1 catalyzing their decarboxylation to produce the final methylketone products. This is consistent with the observations that several-fold more methylketones are found in the tomato trichomes when ShMKS2 is highly expressed and ShMKS1 is expressed at low levels than in the opposite case when ShMKS1 is expressed at high levels but ShMKS2 is expressed at low levels (but methylketone levels are much higher than either case when both genes are highly expressed) (Ben-Israel et al., 2009). This is easily rationalized since ShMKS2 activity is required for production of 3-ketoacids while ShMKS1 activity is limited to decarboxylation, and some decarboxylation may occur spontaneously both *in planta* and in *E. coli* given sufficient time (although ShMKS1 activity substantially increases methylketone production in both cases). Our earlier observation that small amounts of 2-tridecanone were produced *in vitro* when a solution containing enzymatically prepared 3-ketomyristoyl-ACP was incubated with ShMKS1 can now be explained as the result of the decarboxylating activity of ShMKS1 on the free 3-ketomyristic acid present in the substrate solution, as shown in the present study (Figure 3-8).

3.4.2 The mature ShMKS2 protein is localized to the plastids and is likely to hydrolyze 3-ketoacyl-ACP substrates

The results of transient expression of ShMKS2-GFP fusion protein in tobacco leaves indicated that the ShMKS2 protein localized to the plastids, as has been previously shown for ShMKS1 by *in vitro* chloroplast import studies (Fridman et al., 2005). A plastidic localization is consistent with the hypothesis that ShMKS2 works by competing

for 3-ketoacyl-ACP intermediates formed iteratively during fatty acid elongation in the plastids, rather than from 3-ketoacyl-CoA intermediates formed catabolically during fatty acid degradation in peroxisomes. This conclusion is consistent with the demonstrated *in vitro* thioesterase activity of ShMKS2 with 3-ketomyristoyl-ACP, although 3-ketoacyl-CoA could not be obtained for comparison purposes.

The range of methylketones produced by expressing *ShMKS2* in *E. coli* was very similar to that seen in *S. habrochaites glabratum* trichomes, with 2-tridecanone and 2-undecanone being the most abundant, suggesting that ShMKS2 displays a preference for similar chain-length intermediates in both plants and *E. coli*. Intriguingly, substantial amounts of 2-tridecenone (with one double bond present between C3 and C4) and some 2-pentadecenone (also with one double bond present between C3 and C4) were also observed in *ShMKS2*-expressing *E. coli* cultures (Figure 3-6), whereas these compounds were not detectable in *S. habrochaites glabratum* trichomes (Fridman et al., 2005). The position of the double bond indicates that ShMKS2 is able to act on 2-oxo-4-en-acyl-ACPs that are at least 14 carbons long. Such intermediates could have resulted from the elongation of un-reduced 2-en acyl-ACPs. Whether ShMKS2 acts on such intermediates in *S. habrochaites glabratum* trichomes or whether this activity is simply a peculiarity of its heterologous expression in *E. coli* is not presently known.

3.4.3 Evolution of ShMKS1 and ShMKS2

Low levels of methylketones have occasionally been found in plant species from diverse taxa outside the genus *Solanum* (Jasperson and Jones, 1947; Henricsson et al., 1996), but their mode of synthesis has not yet been determined. In *Solanum*, only *S. habrochaites glabratum* has been reported to synthesize and store high levels of

methylketones (up to 8 mg/leaf FW) in their Type VI glandular trichomes, while the trichomes of the cultivated tomato (*S. lycopersicum*) contain methylketones at levels that are about 1,000-fold lower (Ben-Israel et al., 2009). We also previously showed that the expression of both *SlMKS1* and *SlMKS2* genes in *S. lycopersicum* trichomes is considerably lower than in their wild-type cousins (Ben-Israel et al., 2009). The presence of proteins in species outside *Solanum* with homology to MKS1 and MKS2 raises the question of whether such proteins are involved in methylketone biosynthesis, albeit at very low rates, and if not, how the *Solanum* MKS1 and MKS2 proteins cooperatively acquired the catalytic ability to biosynthesize considerable amounts of methylketones.

It is possible that regardless of the original function of *MKS2*-like genes, simply increasing the expression of such a gene possessing a low level of an alternative activity (with a concomitant increase in fatty acid biosynthetic flux) will lead to production of some methylketones (for example, high-level expression in *E. coli* of the *ShMKS2* homologs from *Arabidopsis thaliana* also leads to substantial methylketone production). However, this will also lead to accumulation of 3-ketoacids, which could interfere with fatty acid biosynthesis, and perhaps MKS1 with evolved 3-ketoacid decarboxylation activity was selected because it conferred an advantage to the plant by decomposing such acids. It is interesting to note that unlike MKS2-like proteins from Arabidopsis, which catalyze a similar reaction to ShMKS2 in *E. coli*, ShMKS1 and its ortholog SlMKS1a, are fundamentally different from the other SlMKS1 and MKS1-like proteins in other species, in that the first two are missing what at first glance appears to be a catalytically essential Ser (at position 87 in ShMKS1), part of the catalytic triad necessary for the *α/β*-hydrolase activity of many proteins in the *α/β*-hydrolase superfamily (Hotelier et al.,

2004; Forouhar et al., 2005). Although ShMKS1 and SlMKS1a clearly belong to this family, they have an Ala substitution at this position and are therefore unlikely to possess hydrolase activity. It thus appears that a bona fide MKS1 evolved recently in the Solanum lineage by acquiring decarboxylase activity and shedding the hydrolase activity (Auldridge et al., 2012).

3.5 MATERIALS AND METHODS

3.5.1 Bioinformatics

Homologs of *ShMKS1* and *ShMKS2* were identified by BLAST search of the "Tomato WGS Scaffolds Prelease (previous)" data set (http://solgenomics.net/). The genomic sequences identified in this search were checked (by BLAST) with the EST database (http://bioinfo.bch.msu.edu/trichome_est) from the trichomes of *S. lycopersicum*. The positions of exons were determined by comparisons with ESTs directly derived from these genes or, in the absence of ESTs, from a comparison with *ShMKS1* and *ShMKS2* cDNAs, respectively. Protein sequence comparisons were performed with the CLUSTAL_X protocol (Thompson et al., 1997).

3.5.2 Gene isolation

A full-length cDNA of *ShMKS2* was isolated by RT-PCR using the oligonucleotides 5'- ATGTCTCATTCGTTCAGCA-3' and 5'-GAGATGATGTTGTACACCGCAACT-3' with total RNA from *S. habrochaites*. The genomic sequence of *ShMKS2* was obtained using total DNA as the templates. The promoter sequence of *ShMKS2* was isolated by PCR using the oligonucleotides 5'-CTGTGGCAATTGTTAATTGGTGGGAGT -3' and 5'-

GAGCGGGAGTTGCCGGTGAG -3'. The genomic sequence of *ShMKS1* was obtained by PCR with nucleotides 5'- ATGGAGAAAAGCATGTCGCCA-3' and 5'-TTTATACTTGTTAGCGATGCTTAGAAGAGT-3'. All PCR reactions employed KOD hot start polymerase (Novagen). Products were spliced into the pGEM-T easy vector (Promega) and sequenced.

3.5.3 5' RACE

5' RACE procedure used the SMART RACE cDNA amplification kit (Clontech laboratories) with SuperScript II Reverse Transcriptase and anchored Oligo(dT)20 (Invitrogen). Two independent experiments with different primers were performed for each gene with RACE-ready cDNA synthesized from total RNA from the leaves. Products were spliced into the pGEM-T easy vector (Promega) and sequenced.

3.5.4 Genome walking

Isolation of the promoter region of *ShMKS1* was done with the GenomeWalker Universal Kit (Clontech, Inc.) according to the manufacturer's instructions.

3.5.5 Constructs for subcellular localization

Full-length *ShMKS2* cDNA and *ShMKS2* without the coding region of exon 1 (starting with the first ATG codon in exon 2) were amplified by KOD polymerase to add *Bgl* II and *Sal* I restriction sites and spliced into pSAT6A-EGFP-N1 (Tzfira et al., 2005). The expression cassettes were digested by PspI and ligated to pPZP-RCS2 binary vector and transferred into *Agrobacterium tumefaciens* strain EHA105 (Tzfira et al., 2005).

3.5.6 Transient expression in *Nicotiana benthamiana* and confocal microscopy

Agrobacterium tumefaciens cells were grown in a shaker-incubator at 30^{o}C at 200 rpm in LB broth supplemented with 200 µg per mL spectinomycin and 200 ug per mL streptomycin until the optical density of the culture at 600 nm reached 0.7-0.9. Bacteria were pelleted by centrifugation at 5000 rpm for 10 min at room temperature, and resuspended to OD 0.4 in fresh infiltration buffer containing 10 mM $MgCl_2$ and 0.1 µM acetosyringone. The resulting mix was diluted with infiltration buffer to OD of 0.1 and infiltrated into the abaxial air spaces of 4-6 week-old *N. benthamiana* plants by a syringe, as previously described previously (Yang et al., 2000). The plants were then returned to the growth chamber for 48-72 hours for an optimal expression of the gene.

To test for the localization of the ShMKS2 protein, the infiltrated tobacco leaves were dissected and mounted on a microscope slide with distilled water and examined using a Leica SP5 confocal system and a 63x (1.3NA) glycerin immersion lens. eGFP was visualized using an argon gas 488 nm laser, a RP500 dichroic mirror, and PMT detection from 500-530 nm. Chloroplast fluorescence was visualized using the same argon gas 488 nm laser, a RP500 dichroic mirror, and PMT detection from 650 nm long pass.

3.5.7 Expression of *ShMKS1* and *ShMKS2* in *E. coli*

The construction of bacterial expression vectors for expressing *ShMKS1* and *ShMKS2* separately were previously reported (Ben-Israel et al., 2009). The expression vectors were introduced into *E. coli* BL21 Star (DE3) (Invitrogen) cells, and gene expression was induced by the addition of 0.5 mM IPTG after the culture optical density at 600 nm had reached 0.65. After induction with IPTG and growth at 18^{o}C overnight, the

cells expressing *ShMKS1* or *ShMKS2* were centrifuged at 5000 rpm for 15 minutes, and 1-ml aliquots of the spent medium were placed in individual vials for further analysis.

3.5.8 GC-MS analysis of spent medium of *E. coli* cells expressing *ShMKS2*

Aliquots (1 ml) of the spent medium of *E. coli* expressing *ShMKS2*, obtained by centrifuging the culture solution after the incubation time at 5,000 rpm for 15 min and collecting the solution without the cells, were treated in the following ways: 1. Incubated with 40 µ1 (3 µg) of purified MKS1 in phosphate buffer (12.5 mM NaH_2PO_4, 125 mM NaCl, 2mM DTT, pH 6.8) for 2 hrs at 30°C. 3. Incubated at 75°C for 30 min followed by 30 min at 30°C. 4. Incubated with 1 ml of 2M H_2SO_4 at 75°C for 30 min followed by 30 min at 30°C. After the various treatments, 1 ml of hexane containing 5 ng/µl linalool as an internal standard was added and the resulting mixture was vortexed and centrifuged at 5000 rpm for 10 min. Two µ1of the resulting extract were injected into the GC-MS for determination of methylketones. GC-MS and product analysis were performed as described previously (Ben-Israel et al., 2009).

3.5.9 Affinity purification of ShMKS1 and ShMKS2

Affinity purification of His-tagged ShMKS1 and ShMKS2 was as previously reported (Ben-Israel et al., 2009). After elution from the Nickel-agarose column, the proteins were analyzed by SDS-PAGE and dialyzed against 50 mM phosphate buffer pH 6.8, 500 mM NaCl, and 1M $(NH_4)_2SO4$, and 2 mM DTT. ShMKS1 purity was estimated at 99 %, and ShMKS2 purity was estimated at 6 %.

3.5.10 Decarboxylase activity assays

A typical decarboxylase assay consisted of a 500 µl reaction solution containing ShMKS1 or ShMKS2 (2.5 µg), 3-ketomyristic acid (0.1 mM), and 1,3-bis(tris(hydroxymethyl)methylamino) propane – Na^+ (20 mM, pH 7.0). For measuring kinetic parameters, substrate concentrations ranged from 5 to 75 µM. Assays were performed at 23^{o}C for 10 min after addition of protein. Reactions were quenched by addition of 25 µl 3 M NaOH and the methylketone products were extracted with 500 µl hexane. For reaction normalization, a standard concentration of 2-undecanone was added prior to extraction to a final concentration of 4 µM. Reaction products (5 µl) were analyzed by a modified procedure described in (O'Maille et al., 2004) using a Hewlett–Packard 6890 gas chromatograph (GC) coupled to a 5973 mass selective detector (MSD) equipped with an HP-5MS capillary column (0.25 mm i.d. 30 m length with 0.25 µm film thickness) (Agilent Technologies). Product quantification was performed using total ion monitoring (TIM) mode where all ions in the mass spectrum contribute to the measured response. The GC was operated at a He flow rate of 1.5 mL/min, and the MSD was operated at 70 eV. Splitless injections (5 µL) were performed with an inlet temp of 280°C. The GC was programmed with an initial oven temp of 60°C (2-min hold), which was then increased 5°C/min up to 200°C, followed by a 50°C/min ramp until 280°C (5-min hold). A solvent delay of 8.5 min was included prior to the acquisition of the MS data. 2-Tridecanone was quantified by integration of peak areas using Enhanced Chemstation (version B.01.00, Agilent Technologies). The GC–MS instrument was

calibrated with an authentic undecanone standard included in the quenched reactions prior to hexane extraction.

3-Ketomyristic acid was prepared from methyl 3-oxotetradecanoate (1 mmol) to which was added 6 ml 3.0 M aqueous NaOH in addition to several drops of THF to aid in dissolution of the esterified starting material. The mixture was stirred at 23 °C for 12 hr. The reaction mixture was then diluted with 10 ml water and acidified to pH 2-3 by adding 3 M HCl dropwise while monitoring pH. The acidified mixture was next extracted 5 times with 30 ml methylene chloride each. The organic phases were pooled, washed with saturated NaCl and then dried using anhydrous sodium sulfate. The methylene chloride solvent was removed under reduced pressure yielding an opaque yellowish powder. This powder was purified using a normal phase silica gel column after dissolution in a minimal amount of column solvent [methylene chloride-methanol (3:1)] to afford 3-ketomyristic acid.

3.5.11 Thioesterase activity assays

3-Ketomyristoyl-ACP was synthesized in a 500 µl reaction volume containing 1,3-bis(tris(hydroxymethyl)methylamino) propane (20 mM, pH 7.0), malonyl-CoA (0.2 mM), lauroyl-CoA (0.2 mM), *Sh*ACP (0.1 mM), *Ec*FabD (10 µg) and *Mt*FabH (10 µ g). After 5 hours at 37 ^{o}C, the reaction solution was used as the substrate solution for subsequent treatments. A 20 µl solution containing 2.5 µg of enzyme (or buffer only) was added, and the reaction was incubated for 30 min at 23 ^{o}C, extracted with hexane directly or first treated with acid and heat and then extracted with hexane. Hexane extracts were then analyzed by GC-MS.

Chapter 4 Heterologous expression of methylketone synthases 1 and 2 leads to accumulation of methylketones and myristic acid in Arabidopsis and tobacco but not in tomato plants[3]

4.1 ABSTRACT

Some plants produce methylketones as potent defense compounds against various insects. *Solanum habrochaites*, a relative of the cultivated tomato (*S. lycopersicum*), synthesizes large amounts of 2-methylketones in its glandular trichomes, but tomato trichomes contain little or no methylketones. Two enzymes, methylketone synthases 1 and 2 (MKS1 and MKS2), are required to convert *β*-ketoacyl-ACP intermediates of the fatty acid biosynthetic pathway to methylketones. ShMKS2 is a thioesterase that hydrolyzes *β*-ketoacyl-ACP, and ShMKS1 is a decarboxylase that converts the resulting 3-ketoacid to 2-methylketones. We introduced *ShMKS2* by itself or together with *ShMKS1* to Arabidopsis, tobacco and tomato under the control of the *35S*, *Rbcs*, and tomato trichome-specific promoters. Young tobacco and Arabidopsis plants expressing both genes under the control of 35S and Rbcs promoters produced methylketones in their leaves but had serious growth defects, and as plants mature they ceased to produce methylketones. Tobacco plants but not Arabidopsis or tomato plants expressing only *ShMKS2* under the 35S promoter also synthesized methylketones, but at a lower rate.

[3] This work has been submitted as Geng Yu, Eran Pichersky. Heterologous expression of methylketone synthases 1 and 2 leads to accumulation of methylketones and myristic acid in Arabidopsis and tobacco but not in tomato plants. Plant Physiol.

Transgenic tomato plants expressing *ShMKS1* and *ShMKS2* under trichome-specific promoters did not produce elevated levels of methylketone. These results suggest that increases in methylketone production in plants will require the targeting of the pathway to self-contained structures in the plant that also have high flux of fatty acid biosynthesis.

4.2 INTRODUCTION

The wild tomato species *Solanum habrochaites* ssp *glabratum* (previously known as *Lycopersicon hirsutum* ssp *glabratum*) has long been known for its resistance to a wide spectrum of arthropod pests, including spider mites (*Tetranychus urticae*), glasshouse whitefly (*Trialeurodes vaporariorum*), tomato fruitworm (*Heliothis zea*), Colorado potato beetle (*Leptinotarsa decimlineata*), tomato pinworm (*Keiferia lycopersicela*) and tobacco hornworm (*Manduca sexta*), due to its production of methylketones (Maluf et al., 1997). Methylketones, mainly in the form of 2-undecanone and 2-tridecanone, are found in prodigious amounts in leaves of *S. habrochaites* and exhibit strong pesticidal activity, with 2-tridecanone as the most effective compound (Williams et al., 1980; Antonious et al., 2003). The toxicity of 2-tridecanone to arthropod was confirmed by a positive association of its concentration *in vivo* in the *M. sexta* or *Scrobipalpuloides absoluta* with the lethality rate exhibited by these two insects (Fery and Kennedy, 1987; Gonçalves et al., 1998). Since cultivated tomato (*S. lycopersicum*, previously known as *Lycopersicon esculentum*) plants possess very low levels of methylketones, and breeding insect resistant tomato varieties that accumulate methylketones has not been successful (Hartman and St. Clair, 1998), understanding the biosynthetic mechanism in *S. habrochaites* may help to engineer methylketone-based defenses into cultivated tomatoes.

Biosynthesis and storage of methylketones are confined to dedicated structures in *S. habrochaites* termed type VI trichomes (Antonious et al., 2003; Fridman et al., 2005). Trichomes are small protrusions of epidermal originated cells on surface of aerial parts of about one third vascular land plants (Luckwill, 1943; Schilmiller et al., 2008). In the *Solanum* genus, seven types of trichomes have been reported in various tomato species, with type VI trichome found in all those species (Luckwill, 1943; McDowell et al., 2011).

Type VI trichomes in *S. habrochaites* have a multi-cell stalk and four gland cells on the apex of the stalk, and are densely present on the surfaces of leaves and stems (Ben-Israel et al., 2009). Methylketones accumulate in large amounts in type VI trichomes, specifically in the apical gland cells, with concentrations of up to 8 mg per gram fresh leaf weight (FLW) observed (Fridman et al., 2005). Analysis of a type VI-specific expressed sequence tag (EST) database from *S. habrochaites* (accession number PI126449), which is a methylketone-producing line, followed by gene expression comparison with a trichome EST database from *S. habrochaites* accession number LA1777, which does not produce methylketones, identified the gene *S. habrochaites* methyl ketone synthase I (*ShMKS1*) as correlated methylketone production (Fridman et al., 2005). Additional genetic and genomic analyses identified a second gene, designated *ShMKS2* (*S. habrochaites* methyl ketone synthase 2), whose expression was also positively related with methylketone production (Ben-Israel et al., 2009).

ShMKS1 and *ShMKS2* are specifically expressed in the gland cells of type VI trichomes in *S. habrochaites*, and both proteins are targeted to the plastids (Yu et al., 2010; Akhtar et al., 2013; Yu et al., Submitted). *In vitro* biochemical assays and heterologous expression in *Escherichia coli* demonstrated that the ShMK2 protein is a

thioesterase that hydrolyzes *β*-ketoacyl-ACPs, primarily C12 and C14 intermediates of the fatty acid biosynthetic pathway, to give the corresponding 3-keto acids, and that ShMKS1 is a decarboxylase that converts the 3-keto acids to the corresponding C11 and C13 2-methylketones 2-undecanone and 2-tridecanone (Yu et al., 2010). The proscription of methylketone biosynthesis to trichomes could be an adaptation that concentrates these compounds on the surface of the leaves and easily delivers them onto any foreign animal that makes contact with the leaves, and also to isolate such toxic compounds from the rest of the plant (Schilmiller et al., 2008). In addition, since the enzymes that make methylketones compete with substrates for fatty acid biosynthesis, synthesis in the trichomes could be a way to avoid competition for substrates in all other plant cells.

Transcripts of *ShMKS1* and *ShMKS2* in *S. habrochaites* plants are in 800-fold and 300-fold greater abundance than the levels of the transcripts of the orthologous genes, *SlMKS1a* and *SlMKS2c*, in the cultivated tomato (Fridman et al., 2005; Ben-Israel et al., 2009). Consistent with this observation, methylketones in the trichomes of cultivated tomato plants are present at low and often undetectable levels. However, the SlMKS1a and SlMKS2c proteins are highly similar to the *S. habrochaites* proteins (95% and 91% identical), and expressing the cultivated tomato *SlMKS2c* gene in *E. coli* resulted in similar profile of methylketones production as expressing *ShMKS2* (Ben-Israel et al., 2009). It was therefore deemed possible that the extremely low levels of methylketone production in the trichomes of cultivated tomato are due mainly to the low levels of expression of its *MKS*1 and *MKS2* genes, although other genetic factors have been implicated in enhanced methylketone production in wild tomatoes (Ben-Israel et al 2010). To test this hypothesis, we introduced *ShMKS1* and *ShMKS2* genes into cultivated tomato

under the control of their own promoters, which are active specifically in tomato trichomes. In addition, to test if methylketone production could be introduced to non-*Solanum* species and non-trichome tissues, tobacco (*Nicotiana tabacum*) and Arabidopsis plants were transformed with *ShMKS1* and *ShMKS2* (or both) under the control of constitutively active promoters. Here we report that constitutive expression of the two genes in transgenic tobacco and Arabidopsis, but not in tomato, resulted in methylketone production. Constitutive expression of *ShMKS2*, either with or without *ShMKS1*, in the three plant species tested here led to myristic acid accumulation *in planta*. However, trichome-specific expression of methylketone synthases 1 and 2 did not lead to detectable accumulation of methylketones in the trichomes of transformed cultivated tomatoes.

4.3 RESULTS

4.3.1 Transgenic Arabidopsis plants overexpressing *ShMKS2*, but not *ShMKS1*, exhibit lesions

To generate transgenic Arabidopsis lines expressing *ShMKS1* or *ShMKS2*, we spliced these two genes into expression vectors downstream of the cauliflower mosaic virus *35S* promoter (designated *p35S::ShMKS1* and *p35S::ShMKS2* respectively). These vectors were then introduced into wild-type *A. thaliana* (ecotype Columbia-0) by *Agrobacterium*-mediated transformation. To generate transgenic Arabidopsis lines expressing both *ShMKS1* and *ShMKS2*, the ORF of *ShMKS1* was fused to a RuBisCo (Ribulose-1, 5-bisphosphate carboxylase oxygenase) small subunit promoter (designated *pRbcs*) and added to the *p35S::ShMKS2* expression vector. The vector carrying both genes, designated *p35S::ShMKS2-pRbcs::ShMKS1,* was then transferred into wild-type Arabidopsis (Col-0). For all transformants, the presence of intact ORF sequences of

ShMKS1 and/or *ShMKS2* was verified by genomic DNA PCR. RT-PCR was also performed to confirm the expression of these two genes, and only lines with clear gene expression were chosen for further analysis. Four independent lines of transformants expressing *ShMKS1*, seven independent lines expressing *ShMKS2* and four independent lines expressing both genes were selected. Seeds from the transgenic lines were collected, germinated on Kanamycin-containing MS (Murashige and Skoog) plates, and further analyzed.

All the Arabidopsis plants expressing *p35S::ShMKS1* were morphologically indistinguishable from wild-type Arabidopsis (Figure 4-1A,D,G,H,K,L). On the other hand, both the *p35S::ShMKS2* and the *p35S::ShMKS2-pRbcs::ShMKS1* expressing lines exhibited growth retardation and leaf lesions by 8 days after germination (8 DAG). The first pair of true leaves of the *p35S::ShMKS2* plants initiated but was not capable of expanding fully to their maturity (Figure 4-1B,E), and they remained colorless, while the cotyledons also developed regional chlorosis (Figure 4-1B,E). In plants expressing both *ShMKS1* and *ShMKS2*, the first true leaves developed properly but the cotyledons showed chlorosis before the second pair of true leaves emerges (Figure 4-1C,F).

The majority of seedlings expressing *ShMKS2* and both *ShMKS1* and *ShMKS2* could not survive after transfer to soil, and those that survived grew at much slower rate than did wild type or plants expressing *ShMKS1* (Figure 4-1G-J). However, the severity of the chlorosis abated as the plants developed, although these plants still had a reduced growth rate (Figure 4-1K-N).

Figure 4-1. Growth phenotypes of transgenic Arabidopsis plants expressing *ShMKS1*, *ShMKS2*, or both genes

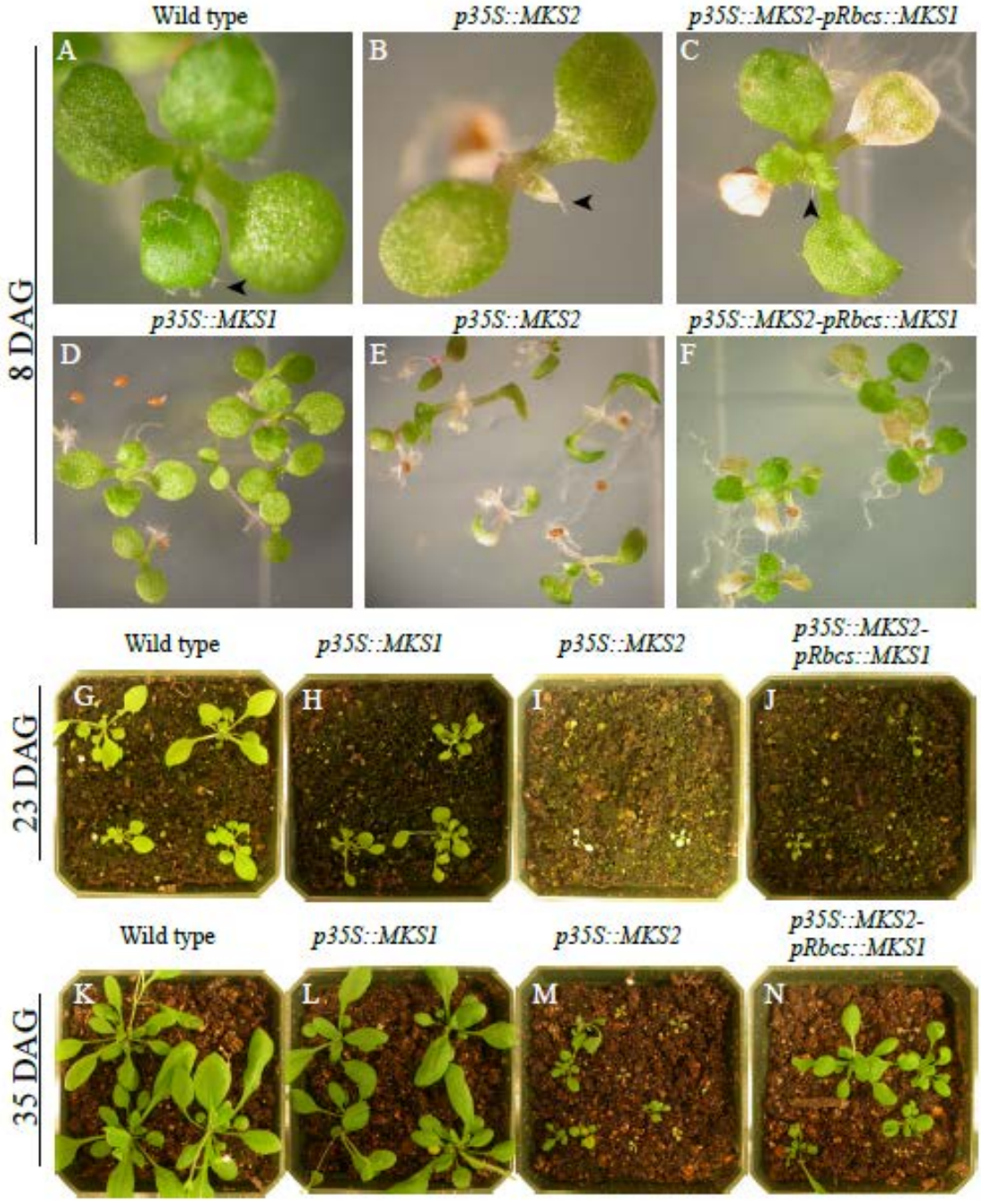

Figure 4-1. Growth phenotypes of transgenic Arabidopsis plants expressing ShMKS1, ShMKS2, or both genes. A-F, 8 day-after-germination (DAG) representative pictures of Arabidopsis seedlings grown on plates. A, wild type; B and E, *p35S::ShMKS2* lines; C and F, *p35S::ShMKS2-pRbcs::ShMKS*1 lines; D, *p35S::ShMKS1*. G to I are pictures of 23 DGA plants, K to N are pictures of 35 DAG plants. G and K, wild-type; H and L, *p35S::ShMKS1* lines; I and M, *p35S::ShMKS2* lines; J and N, *p35S::ShMKS2-pRBcs::ShMKS1* lines. The length of the square pots is 9 cm. True leaves are those with trichomes, as indicated by the arrow heads in the three panels in the top row.

4.3.2 Production of myristic acid and other fatty acids in transgenic Arabidopsis overexpressing *ShMKS2*

Because ShMKS2 is a thioesterase that utilizes fatty acids biosynthetic intermediates, and fatty acids are involved in multiple aspects of plant growth (Ohlrogge and Browse, 1995), the observed growth defects of transgenic Arabidopsis overexpressing *ShMKS2* are likely due to interruption of fatty acids biosynthesis. To test this, both free fatty acids and O-acyl lipids from fresh leaves of seedlings at 8 DAG, and from rosette leaves at 35 DAG from various transgenic lines were trans-methylated or acid-catalyzed trans-esterified, and the fatty acids methyl esters (FAME) extracted and analyzed by gas-chromatograph and mass-spectrum (GC-MS). For both transgenic and wild type plants, higher concentrations of fatty acids per gram FLW (fresh leaf weight) were detected at 8 DAG than at 35 DAG (Table 4-1). Plants overexpressing *ShMKS1* showed minor differences in each fatty acid concentration or in total fatty acid concentration compared to wild-type. Both the *p35S::ShMKS2* and *p35S::ShMKS2-pRbs::ShMKS1* lines had elevated levels of palmitic acid and myristic acid at 8 DAG, with the lines expressing both genes having the highest level (Table 4-1). At 35 DAG, trace amounts of myristic acid were detected only in plants overexpressing both *ShMKS1* and *ShMKS2*, and only minor differences in the relative levels of other fatty acids were observed between different plant lines (Table 4-1).

Table 4-1. Fatty acids compositions of transgenic and wild type Arabidopsis plants in early and late developmental stages

	Fatty acids (mmol/gram fresh leaf weight)					
Plant lines	C14	C16	C16-3	C18-1	C18-3	Total
Wild type Y[1]	N.D.[2]	10.7±1.76[3]	12.2±3.55	7.0±1.09	36.5±5.53	66.4±3.08
Wild type O[4]	N.D.	5.8±0.98	4.2±0.70	3.9±0.49	17.7±0.38	31.6±0.60
p35S::ShMKS1 Y	N.D.	13.7±0.28	12.8±3.04	8.0±2.42	44.2±5.95	78.8±3.18
p35S::ShMKS1 O	N.D.	4.0±0.24	3.9±0.19	2.4±0.76	14.8±2.41	25.1±1.14
p35S::ShMKS2 Y	0.2±0.05	15.6±2.10	17.7±5.46	9.5±1.40	44.4±2.45	87.5±2.90
p35S::ShMKS2 O	N.D.	4.9±1.29	4.1±0.83	3.3±1.42	15.8±4.06	28.2±2.04
p35S::ShMKS2-pRbc::ShMKS1 Y	1.7±0.70	23.5±3.09	17.9±7.09	10.8±1.66	47.0±5.08	103.3±4.22
p35S::ShMKS2-pRbc::ShMKS1 O	0.1±0.01	4.6±0.45	3.5±0.44	2.6±0.70	13.8±1.76	24.6±0.89

[1] Y, young plants, 8 DAG
[2] N.D., not detected
[3] The standard errors were calculated from three biological measurements
[4] O, old plants, 35 DAG

4.3.3 Production of Methylketones in transgenic Arabidopsis plants expressing both *ShMKS1* and *ShMKS2*

Volatiles from 8 DAG transgenic seedlings growing on plates were collected by solid phase micro-extraction (SPME) fiber and analyzed by GC-MS. Wild-type Arabidopsis seedlings emitted very little volatile compounds (Figure 4-2). Three compounds were easily distinguished from lines expressing both *ShMKS1* and *ShMKS2*, but not from plants expressing only *ShMKS1* or only *ShMKS2*, compared to wild-type plants (Figure 4-3A). Compounds 1 and 3 were identified as 2-tridecanone and 2-pentadecanone, respectively, as their retention time and MS both matched that of the standards (Figure 4-3). Compound 2 had a similar MS profile to those of straight-chain 2-alcohols, but had a retention time of 14.3 min and a molecular weight of 199, which is between the standards of 2-dodecanol (13.1 min and MW 185) and 2-tetradecanol (15.3

min and MW 213, Figure 4-3B). Thus, compound 2 was likely to be 2-tridecanol. The methylketone production rate by 8 DAG seedlings expressing both *ShMKS1* and *ShMKS2* was roughly 1.5 μmol per gram FLW in a 5-hour period (Table 4-2). However, no methylketone emission could be detected from any transformants after they were transferred to soil (data not shown).

Figure 4-2. Chromatography of Wild-type Arabidopsis (A) and *S. habrochaites* PI126449 (B)

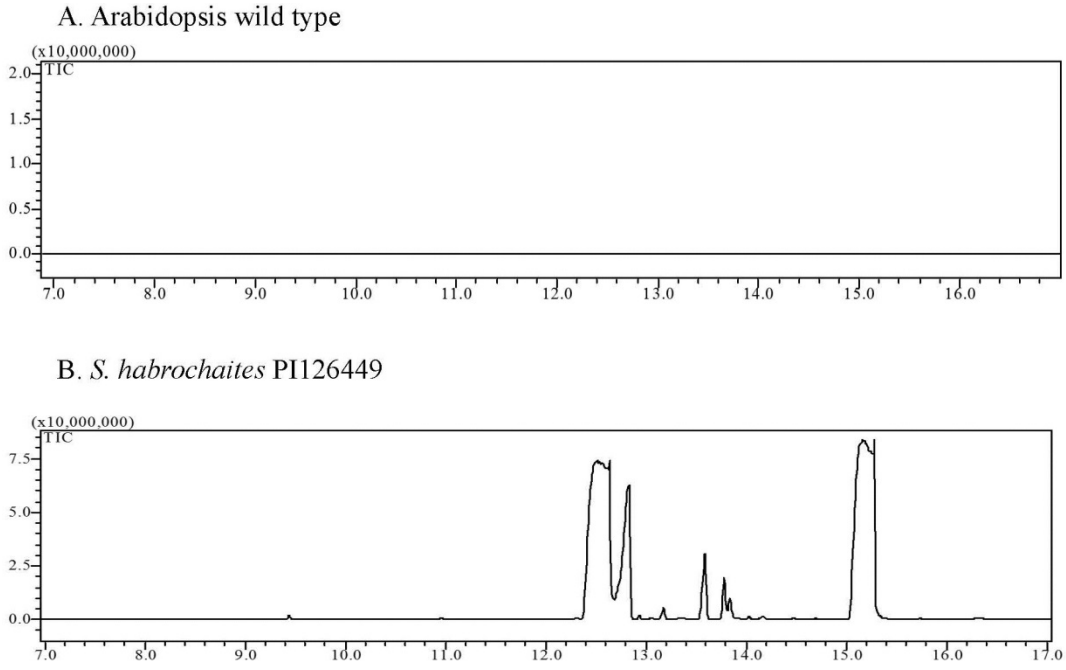

Figure 4-2. Chromatography of Wild-type Arabidopsis (A) and *S. habrochaites* PI126449 (B). Volatiles were collected by the needle of a solid phase micro-extraction (SPME) device though a small hole on a screw lid that secures a vial containing 100 mg of leaves for 5 hours. Next, the needle was withdrawn and injected into the GC device.

Figure 4-3. Chromatography and mass spectra of transgenic Arabidopsis volatiles

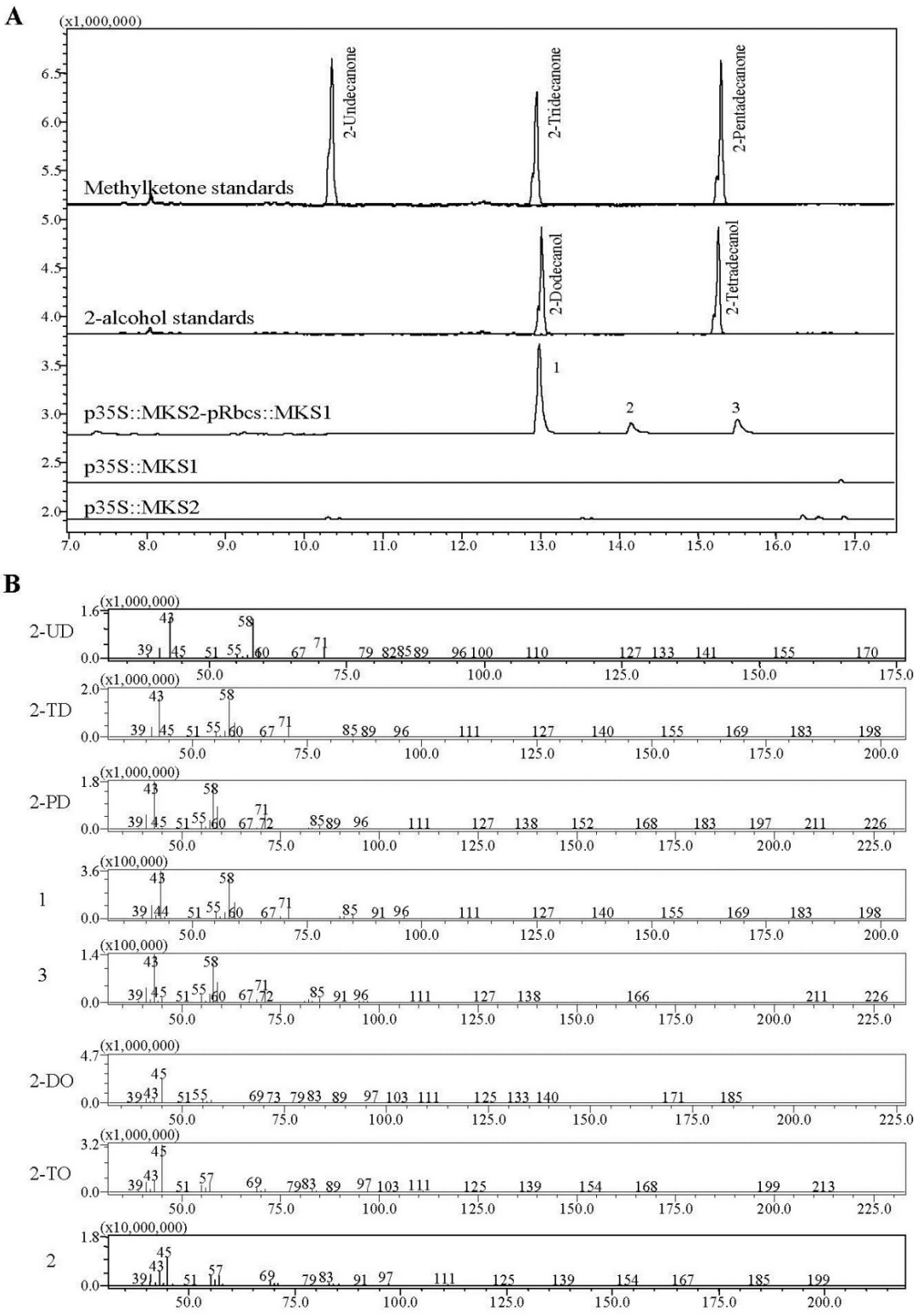

Figure 4-3. Chromatography and mass spectra of transgenic Arabidopsis volatiles. A. The top chromatograph line shows the retention time for methylketone standards of 2-undecanone, 2-tridecanone and 2-pentadecanone. The second top chromatograph line shows the retention time of 2-dodecanol and 2-tetradecanol. The bottom three chromatograph lines are 5 h volatile collection of transgenic Arabidopsis of *p35S::ShMKS2-pRbcs::ShMKS2*, *p35S::ShMKS2* and *p35S::ShMKS1* lines. B. Mass spectra of the compounds indicated in A. 2-UD, 2-undecanone standard; 2-TD, 2-tridecanone standard; 2-PD, 2-pentadecanone standard; 2-DO, 2-dodecanol standard; 2-TO, 2-tetradecanol standard. Compounds 1, 2 and 3 in (A) were identified as 2-tridecanone, 2-tridecanol and 2-pentadecanone, respectively.

Table 4-2. The amount of methylketones emitted by transgenic plant seedlings (8 DAG) in a 5-hour period

	Methylketone amount (μmol/g fresh leaf weight/5h)			
Plant lines	2-Undecanone	2-Tridecanone	2-Tridecanol	2-Pentadecanone
Arabidopsis *p35S::ShMKS2-pRbcs::ShMKS1*	N.D.[1]	1.08±0.96	0.35±0.16	0.41±0.22
Tobacco *p35S::ShMKS2*	N.D.	2.06±0.22	0.36±0.03	0.40±0.05
Tobacco *p35S::ShMKS2-pRbcs::ShMKS1*	0.71±0.03	8.62±1.40	1.58±0.42	2.02±0.46

[1]N.D., not detected

4.3.4 Analysis of transgenic tobacco plants overexpressing *ShMKS1* and/or *ShMKS2*

To further examine if over-expression of *ShMKS1* and *ShMKS2* is sufficient to cause methylketone production in other plant species, the same constructs, *p35S::ShMKS2-pRbcs::ShMKS1* and *p35S::ShMKS2*, were introduced into wild-type *N. tabacum*. *p35S::ShMKS1* was not introduced into tobacco because the Arabidopsis lines expressing *ShMKS1* showed no detectable changes. Five independent tobacco lines overexpressing both *ShMKS1* and *ShMKS2* and four lines overexpressing only *ShMKS2*

were selected. Transgenic tobacco plants with either of the two constructs displayed chlorosis and malformation of the first pair of true leaves when growing on kanamycin-containing MS plates (Figure 4-4). Tobacco plants expressing both *ShMKS1* and *ShMKS2* displayed a larger area of chlorosis at 8 DAG and stunted growth at 45 DAG (Figure 4-4A-C and G-I). The first pair of true leaves failed to fully develop by 20 DAG for both constructs (Figure 4-4D-F).

Figure 4-4. Growth phenotypes of transgenic tobacco plants expressing *ShMKS2* or co-expressing *ShMKS1* and *ShMKS2*

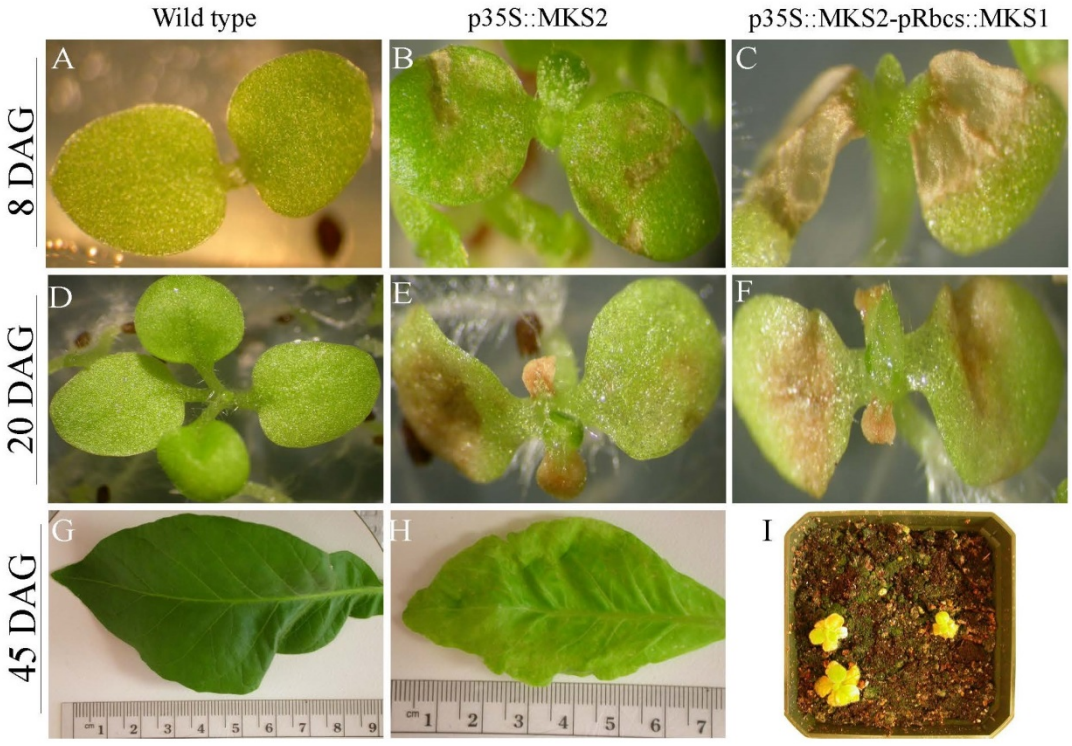

Figure 4-4. Growth phenotypes of transgenic tobacco plants expressing *ShMKS2* or co-expressing *ShMKS1* and *ShMKS2***.** Left column, wild-type tobaccos; middle column, *p35S::ShMKS2* transformed tobaccos; right column, *p35S::ShMKS2-pRbcs::ShMKS1* transformed tobaccos. A-C, 8 DAG tobacco seedlings; D-F, 20 DAG tobacco seedlings; G-I, 45 DAG tobacco leaves. G,H, the 9th true leaf; I, whole plants. The length of the square pot is 9 cm.

Fatty acids from fresh leaves of 8 DAG seedlings were transmethylated and extracted, then analyzed by GC-MS. Tobacco lines expressing only *ShMKS2* had higher levels of myristic acid and palmitic acid per gram FLW than tobacco lines overexpressing both *ShMKS1* and *ShMKS2* (Table 4-3). However, tobacco lines expressing only *ShMKS2* had lower amounts of Δ2 and Δ3 unsaturated C18 fatty acids and lower overall fatty acid levels compared to wild-type plants (Table 4-3). Tobacco plants expressing both *ShMKS1* and *ShMKS2* had only trace amounts of myristic acid and displayed minor differences in fatty acids profiles compared to wild-type plants (Table 4-3).

Table 4-3. Fatty acids compositions of transgenic and wild type tobacco seedlings (8 DAG)

	Fatty acids (mmol/g fresh leaf weight)						
Plant lines	C14	C16	C18	C18-1	C18-2	C18-3	Total
Wild type	N.D.[1]	0.75±0.39	0.14±0.10	0.08±0.05	1.21±0.77	3.65±1.43	5.83±0.57
p35S:ShMKS2	0.31±0.09	1.40±0.44	0.31±0.11	0.10±0.03	0.89±0.20	0.79±0.32	3.79±0.16
p35S:ShMKS2-pRbcs1:ShMKS1	0.07±0.03	1.12±0.38	0.30±0.12	0.16±0.08	2.15±0.50	3.01±1.88	6.81±0.70

[1]N.D., not detected

Volatile analysis of tobacco seedlings expressing both *ShMKS1* and *ShMKS2* revealed that they emitted methylketones (Figure 4-5). In addition to the same three compounds observed in Arabidopsis (2-tridecanone, 2-tridecanol and 2-pentadecanone), tobacco volatiles contained small amounts of several other compounds (numbered 4-8 in Figure 4-5A). Compound 4 was identified as 2-undecanone by its mass spectrum (Figure 4-5B). Compounds 5-8 could not be identified. In contrast to Arabidopsis, where transgenic plants expressing only *ShMKS2* showed no methylketones emission, the

tobacco counterpart lines did emit some methylketones (Figure 4-5A), with 2-tridecanone, 2-tridecanol and 2-pantadecanone as the three major compounds. However, the transgenic tobacco plants expressing both genes had a faster methylketone production rate per FLW than the tobacco lines expressing only *ShMKS2*, 11.35 μmol/g fresh leaf weight/5h vs. 1.49 μmol/g fresh leaf weight/5h (Table 4-2).

Figure 4-5. Chromatography and mass spectrum of transgenic tobacco volatiles

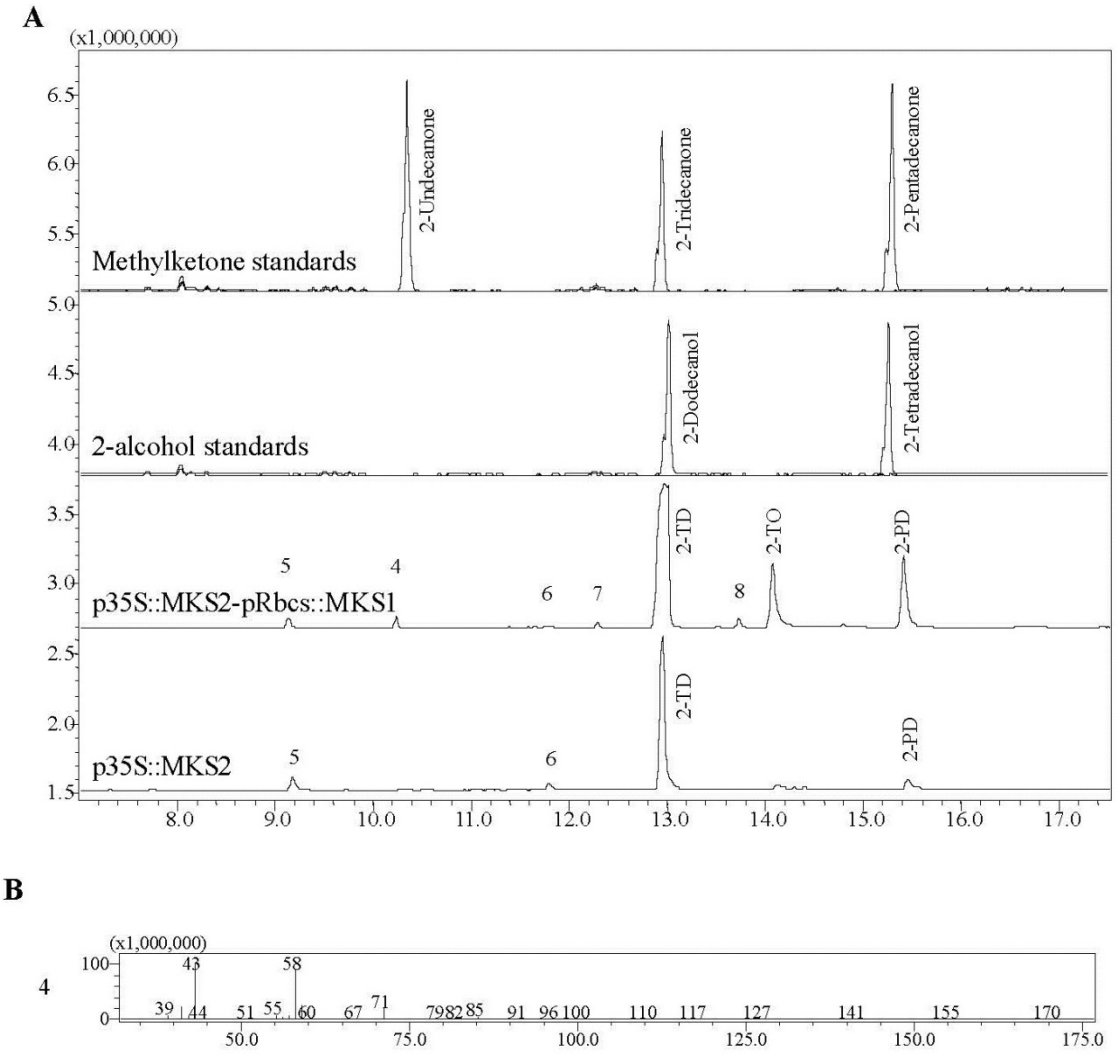

Figure 4-5. Chromatography and mass spectrum of transgenic tobacco volatiles. A. The top two chromatographs are the same as shown in Figure 2. The bottom two lines are chromatographs of 5 h volatile collection of a tobacco line expressing both *ShMKS1* and *ShMKS2* and a line expressing only *ShMKS2*. 2-TD, 2-tridecanone, 2-TO, 2-tridecanol, 2-PD, 2-pentadecanone. B. Mass spectrum of compound 4 in A. Compound 4 was identified as 2-undecanone.

The transgenic tobacco plants also suffered high lethality rate when transferred to soil and displayed growth retardation similar to the Arabidopsis *ShMKS2*-expressing lines (data not shown). At 45 DAG, tobacco plants expressing only *ShMKS2* showed attenuated chlorosis but fully extended leaves, whereas tobacco plants expressing both *ShMKS1* and *ShMKS2* still displayed severe growth retardation (Figure 4-4G-I).

4.3.5 Production of transgenic tomatoes with trichome-specific expression of *ShMKS1* and *ShMKS2*

We recently showed that the promoter of *ShMKS1* is highly active in type VI trichomes of cultivated tomato (Akhtar et al., 2013). To test the activity of the *ShMKS2* promoter in cultivated tomato, we made a construct that fused 1.5 kb upstream region of *ShMKS2* with GFP and transformed tomato plants. In these transgenic plants, strong GFP signal was observed in type VI trichomes as well as some weak GFP signal in other types of trichomes (Figure 4-6A-C), indicating that the *ShMKS2* promoter is strongly active in these trichomes in the cultivated tomato as well. Next, we generated transgenic tomato plants carrying both *ShMKS1* and *ShMKS2* each under the control of its own promoter using *pShMKS2::ShMKS2-pShMKS1::ShMKS1* construct (see Materials and methods). In addition, plants carrying only *ShMKS2* under the control of its own promoter (*pShMKS2::ShMKS2*) and plants carrying *p35S::ShMKS2* were also obtained. RT-PCR analysis of total RNA from isolated trichomes of transgenic tomato lines expressing

pShMKS2::ShMKS2-pShMKS1::ShMKS1 showed increased transcript levels of both *ShMKS2* and *ShMKS1* in trichomes but not in the leaves without trichomes (Figure 4-6D).

Figure 4-6. Trichome-specific expression of *ShMKS1* and *ShMKS2*

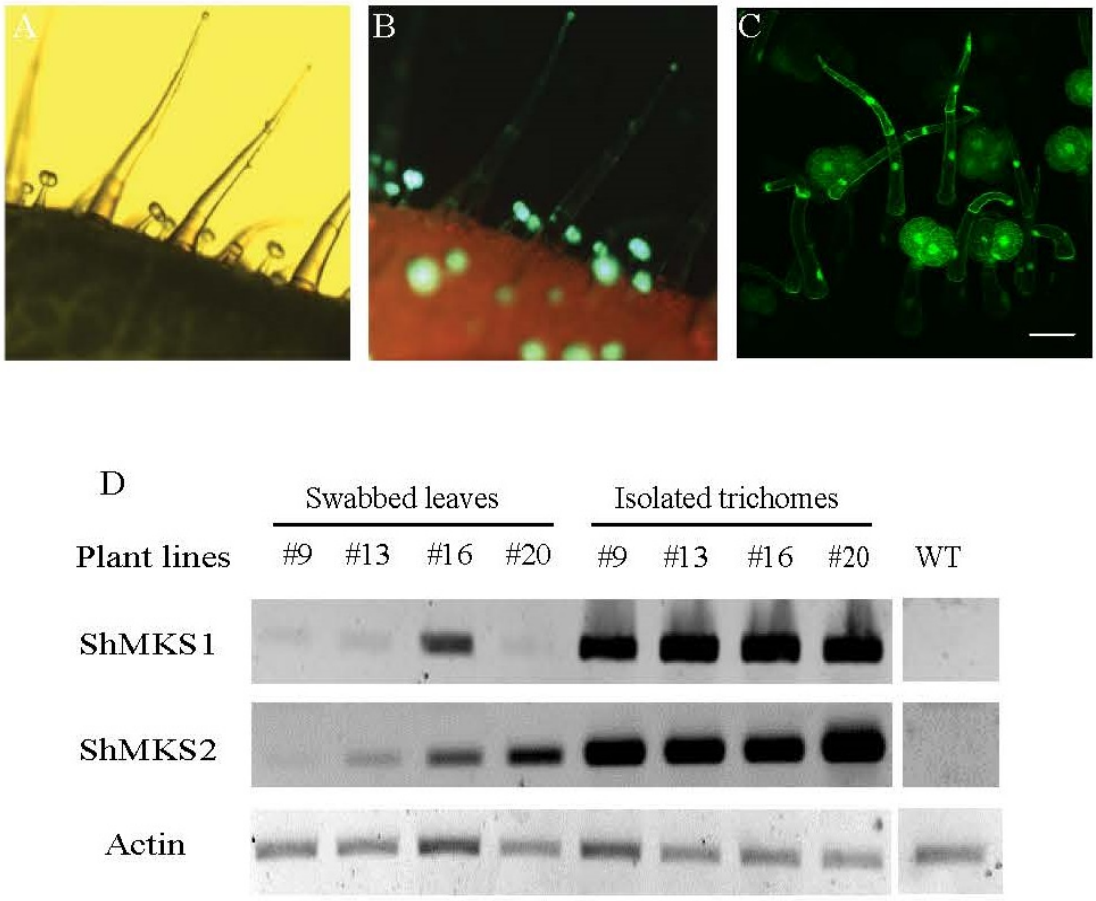

Figure 4-6. Trichome-specific expression of *ShMKS1* and *ShMKS2*. A and B, fluorescence anatomy microscope view of tomato expressing *pShMKS2::GFP*, under bright-field view (A) and green-fluorescence view (B). C. Confocal microscope view of *pShMKS2::GFP* transformed tomato leaves showed the GFP signal only in the trichomes. D. RT-PCR amplification of total RNA isolated from swabbed leaves (trichome removed), isolated trichomes and wild-type tomato whole leaf. The numbers with # indicate different independent lines of double-gene overexpression lines. *ShMKS1*, *ShMKS2* and *actin* specific oligonucleotides were used for amplification form different RNA samples.

All the tomato transgenic lines, six expressing only *ShMKS2* and seven expressing both *ShMKS1* and *ShMKS2* in the trichomes, were morphologically indistinguishable from wild type tomatoes and they emitted no methylketones from their trichomes (data not shown). In addition, no changes in fatty acid profiles or myristic acid were detected from the leaves of plants expressing the two *ShMKS* genes in trichomes (Table 4-4). However, tomato plants expressing *p35S::ShMKS2* showed lesions in their leaves (Figure 4-7), accumulation of myristic acid and elevated level of palmitic acid (Table 4-4), but they emitted no methylketones (data no shown).

Table 4-4. Fatty acids compositions of transgenic and wild type tomato seedlings (15 DAG[1])

	Fatty acids methyl esters amount (mmol/g fresh-leaf-weight)						
Plant lines	C14	C16	C16:3	C18	C18:2	C18:3	Total
S. habrochaites PI126449	0.06±0.01	2.02±0.05	0.61±0.09	0.29±0.01	2.72±0.18	4.99±0.86	10.68±0.33
S. lycopersicum	N.D.[2]	1.98±0.26	0.69±0.14	0.14±0.01	1.59±0.13	5.50±0.55	9.90±0.20
P35S:ShMKS2	0.10±0.01	2.13±0.11	0.28±0.05	0.18±0.01	1.52±0.12	3.65±0.33	7.75±0.13
pShMKS2:ShMKS2	N.D.	1.56±0.15	0.30±0.05	0.13±0.01	0.82±0.16	3.03±0.30	5.83±0.11
pShMKS2:ShMKS2 - *pShMKS1:ShMKS1*	N.D.	1.40±0.07	0.27±0.01	0.12±0.01	0.77±0.07	2.78±0.04	5.34±0.03

[1] An 8 DAG tomato plant has too little true leaf materials, so 15 DAG seedlings were analyzed.

[2] N.D., not detected

Figure 4-7. Leaf phenotype of transgenic tomato plants expressing *p35S::ShMKS2*

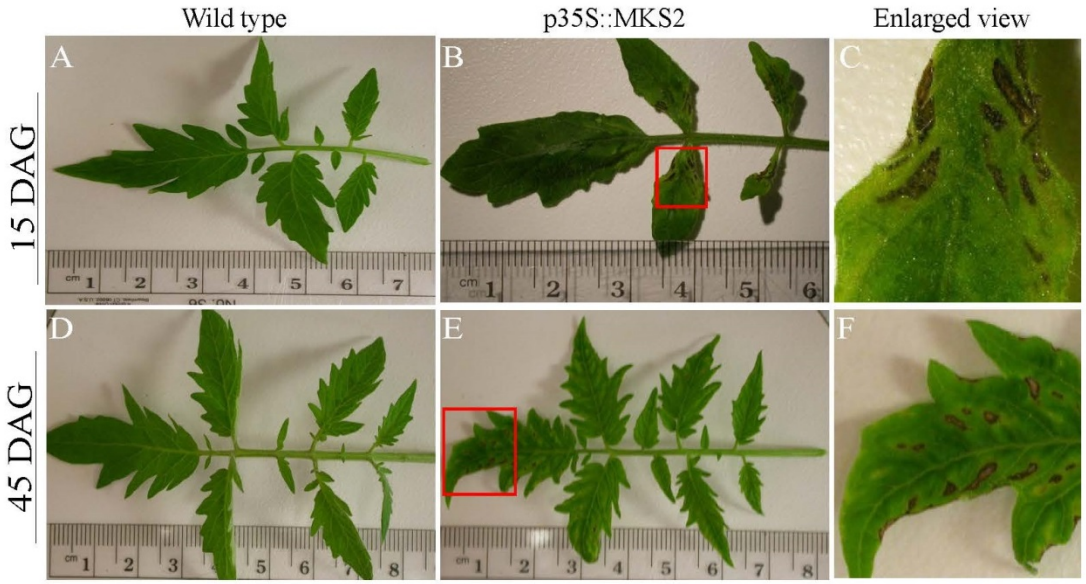

Figure 4-7. Leaf phenotype of transgenic tomato plants expressing *p35S::ShMKS2*. Left column, wild-type tomato leaves; middle column, *p35S::ShMKS2* transformed tomato leaves; right column, enlarged portion of the area enclosed in the red box in the middle column. A-C, the 3rd true leaf at 15 DAG, D-F, the 9th true leave at 45 DAG.

4.4 DISCUSSION

4.4.1 Trichome specific co-expression of both *ShMKS1* and *ShMKS2* is insufficient for methylketone production in cultivated tomato

Methylketone production and accumulation in wild tomato *S. habrochaites* PI126449 renders it resistant to many of cultivated tomato pests (Gonçalves et al., 1998). Introducing methylketone production into cultivated tomato is thus a desired goal for tomato breeders. However, previous tomato breeding efforts involving interspecific crosses showed it was difficult to breed tomato lines with high levels of methylketone production in their trichomes (Hartman and St. Clair, 1998), consistent with early genetic

analysis showing that the methylketone production trait in tomato was associated with multiple gene loci (Zamir et al., 1984). Recent genetic and molecular analyses of *S. lycopersicum* X *S. habrochaites* hybrid progeny showed that that transcripts of several fatty acids biosynthetic genes were present at lower concentrations in the trichomes of cultivated tomato than in wild tomato, and that maximal levels of methylketones depended on the presence of *S. habrochaites* alleles at several fatty acids biosynthetic gene loci as well as at the *MKS1* and *MKS2* gene loci (Ben-Israel et al., 2009). Yet, because the expression levels of *ShMKS1* and *ShMKS2* in *S. habrochaites* PI126449 are 800-fold or 300-fold higher than in the cultivated tomato and the differences in levels of transcripts of other fatty acid-related genes in the trichomes was only <8-fold between the two species, it was thought possible that the introduction of *ShMKS1* and *ShMKS2* alone might lead to increases in methylketone production in cultivated tomato plants.

Since *ShMKS1* and *ShMKS2* show trichome-specific expression in *S. habrochaites* PI126449, the promoters of these two gene loci were amplified and fused in frame with their own ORF for trichome-specific expression in *S. lycopersicum*. However, no methylketones were detected in cultivated tomato lines expressing both *ShMKS1* and *ShMKS2* in their trichomes. Thus, it appears that introducing only *ShMKS1* and *ShMKS2* into tomato for trichome-specific expression without also increasing fatty acid biosynthetic flux is not sufficient for methylketone production in the trichomes. Successful engineering of tomato trichomal methylketone production will require increases in metabolic flux towards fatty acid biosynthetic pathway in trichomes, possibly by introducing the *S. habrochaites* fatty acid biosynthetic gene loci to cultivated tomato, in addition to co-expression of *ShMKS1* and *ShMKS2*.

4.4.2 Methylketone production in transgenic plants expressing *ShMKS2* under the control of the general promoter 35S, with or without co-expression of *ShMKS1*

Co-expression of *ShMKS1* and *ShMKS2* in Arabidopsis and tobacco led to methylketone production *in planta*, indicating that methylketone production could be engineered in other plant species. In a five-hour period, transgenic tobacco plants expressing both *ShMKS1* and *ShMKS2* produced 11.35 μmol/g FLW methylketones, while the Arabidopsis counterparts produced only 1.49 μmol/g FLW (Table 4-2).

In tobacco, expression of *ShMKS2* alone also led to the synthesis of methylketones at 2.46 μmol/g FLW/5 h, which was 80% lower than the rate of methylketone production by transgenic tobacco plants expressing both *ShMKS1* and *ShMKS2*. On the other hand, transgenic Arabidopsis and tomato plants expressing only *ShMKS2* under the control of the 35 promoter emitted no methylketones (Table 4-2). The action of the thioesterase ShMKS2 on its *β*-ketoacyl-ACP substrates leads to the production of 3-keto acids, which are known to be unstable to and to spontaneously decarboxylate to give the corresponding 2-methyketones (Matiasek et al., 2001). A previous study showed that expressing *ShMKS2* in *E. coli* also leads to the appearance of methylketones (Yu et al., 2010). In addition, the spent medium of the bacteria accumulated 3-keto acids (Yu et al., 2010). These observations demonstrate that the action of the decarboxylase ShMKS1 is not absolutely required to produce methylketones. However, the expression of *ShMKS1* in addition to *ShMKS2* led to a five-fold increase in

methylketone production in transgenic tobacco, consistent with the results of the co-expression of the two genes in *E. coli* (Park et al., 2012).

The fate of the 3-keto acids produced by the action of ShMKS2 *in planta* and not converted to methylketones is not clear. We could not detect any 3-keto acids in Arabidopsis, tobacco and tomato plants expressing *ShMKS2* (data not shown). These plants, except tobacco, emitted no methylketones. Yet, since Arabidopsis as well as tobacco plants expressing both *ShMKS1* and *ShMKS2* emit methylketone, a considerable amount of 3-keto acids must have been generated in plants expressing *ShMKS2* alone. The 3-keto acids may have been catabolized by the plants at a faster rate than their spontaneous degradation rate, which could also explain the lower methylketone productivity of tobacco plants expressing only *ShMKS2* than that of tobacco plants expressing both *ShMKS1* and *ShMKS2*.

4.4.3 ShMKS2 is capable of hydrolyzing myristoyl-ACP in addition to 3-ketoacyl-ACP

Transgenic Arabidopsis and tobacco plants expressing *ShMKS2* together with *ShMKS1* produced similar methylketones, and at similar ratios, as those of the wild tomato *S. habrochaites* PI126449 (Figure 4-2), with 2-tridecanone being the major methylketone in all cases. This suggests that the substrate chain-length selectivity of ShMKS2 is not changed when heterologously expressed in other plant species. Interestingly, production of myristic acid was observed in all three plant species expressing *ShMKS2* under the 35S promoter (Tables 4-1, 4-3, 4-4), indicating that ShMKS2 is capable of hydrolyzing myristoyl-ACP as well. An examination of the trichomes of *S. habrochaites* accession PI126449 showed that low levels of myristic acid

were found in these trichomes as well (Table 4-4). Thus, while ShMKS2 is mostly a *β*-ketoacyl-ACP thioesterase preferring the C14 *β*-ketoacyl-ACP substrate, it is also capable of hydrolyzing a fully reduced myristoyl-ACP although evidently at a lower rate.

4.4.4 Constitutive expression of *ShMKS2 in planta* results in phenotypes similar to those of Arabidopsis mutants with defects in *de novo* fatty acids biosynthesis

The morphological and developmental phenotypes observed for *ShMKS2*-expressing plants were quite similar to those of Arabidopsis mutants that are deficient in *de novo* fatty acids biosynthesis, such as Arabidopsis acyl-ACP thioesterase *FATB* mutant, enoyl-acyl-ACP (ENR) mutant *MOD1* and *β*-ketoacyl-ACP synthase I (*KAS1*) mutant (Mou et al., 2000; Bonaventure et al., 2003; Wu and Xue, 2010), all of which display growth retardation, variegated leaves and reduced fertility. It is also interesting to note that the deleterious effects of the *KASI* mutant were alleviated when the plants matured (Wu and Xue, 2010), similar to our observations with older Arabidopsis and tobacco plants expressing *ShMKS2* (Figure 4-1K-N, Figure 4-4G-I). The phenotypes observed are consistent with the function of ShMKS2, which is a thioesterase utilizing the intermediates in fatty acids biosynthesis as substrates (Yu et al., 2010). Because ShMKS2 prefers *β*-ketomyristoyl-ACP (Yu et al., 2010), constitutive expression of *ShMKS2* will terminate the fatty acids biosynthesis prematurely and reduce its flux, which leads to plant developmental malfunction. In addition, the 3-ketoacids or their metabolites, including methylketones, might also lead to toxic effects on the plant, which could explain the persistent developmental problems seen in these plants, particularly transgenic tobacco plants expressing both *ShMKS1* and *ShMKS2* that made high levels of methylketones (Figure 4-4G-I). Thus, attempts to engineer high-level production of

methylketones throughout the plant are likely to fail. To avoid the deleterious effects caused by interference with fatty acid biosynthesis and the production of methylketones, the expression of *ShMKS1* and *ShMKS2* must be directed to a small, self-contained structure such as trichomes or root nodules.

4.5 MATERIALS AND METHODS

4.5.1 Plasmids construction

Plasmids for plant transformation were constructed by employing the pSAT vector system (Tzfira et al., 2005). The full length coding sequences were end-fused in frame with *Eco*RI and *Sal*I restriction sites for *ShMKS1* and *Bgl*II and *Sal*I sites for *ShMKS2* by PCR. Next, these constructs were spliced into *pSAT4A-35S* to build *p35S::ShMKS1* and *p35S::ShMKS2*. The *ShMKS1* fragment was also transferred to *pSAT6A-rbc* to form *pRbcs::ShMKS1*. The 35S promoter of *p35S::ShMKS2* was exchanged for the promoter sequence of *ShMKS2* by *Age*I and *Bgl*II to build the *pShMKS2-MKS2* construct. The *pShMKS1-MKS1* construct was made by switching the *Rbcs* promoter with the promoter sequence of *ShMSK1* with *Age*I and *Eco*RI. The *pShMKS2::GFP* construct was obtained by splicing the promoter sequence of *ShMSK2* into *Age*I and *Bam*HI restriction sites of *pSAT6A-EGFP* construct to replace the 35S promoter (Tzfira et al., 2005).

All the pSAT4A expression cassettes were spliced into the I-*Sce*I site and pSAT6A expression cassettes were spliced into PI-*Psp*I site of the binary vector *pPZP-RCS2* (Tzfira et al., 2005). For *p35S::ShMKS2-pRbcs::ShMKS1* and

pShMKS1::ShMKS1-pShMKS2::ShMKS2, two expression cassettes were spliced in the same binary vector.

4.5.2 Plant materials and generation of transgenic plants

For Arabidopsis, wild type (Columbia 0) or transgenic seeds were surface sterilized with 70% ethanol for 10 minutes and rinsed with 100% ethanol to remove water. Then the seeds were left in open tubes at 37 ℃ to completely dry up the ethanol. The dry seeds were spread on half-strength Murashige and Skoog (Sigma-Aldrich Co. LLC.) phytoblend (Calsson labs Inc.) plates containing 1% sucrose. After 3 days of stratification at 4 ℃, plates were left in a growth room that was adjusted to 14 h light/10 h dark at 22 ℃. The same growth room condition was used for all plants mentioned in this study, unless otherwise stated. Seedlings 8 days after germination (DAG) were harvested for fatty acids analysis and the rest were transferred onto soil. Plants in soil were harvested at 35 DAG again.

For tobacco, *N. tabacum* cv. SR1, seeds were surface sterilized with 40% colorox bleach (Colorox Co.) for 15 mins and rinsed with sterilized water, then transferred (with some water) onto half-strength MS plates. Tobacco seedlings were transferred to soil at 15 DAG after volatiles had been collected.

For tomato, *S. lycopersicum* MP1 seeds were sterilized with 50% bleach for 20 min and then with 70% ethanol for 10 min. The seeds were then dried and spread onto half-strength MS plates.

All the binary vectors mentioned above were introduced into *Agrobacterium tumefaciens* strain EHA105 by heat-shock method. Arabidopsis was transformed by the

flower dip method (Clough and Bent, 1998). Tobacco transformation followed a leaf-disc infection method (McCormick, 1991). Tomato transformation was conducted at the Plant Transformation Core Research Facility (PTCRF) at the University of Nebraska Lincoln following the protocol detailed on their website (http://biotech.unl.edu/node/211).

For selection of transgenic plants, Arabidopsis, tobacco and tomato plants were selected on half strength MS plates containing 50 mg/L, 200 mg/L and 200 mg/L Kanamycin, respectively. *ShMKS1* was amplified by oligonucleotide-pair 5'-GAATTCATGGAGAAAAGCATGTCGCC-3' and 5'-GCGTCGACTTTATACTTGTTAGCGATGCT-3' and *ShMKS2* was using 5'-AGATCTATGAGTGATCAGGTCTATCACCATGA-3' and 5'-GTCGACTTTAGAGATGATGTTGTACACCGCA-3'

4.5.3 Trichome isolation and swabbed leaves

Trichomes were isolated by gently shaking the leaves taken right out of liquid nitrogen in a 50 mL corning tube. After discarding the leaves, the frozen trichomes were washed off the tube by the RNA extraction buffer and ready for isolation.

Trichomes on leaves were removed by using cotton swabs to bush the leaf surface several times.

4.5.4 RT-PCR analysis of transgenic plants

RNA was extracted from the leaves of wild type and the transgenic plants using EZNA. Plant RNA kit (Omega Bio-tek Inc.). During the RNA isolation, DNA was

removed by On-Column DNase I Digestion Set (Sigma-Aldrich Co. LLC) and then the purified RNA was reverse transcribed by High Capacity cDNA Reverse Transcription Kit (Life technologies Co.). PCR reactions were performed by GoTaq® Green Master Mix (Promega Co.).

4.5.5 Plant imaging

Plant phenotype images were taken through a Leica MZ6 anatomy microscope or directly by a Nikon camera (Coolpix 4500). The fluorescence anatomy microscope was Leica MZFLIII, with all factory-equipped devices and default settings for bright field and green fluorescence imaging. Confocal microscopy was detailed in Yu et al., 2010, with images taken through a 20* objective lens.

4.5.6 Fatty acids and methylketone analysis by GC-MS

Fatty acids analysis was followed the procedure described by Browse et al. (Browse et al., 1986) with small modification. Briefly, around 100 mg of fresh seedlings or mature plant leaf material were weighed and submerged in 1 mL 1N Methanol-HCL anhydrous solution (with 0.1 mM pentadecanoic acid as an internal standard) for 2 hours at 60 ℃ for fatty acids extraction and derivazation. After the reactions were cooled down to room temperature, 1 mL of 0.9% NaCI was added and then the mixture was extracted by 300 μL hexane. The hexane layer was used directly for GC-MS analysis. The hexane extract (3 μL) was injected in splitless mode into Shimadzu GCMS-QP5000 system equipped with an EC-WAX column (thickness, 0.25 μm; diameter, 0.32 mm; length, 30 m). The GC initial oven temperature was 50 ℃, then immediately ramped to 250 ℃ at

5 ℃/min and held at 250 ℃ for 1 min. The injector and detector were set at 250 ℃ and 280 ℃, respectively. The quantities of each fatty acid were calculated by comparing the peak areas with that of the pentadecanoic acid.

Methylketones were analyzed on the same GC-MS machine, but the oven temperature was held at 50 ℃ for 2 min, then ramped at 10 ℃/min to 190 ℃, followed by 5 ℃/min to 220 ℃ and finally 10 ℃/min to 275 ℃. Temperature was held at 275 ℃ for 3 min. Helium was the carrier gas used and the flow rate was 1.4 mL/min for both methods.

The identities of fatty acids, methylketones and the medium-chain 2-alcohols were determined by comparing their retention times and mass spectra with those of authentic chemical standards. The identity of 2-tridecanol was inferred from comparing its mass spectrum and retention time with that of 2-dodecanol and 2-tetradecanol.

Chapter 5 Specificity of two trichomal gland cell promoters from *Solanum* across species boundaries and evidence for the presence of plastids in gland cells[4]

5.1 ABSTRACT

5.1.1 Background

Many plant species have trichomes, the epidermal protuberances covering their aerial surface. Trichomes exhibit tremendous morphological diversity but can be roughly classified as glandular or non-glandular, with the glandular ones containing stalk cells topped by one or more gland cells. The glandular trichomes of tomato and tobacco have been heavily studied for their capacity to synthesize several classes of specialized chemicals. Cultivated tomato (*Solanum lycopersicum*) has two major types of glandular trichomes, called type I/IV and type VI. Tobacco (*Nicotiana tabacum*) also has two morphologically distinguishable types of glandular trichomes, long and short ones. Although a few trichomal gland cell-specific promoters have been isolated from both species, whether these promoters can confer specific expression in trichomal gland cells of other species has not yet been explored.

---

[4] This work has been submitted as Geng Yu, Anthony L. Schilmiller, Robert L. Last, Eran Pichersky. Specificity of two trichomal gland cell promoters from *Solanum* across species boundaries and evidence for the presence of plastids in gland cells. BMC Plant Biology. My contribution was to design, conduct experiments except the ones studying *pAT2* activity and drafting the manuscript.

5.1.2 Results

Reported here is an examination of the expression patterns of two trichomal gland cell-specific promoters, one from *S. lycopersicum* and the other from the wild tomato species *S. habrochaites*, in transgenic tomato, tobacco, and Arabidopsis. We show that a type VI trichomal gland cell-specific promoter from *S. habrochaites* is also Type VI trichomal gland cell-specific in the closely related species *S. lycopersicum* but is not active in tobacco glands. In contrast, the type I/IV trichomal gland cell-specific *S. lycopersicum* promoter tested is active in the gland cells of the long glandular trichomes of tobacco. Neither of these two trichomal gland cell-specific *Solanum* promoters is active in Arabidopsis.

5.1.3 Conclusions

The two tomato trichomal promoters tested here maintained their specificity when expressed in other plant species and they can be used to drive trichome specific gene expression in Solanaceae.

5.2 BACKGROUND

The aerial parts of many plants develop small protuberances called trichomes. Some of these trichomes have “gland” cells at their tips that synthesize, store, and secrete specialized metabolites that may aid in the defense of the plants against herbivores and pests (Wagner, 1991; Tissier, 2012; Schilmiller et al., 2012b). Other trichomes do not contain glands and appear to act as physical barriers (Eisner et al., 1998). For example, *Arabidopsis thaliana* plants produce either branched or un-branched unicellular, non-glandular, needle-like trichomes (Szymanski et al., 1999).

Plant species in the Solanaceae family exhibit many types of morphologically different glandular and non-glandular trichomes (Adedeji et al., 2007). A variety of specialized compounds have been found in these glands, including diterpenes in tobacco species (genus *Nicotianae*), mono- and sesquiterpenes, flavonoids, methylketones in tomato species (genus *Solanum*), and acyl sugars in *Nicotianae* species, petunia, *Datura*, and *Solanum* species (McCaskill and Croteau, 1999; Wagner et al., 2004; Schilmiller et al., 2008).

Detailed morphological descriptions of glandular trichomes have been provided for tobacco (*Nicotiana tabacum)*, cultivated tomato (*Solanum lycopersicum*), and other species (Werker, 2000). In tomato, the two main glandular trichomes are type I/IV (originally considered as two different types, but recently shown to be equivalent in their chemical content, see McDowell *et al*. 2011) and type VI. A type I/IV glandular trichome has a long and slender stalk made up of several cells in tandem and a single gland cell sitting on top of the stalk (and having a shape reminiscent of a upside-down round-bottomed flask). Type VI glandular trichomes consist of a unicellular long stalk cell followed by a short cell that flares out, and four gland cells sitting on a short connecting cell and arranged as a sphere. *Solanum* type I/IV glands have been shown to synthesize and secrete acyl sugars and methylated flavonoids (Schmidt et al., 2011; Schilmiller et al., 2012a), whereas type VI glands appear to specialize in synthesizing and storing terpenes and methylketones (Fridman et al., 2005; Ben-Israel et al., 2009; Schilmiller et al., 2012a). Tomato also has a third, less abundant type of multicellular glandular trichome, type VII, but this gland does not seem to be involved in the synthesis of specialized metabolites (McDowell et al., 2011).

Tobacco species have a long glandular trichome with a resemblance to *Solanum* type I/IV shown to synthesize and secrete acyl sugars but in some species is also involved in the synthesis of terpenes (Keene and Wagner, 1985; Wang and Wagner, 2003; Slocombe et al., 2008; Cui et al., 2011). Tobacco species do not have a glandular trichome that is similar to *Solanum* type VI, but they have a short, multicellular trichome that appears similar to *Solanum* type VII (Luckwill, 1943), and that was reported to make and secrete low-molecular mass proteins (Shepherd et al., 2005).

Interest in the biosynthetic capacity of glandular trichomes in the Solanaceae and elsewhere led to the identification of genes encoding enzymes involved in the synthesis of terpenes, methylketones, acyl sugars and flavonoids (Fridman et al., 2005; Ben-Israel et al., 2009; Schmidt et al., 2011; Kim et al., 2012; Schilmiller et al., 2012a). These genes are typically expressed only or mostly in the gland cells of a single type of trichome due to the promoters of the genes. For example, several promoters demonstrated specificity for the gland cells of long glandular trichome of tobacco by using *promoter::GFP* or *promoter::GUS* fusion constructs. These promoters are of the gene encoding *Nicotiana tabacum* lipid transfer protein 1 (*LTP1*) (Choi et al., 2012), and of the *N. tabacum* (cv. T.I. 1068) *CYP71D16* gene encoding a cytochrome P450 oxidoreductase active on the diterpene cembratriene-ol to produce cembratriene-diol (Wang et al., 2002), and of the *N. sylvestris* gene encoding the cembratriene-ol synthase (Ennajdaoui et al., 2010). Similar experiments in tomato showed specificity of the promoter of the acyltransferase gene *SlAT2* for the gland cell of type I/IV trichomes and the gland cells of type VI trichomes for the *Solanum habrochaites ShMKS1* gene (Bleeker et al., 2012; Akhtar et al., 2013).

However, whether the cell-type specific activity of these promoters is maintained across genera has not been reported.

Promoters that are completely specific for the gland cells of glandular trichomes could be useful for engineering new biochemical capacities in specific trichome types (Tissier, 2012). At present, however, the ability of such promoters to confer activity in glandular trichomes in distally related species is unclear. Predicting which glandular trichome-specific promoters would be active in the trichomal gland cells of other species is further complicated by our inability to determine whether similarly looking glandular trichomes in different species are homologous structures.

Several reports showed that the promoters of the MYB2 transcription factor gene and the lipid transfer protein genes that are specifically expressed in cotton (*Gossypium hirsutum*) fiber cells are active in transgenic tobacco exclusively in the long glandular trichomes (Liu et al., 2000; Shangguan et al., 2008). Besides the difficulty in defining cotton fiber cells as glandular trichomes, these promoters were expressed in all cells of the trichome, and when the *GhMYB2* promoter was tested in Arabidopsis it was also active in the non-glandular trichomes of that species (Shangguan et al., 2008). In addition, Gutierrez-Alcala et al. (2005) used the promoter of an *OASA1* gene, which encodes a cytosolic *O*-acteylserine(thiol)lase that is expressed in the Arabidopsis non-glandular trichomes, to drive expression of GFP or GUS in glandular trichomes in mint (*Mentha piperita* L.) and tobacco (*N. tabacum* cv W38) (Gutiérrez-Alcalá et al., 2005). In mint, both peltate and capitate glands actively expressed GFP or GUS, with some expression seen in the stalk of peltate trichomes as well but no expression was seen in other tissues. In tobacco, the GUS expression appeared to be stronger in the gland cells of

both long glandular trichomes and short multicellular glandular trichomes (Gutiérrez-Alcalá et al., 2005).

Many of the biochemical pathways responsible for the synthesis of specialized metabolites in Solanaceae trichomes are known or predicted to occur in the plastids, such as the synthesis of terpenes (Falara et al., 2011) and methylketones (Yu et al., 2010). However, while the presence of plastids in peltate glands of mint and their involvement in terpene biosynthesis was exhaustively documented (Turner et al., 1999), no information on the presence of plastids in any Solanaceae glandular trichomes has been reported.

Here, we report the examination of the activity of two promoters of type I/IV or VI glands of either *S. lycopersicum* or *S. habrochaites* in transgenic tomato, tobacco and Arabidopsis. We show that a type I/IV-specific promoter from *S. lycopersicum* is active in the gland cells of the long glandular trichomes of tobacco, which are morphologically similar to *Solanum* type I/IV glandular trichomes. On the other hand, a type VI trichome-specific promoter from *S. habrochaites* is active in type VI gland cells in *S. lycopersicum*, but is not active in the glandular trichomes of tobacco, which do not have a trichome type similar to the *Solanum* type VI glandular trichome. Neither of the two *Solanum* glandular trichome promoters tested here is active in Arabidopsis trichomes, which are exclusively non-glandular. These results suggest that much divergence has occurred in glandular trichomes across species, but that some types of glandular trichomes within plant families may have maintain similar organellar and cellular identities. In addition, our results provide evidence for the presence of plastids inside the gland cells of both type I/IV and VI trichomes in tomato.

5.3 RESULTS

5.3.1 The tomato type I/IV *SlAT2* promoter is active in the gland cells of tobacco long glandular trichomes, but not in Arabidopsis trichomes

The *Acyltransferase 2* gene (*SlAT2*, Solyc01g105580) from *S. lycopersicum* encodes a BAHD acyltransferase involved in acyl sugar biosynthesis in type I/IV glands (Schilmiller et al., 2012a). A 1486-bp fragment of the *SlAT2* promoter was previously shown to drive the expression of GFP in type I/IV trichomes in transformed tomato plants (Schilmiller et al., 2012a). We confirmed this activity by observing GFP signal in the glandular tips of type I/IV trichomes during all the developmental stages in aerial parts of the plants (Figure 5-1A). The gland cell of tomato type I/IV trichome is a single cell on top of the multicellular stalk (Figure 5-1B).

Figure 5-1. Activity of tomato glandular trichome specific promoters shown by *promoter::GFP* expression in different plant species

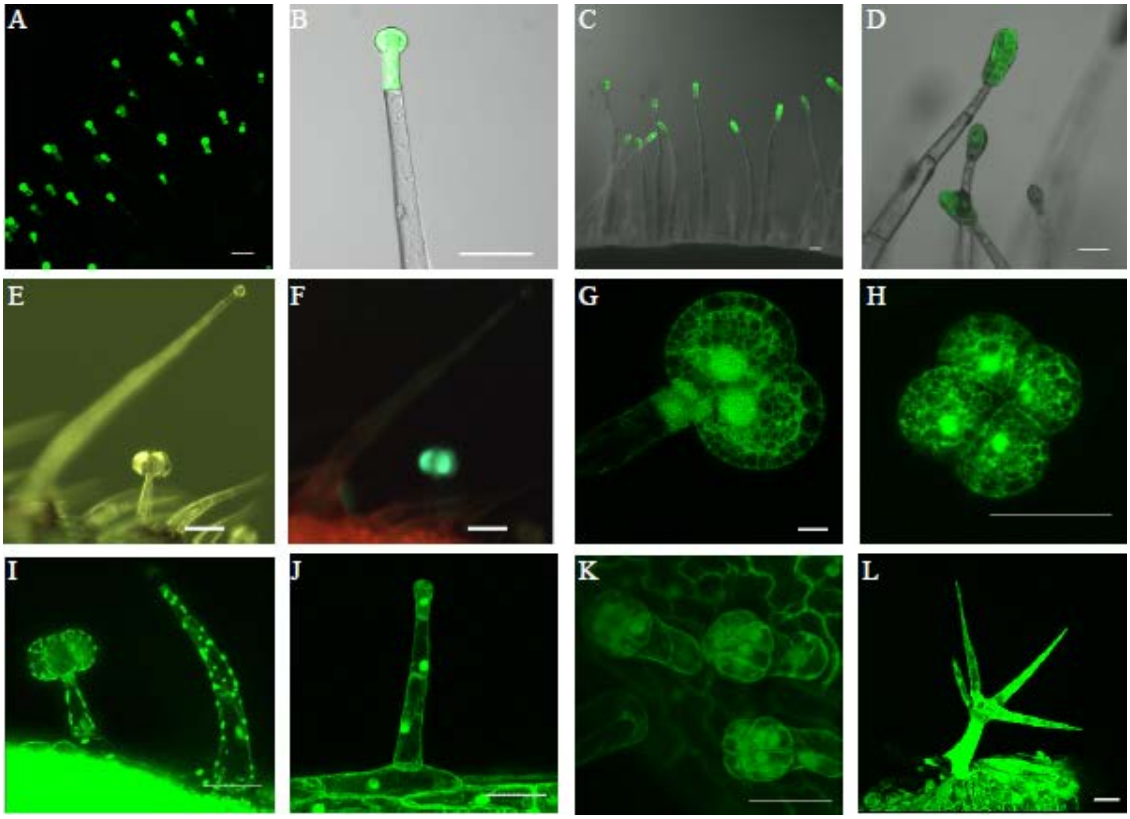

Figure 5-1. Activity of tomato glandular trichome specific promoters shown by promoter::GFP expression in different plant species. (A) and (B), *pAT2::GFP* transformed tomato type I/IV trichomes, with (B) showing a close view. (C) and (D), *pAT2::GFP* transformed tobacco long glandular trichomes, with (D) showing a close view. (E) and (F), *pShMKS1::GFP* transformed tomato type VI trichomes (G) and (H), Close side-view and top-view, respectively, of the *pShMKS1::GFP* transformed tomato type VI trichome. (I) *p35S::tpGFP* transformed tomato trichomes (a *p35S::tpGFP* transformant is shown here because *p35S::GFP* tomato transformants were not available.) (J) *p35S::GFP* transformed tobacco long glandular trichome. (K) *p35S::GFP* transformed tobacco short glandular trichomes. (L) *p35S::GFP* transformed Arabidopsis non-glandular trichome. All images except (E) and (F) are taken by confocal microscope. (E) and (F) are taken by fluorescence anatomy microscope with (E) in bright field view and (F) in green fluorescence view. The settings for both confocal microscope and fluorescence anatomy microscope are detailed in Materials and Methods. Length of scale bars shown in figures is 50 μm.

The *pAT2::GFP* (see 5.6 Methods) construct was transformed into *N. tabacum* SR1 and the GFP signal was visualized by confocal microscope. In these experiments

GFP expression was observed specifically in the long glandular trichomes (Figure 5-1C, D). Unlike tomato type I/IV glandular trichomes, tobacco long glandular trichomes have multiple gland cells stacked in tandem on the multicellular stalk, and the promoter of *SlAT2* appears to be active in all those gland cells, but not in the stalk cells (Figure 5-1D). Arabidopsis plants carrying *pAT2::GFP* showed no GFP signal in trichomes or any other cell types (data not shown).

5.3.2 The *ShMKS1* promoter is active in the gland cells and the apex of the stalk of type VI glandular trichomes in *S. lycopersicum*, but not in tobacco or Arabidopsis trichomes

It was previously shown that *S. habrochaites* PI126449 plants produce high levels of methylketones in their type VI glands (Fridman et al., 2005). The mRNA of *S. habrochaites* methylketone synthase I (*ShMKS1*), a gene encoding a decarboxylase involved in methylketone biosynthesis (Yu et al., 2010; Auldridge et al., 2012), was one of the most abundantly expressed sequences in the *S. habrochaites* PI126449 total (combined) trichome Expressed Sequence Tag (EST) database (Fridman et al., 2005). In contrast, cultivated tomato has several *ShMKS1* homologs, but none of them are highly expressed in its trichomes, as evidenced by their extremely low presence in total (combined) trichome EST databases of *S. lycopersicum* M82 (http://www.trichome.msu.edu/). RT-PCR with *ShMKS1*-specific primers and total RNA extracted from hand-picked type I/IV or type VI gland cell preparations from *S. habrochaites* show that this gene is expressed in type VI, but not in type I/IV trichomes (Figure 5-2).

Figure 5-2. RT-PCR amplification of ShMKS1 and actin mRNA from collections of purified *S. habrochaites* gland cells of type I/IV and type VI glandular trichomes

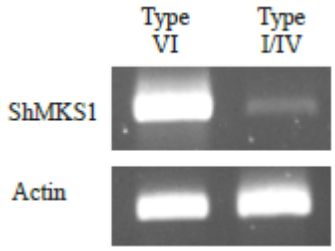

Figure 5-2. RT-PCR amplification of *ShMKS1* and *actin* mRNA from collections of purified *S. habrochaites* gland cells of type I/IV and type VI glandular trichomes. VI or I/IV indicate the RNA source from type VI or I/IV trichomes, respectively. The mRNA amplified is indicated on the left. Amplification of actin transcript was done to demonstrate similar amount RNA template was used. The number of PCR cycles was set at 40 to reach saturation.

Recently, a 1.5 kb long genomic region directly upstream from the initiating methionine codon of *ShMKS1*, isolated by chromosome walking from the genome of *S. habrochaites* PI126449 was shown to direct the expression of a GFP marker gene exclusively in the type VI trichome gland cells of the cultivated tomato *S. lycopersicum* (Bleeker et al., 2012; Akhtar et al., 2013). Here we confirm type VI trichome gland cell-specific expression by fluorescence anatomy microscopy (Figure 5-1E, F) and by confocal microscopy of transgenic tomato plants expressing GFP driven by *ShMKS1* promoter (Figure 5-1G, H). The data indicate that the *ShMKS1* promoter is active in all four gland cells as well as in the apex cell of the trichome stalk that connects the four gland cells.

To test whether the *ShMKS1* promoter is also active in tobacco trichomes, we obtained a total of 20 lines of tobacco plants transformed with the *pShMKS1::GFP*

construct that was used to transform *S. lycopersicum*. In contrast to the situation in tomato, no GFP signal could be observed in any of the transgenic tobacco plants. Arabidopsis plants carrying the same construct also showed no GFP signal in trichomes or any other cells (data not shown).

As expected, tomato, tobacco, and Arabidopsis plants carrying a *p35S::tpGFP* (See 5.6 Methods) or *p35S::GFP*, obtained for control, showed GFP signal in all types of trichomes as well as in other types of cells (Figure 5-1H-K).

5.3.3 Plastids are found in the gland cells of both type I/IV and type VI glandular trichomes in tomato

Some pathways of specialized metabolism such as methylketone and terpene biosynthesis occur in the plastids (Yu et al., 2010; Falara et al., 2011). Because both type I/IV and type VI trichome gland cells do not accumulate detectable amounts of chlorophyll or other known plastid-associated pigments, the extent of plastid distribution in these cells was not clear. To examine the presence of plastids in the gland cells of tomato glandular trichomes, we modified the constructs *pAT2::GFP* and *pShMKS1::GFP* to add a transit peptide from ribulose bisphosphate carboxylaase/oxygenase (RuBisCo) small subunit, a plastidic protein, fused in frame to the N-terminus of the GFP coding sequence. The resultant *pAT2::tpGFP* and *pShMKS1::tpGFP* vectors were used to produce transgenic tomato plants, which should direct the import of GFP into plastids (Nagegowda et al., 2008).

In plants expressing *pAT2::tpGFP*, the GFP signal was found in dots that co-localized with red auto-fluorescence signal from chlorophyll inside type I/IV trichome

gland cells (Figure 5-3A-C). The transgenic tomato plants carrying *pShMKS1::tpGFP* also showed green dots in gland cells, but not in stalk, of type VI glandular trichomes (Figure 5-3G-I). For both tomato type I/IV and type VI trichomes, the observed auto-fluorescence in the gland cells was faint.

GFP signal in trichomes of transgenic tomato lines expressing the *p35S::tpGFP* also appeared as plastid-like dots (Figure 5-3D-F and J-L). The GFP signal was observed in the gland cells of both type I/IV and type VI glandular trichomes (Figure 5-3D-F and J-L) in addition to other leaf tissues.

5.4 DISCUSSION

5.4.1 Activity of *Solanum* trichomal gland cell-specific promoters across species

While the available promoters specifically expressed in gland cells of glandular trichome are important for biotechnological applications, the activity of such promoters also provides clues to the evolutionary relationships of specific trichomes and specific cells within the trichomes. In this respect, only a few tests of the activity of promoters of trichomes-expressed genes in distantly related species have been reported. All involved promoters that are active in non-glandular trichomes in Arabidopsis or cotton and their heterologous expression (in tobacco and Arabidopsis) was not confined to gland cells (Hsu et al., 1999; Liu et al., 2000; Gutiérrez-Alcalá et al., 2005; Shangguan et al., 2008). In contrast, the promoters of both *SlAT2* and *ShMKS1* showed strict specificity for the gland cells of the original type of trichomes across species, and were not active in any other cells of any other type of trichomes, glandular or not, in all three species tested (tomato, tobacco, and Arabidopsis).

Figure 5-3. Plastid-targeted GFP expression in tomato glandular trichomes visualized by confocal microscope imaging

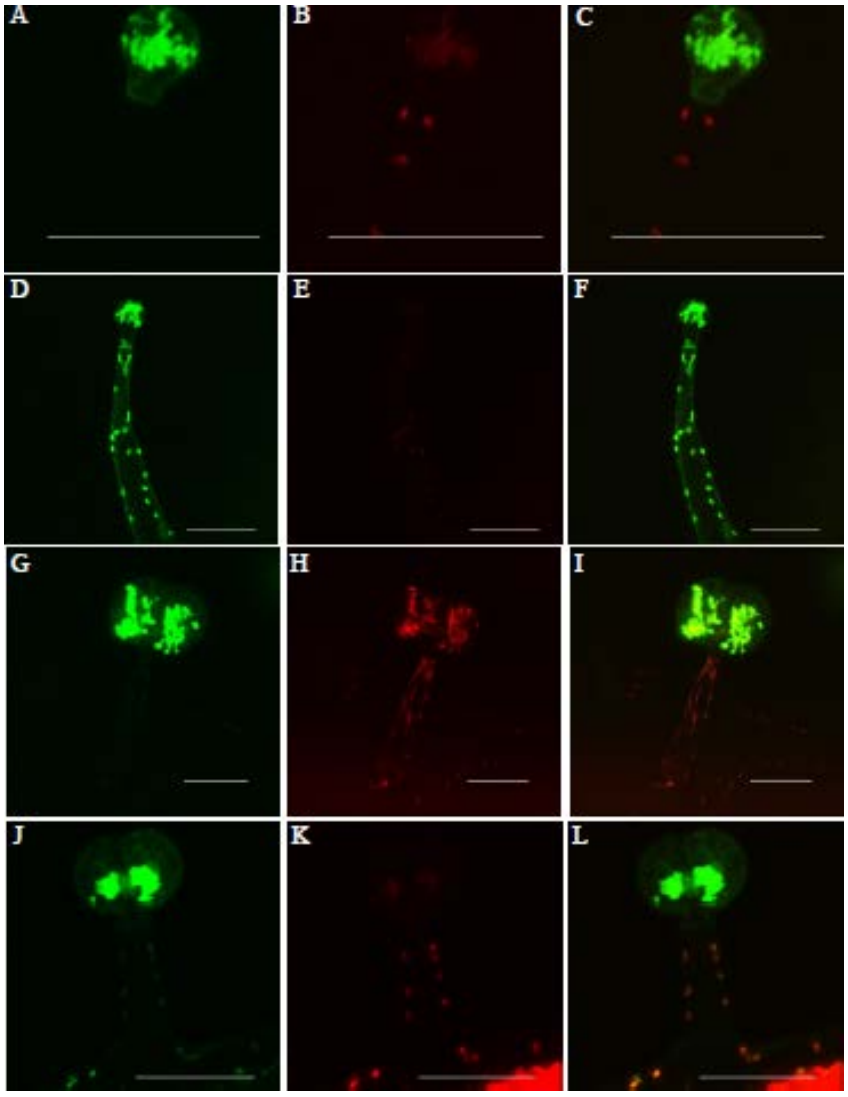

Figure 5-3. Plastid-targeted GFP expression in tomato glandular trichomes visualized by confocal microscope imaging. The left column shows the green fluorescence signal only; the middle column shows the auto-fluorescence signal only; and the right column shows the merged view of the first two columns. (A-C), Type I/IV trichome of the *pAT::tpGFP* transformed tomato. (D-F), Type I/IV trichome of *p35S::tpGFP* transformed tomato. (G-I), Type VI trichomes of *pShMKS1::tpGFP* transformed tomato. (J-L), Type VI trichomes of *p35S::tpGFP* transformed tomato. Scale bar = 50 μm.

Our data indicate that the promoter of *SlAT2* is active in gland cells of type I/IV trichomes of tomato as well as in the gland cells of the long glandular trichomes in tobacco. This observation suggests that the gland cells of the tomato type I/IV trichome and the tobacco long glandular trichomes might employ a similar set of *trans*-activating factors for the *SlAT2* promoter. The result suggests that these structures may share a common developmental origin, although their morphologies are somewhat different, *i.e.* the tobacco long glandular trichome has multiple gland cells (Figure 5-1C, D) while the tomato type I/IV trichome has a single gland cell at the tip of the stalk (Figure 5-1A, B).

We observed that the *ShMKS1* promoter, which is specific for the gland cells of type VI trichomes in *S. habrochaites*, was also specific for the gland cells of type VI trichomes in *S. lycopersicum* (Figure 5-1, Figure 5-1). However, the *ShMKS1* promoter was not active in any type of tobacco trichomes, none of which morphologically resembles the *Solanum* type VI glandular trichomes.

5.4.2 The 35S promoter is active in all cells of tomato, tobacco and Arabidopsis

The cauliflower mosaic virus 35S (CaMv 35S) promoter is widely used to over express foreign genes in plants in all tissues and organs (Hull et al., 2000). While the 35S promoter was previously reported to be active in trichomes (Battraw and Hall, 1990; Hsu et al., 1999; Mathur and Chua, 2000; Sunilkumar et al., 2002; Choi et al., 2012), glandular trichomes in particular have not been examined in detail, although several reports have shown that 35S promoter-driven RNAi constructs successfully repressed gene expression in glandular trichomes (Schilmiller et al., 2012a; Akhtar et al., 2013). To examine in detail the activity of the 35S promoter in trichomes, a 35S promoter::GFP construct was transformed into tobacco and Arabidopsis and a *p35S::tpGFP* construct

was transformed into tomato. Our results indicate that the 35S promoter is active in all types of trichomes in tomato, tobacco, and Arabidopsis (Figure 5-1, Figure 5-3)

5.4.3 Evidence for the presence of plastids in gland cells of both type I/IV and type VI trichomes in tomato

Both gland and stalk cells of type I/IV and type VI trichomes expressing a GFP protein with an N-terminal plastid-targeting sequence showed a green signal in discrete dots. Moreover, these green dots co-localized with red dots observed, presumably from chlorophyll auto-fluorescence (Figure 5-3). This observation suggests that gland cells contain chlorophyll-containing plastids, consistent with the previously established plastid localization of the biosynthesis of specific terpenes in the gland cells of type VI trichomes in *S. lycopersicum* (Falara et al., 2011; Lange and Turner, 2013) and the plastid localization of the elongation steps in the biosynthesis of the acyl metabolites that serve the acyl sugar biosynthesis in the glandular trichomes of type I/IV glands (Schilmiller et al., 2012a). However, these gland cells are not visibly green, and the red auto-fluorescence is correspondingly faint (Figure 5-3), indicating that the chlorophyll amount is low in those gland cells. Thus, the photosynthesis activity in those plastids may also be limited due to the low chlorophyll amount.

5.5 CONCLUSIONS

We show here that two *Solanum* trichomal gland cell-specific promoters maintain their specificity when in another plant species. The promoter of *ShMKS1*, which is active specifically in type VI trichomal gland cells, is also active in the same gland cells in *S. lycopersicum*, but is not active in tobacco or Arabidopsis, which do not have this type of gland. The promoter of *SlAT2*, which is active specifically in type I/IV trichomal gland

cells, is active in tobacco gland cells of the long glandular trichomes, which are morphologically similar to the tomato type I/IV trichomes, but is not active in Arabidopsis, which does not have this type of gland either. The two promoters analyzed here can be used to express genes specifically in trichomal gland cells in Solanaceae.

5.6 METHODS

5.6.1 Promoter sequences

The promoter sequences of *SlAT2* and *ShMKS1* were isolated using methods as previously described (Yu et al., 2010; Schilmiller et al., 2012a).

5.6.2 RNA isolation from specific types of *S. habrochaites* trichomes and RT-PCR

Type VI trichomes were isolated using the same method detailed by Schmidt *et al.* (Schmidt et al., 2011). Type I trichomes were picked up by hand with an extra-fine forceps Dumont #5 (Catalog # 11251-30 in Thermo Fischer Scientific Inc.) under an anatomy microscope (Leica MZ6). RNA was isolated from both samples using E.Z.N.A. Plant RNA Kit (Omega Bio-Tek Inc., GA) following the manual with the kit. The oligo-nucleotides used for RT-PCR were: 5'- ATGGAGAAAAGCATGTCGCCA-3' and 5'-TTATTTATACTTGTTAGCGATGCTTAGAAGA-3' for *ShMKS1* and 5'-ATTCCCTGACTGTTTGCTAGT-3' and 5'- TCCAACACAATACCGGTGGT-3' for *actin.*

5.6.3 Constructs for plant transformation

The *p35S::GFP* construct is the same as *pSAT6A-EGFP-N1* (Tzfira et al., 2005). Full-length promoter sequences of *SlAT2* and *ShMKS1* were amplified by KOD polymerase (EMD4Biosciences, Merck, KGaA, Germany) to add *Age* I and *Bam* HI

restriction sites and spliced into *pSAT6A-EGFP-N1* (Tzfira et al., 2005) to replace the 35S promoter, generating *pAT2::GFP* and *pShMKS1::GFP*. The expression cassettes were digested by PI-PspI (New England Biolabs Inc.) and ligated to *pPZP-RCS2* binary vector and transferred into *Agrobacterium tumefaciens* strain EHA105 (Tzfira et al., 2005).

The *RuBisCo* small subunit transit-peptide sequences was amplified by oligonucleotides 5'- GGATCCATGGCTTCCTCTATGCTCTCTT-3' and 5'-AGATCTTGCAGTTAACTCTTCCGCCG-3' from 326-rbcGFP vector (the underlined sequences are recognition sites for *Bam* HI and *Bgl* II, respectively) (Nagegowda et al., 2008) and spliced into the *Bam* HI site in the ready-made promoter::GFP constructs mentioned above, to generate *pAT2::tpGFP, pShMKS1::tpGFP* and *p35S::tpGFP*. After all of the expression cassettes were sequencing verified, they were spliced to the *pPZP-RCS2* binary vector.

5.6.4 Plants transformation

Arabidopsis transformation was employing the floral dip method (Clough and Bent, 1998). Tobacco transformation was following a leaf-disc *A. tumefaciens* infection mediated method (McCormick, 1991). *pAT2::GFP* and all the tpGFP constructs were transformed at Michigan State University (MSU) following the method detailed before (Schilmiller et al., 2012a).

pShMKS1::GFP transgenic tomato plants were generated in Plant Transformation Core Research Facility (PTCRF) at the University of Nebraska following the protocol detailed on their website (http://biotech.unl.edu/node/211).

5.6.5 Microscopy imaging

Confocal microscopy imaging was following the same setting described before (Yu et al., 2010), except the images were taken through a 20* optical object lens with photomultiplier detection from 500 to 530 nm. The confocal images were processed by the LSA AS lite software from Leica Inc. The fluorescence stereo microscope is Leica MZFLIII equipped with a 100 Watt mercury Arc lamp, 505nm dichroic mirror and with GFP excitation (450-490nm) and emission (520-560nm) filter settings, with its original manufactured filter cube and arc lamp for the bright-field and GFP imaging.

Chapter 6 Conclusions and future directions

6.1 PLANTS SYNTHESIZE METHYLKETONES THROUGH *DE NOVO* FATTY ACID BIOSYNTHETIC PATHWAY

Although aliphatic methylketones occur widely in nature, only limited information about their biosynthesis has been obtained, with the exception of acetone biogenesis in mammals. In particular, the first research data on plant methylketone biosynthesis come from the work of Fridman et al in 2005 (Fridman et al., 2005). They showed that methylketone biosynthesis was tightly linked to *de novo* fatty acid biosynthesis, based on two observations (Fridman et al., 2005). First, the transcript levels of several fatty acid biosynthetic genes were higher in the trichomes of the methylketone-producing plants than those in the non-producers (Fridman et al., 2005). Second, ShMKS1 was localized to the plastids, the site of fatty acid biosynthesis in plants (Fridman et al., 2005).

The work presented here builds on and expands Fridman et al.'s initial discovery, with multiple new observations. First, a genetic analysis of the segregated F2 progeny of *S. habrochaites* X *S. lycopersicum* hybrid showed that multiple fatty acid biosynthetic gene loci were required for maximum methylketone production (Chapter 2). In addition, *ShMKS2* was identified in that study to be positively associated with methylketone production in plants (Chapter 2). Second, the protein sequence of ShMKS2 was found to have a plastid-targeting transit-peptide and ShMKS2 was shown to be localized in the plastids by confocal imaging of the ShMKS2-GFP fusion protein in plant leaves (Chapter

3). Third, transgenic plants expressing *ShMKS2* constitutively under the 35S promoter displayed phenotypes similar to the Arabidopsis mutants that are deficient in fatty acid biosynthesis (Chapter 4). Thus, unlike mammals or some fungi species that produce methylketones through fatty acid catabolism, plants synthesize methylketones through *de novo* fatty acid biosynthesis, in which all the acyl intermediates are carried by ACPs.

6.2 SHMKS1 IS A DECARBOXYLASE AND SHMKS2 IS A THIOESTERASE

The function of the genetically identified *ShMKS1* and *ShMKS2* was furthered studied by biochemical approaches. The *in vitro* enzyme assay of purified ShMKS1 with 3-keto myristic acid led to the production of 2-tridecanone at a rate of 2.6 nM/μg protein/min (Chapter 3). Under the same conditions ShMKS2 had a significant lower reaction rate of 0.1 nM/μg protein/min, indicating that ShMKS1 but not ShMKS2 is the decarboxylase responsible for the biosynthesis of methylketones (Chapter 3). The addition of purified ShMKS1 to the spent medium of *E. coli* expressing *ShMKS2* that accumulated 3-keto acids, significantly enhanced the levels of the methylketones extracted and detected by GC-MS (Chapter 3), further supporting the hypothesis that ShMKS1 is the decarboxylase that uses 3-keto acids as substrates to produce methylketones.

However, the spent medium of *E. coli* expressing only *ShMKS1* did not accumulate any 3-keto acids (Chapter 3). Also, the purified ShMKS1 did not exhibit significant activity when supplied with *β*-ketomyristoyl-ACP as substrate *in vitro* (Chapter 3). In addition, the phenotypes of transgenic Arabidopsis and tobacco plants expressing *ShMKS1* were indistinguishable to the wild type plants, while plants

expressing *ShMKS2* displayed phenotypes of severe disruption in fatty acid biosynthesis (Chapter 4). Taken together, ShMKS1 was not found to have thioesterase activity.

On the other hand, purified ShMKS2 was shown to use *β*-ketomyristoyl-ACP as substrate to produce 3-ketomyristic acid, which is detected by GC-MS after its derivatization to 2-tridecanone by acid and heat treatment (Chapter 3). In that assay, the production of 3-ketomyristic acid was 4-fold higher using ShMKS2 than using ShMKS1, the latter of which showed no difference with the negative control (Chapter 3). However, 3-ketomyristic acid accumulated in the reaction of ShMKS2 if heat and acid treatment was not used to derivatize it (Chapter 3). These results suggest that the ShMKS2 is a thioesterase that removes the ACP from *β*-ketoacyl-ACPs to generate 3-keto acids, but is not capable of catalyzing the conversion of 3-keto acids to 2-methylketones. The accumulation of 3-keto acids in the spent medium of *E. coli* expressing only *ShMKS2* also supports that ShMKS2 indeed functions as a thioesterase *in vivo*, but is inactive towards 3-keto acids.

Another observation is that the methylketone biosynthesis in wild tomato *S. habrochaites*, transgenic *E. coli*, Arabidopsis and tobacco plants overexpressing ShMKS2 was characterized by similar ratios of methylketone products of different length (Chapter 3 and Chapter 4), suggesting that ShMKS2 retains its preference for similar chain-length fatty acid intermediates in both plants and *E. coli*. Finally, the presence of myristic acid in *S. habrochaites*, as well as in transgenic plants expressing *ShMKS2* with or without *ShMKS1* (Chapter 4), indicates that ShMKS2 also uses saturated myristoyl-ACP as its substrate. Taken together, ShMKS2 is a thioesterase that uses *β*-ketoacyl-ACPs as its

preferred substrates in both *E. coli* and plant host species, and it can also hydrolyze 14-carbon long acyl-ACP.

In summary, ShMKS1 and ShMKS2 are two enzymes that sequentially catalyze the conversion of the fatty acid biosynthetic intermediates *β*-ketoacyl-ACPs to methylketones in wild tomato species. First, ShMKS2 hydrolyzes the *β*-ketoacyl-ACPs to 3-keto acids. Next, the previously produced 3-keto acids get decarboxylated by ShMKS1 to produce methylketones (Figure 6-1).

Figure 6-1. A schematic reaction sequence for the synthesis of straight-chain methylketones.

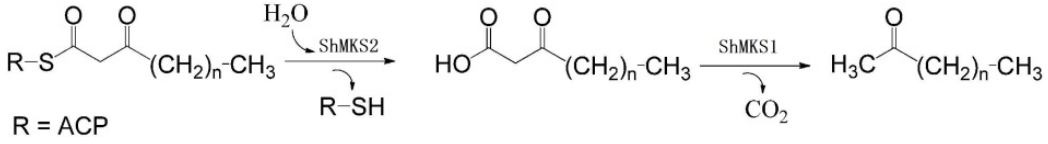

Figure 6-1. A schematic reaction sequence for the synthesis of straight-chain methylketones. *β*-ketoacyl-ACPs, the fatty acid biosynthetic intermediates, are first hydrolyzed by ShMKS2. Then the resulted 3-keto acids are decarboxylated by ShMKS1 (Adopted from Yu et al., 2010).

6.3 EVOLUTION OF SHMKS1 AND SHMKS2

The protein sequences of ShMKS1 and its ortholog, SlMKS1a, are characterized by an alanine residue instead of the conserved serine residue (at position 87 in ShMKS1) in one of the amino acids that constitute the essential catalytic triad of the *α*/*β*-hydrolases superfamily proteins (Hotelier et al., 2004; Forouhar et al., 2005; Auldridge et al., 2012). This substitution makes them fundamentally different from other ShMKS1-like proteins. Indeed, *in vitro* enzymatic assays of ShMKS1 and SlMKS1a with the substrates of AtMES17 (mIAA, mJA and mSA, Yang, et al, 2008), the protein with highest identity to ShMKS1 in Arabidopsis, showed that ShMKS1 and SlMKS1a were not able to hydrolyze

mIAA, mJA or mSA (unpublished data). Thus, ShMKS1 is likely to have acquired the decarboxylase activity recently and, at the same time, essentially lost the *α*/*β*-hydrolase activity (Auldridge et al., 2012).

Proteins with homology to ShMKS2 are found in a wide range of plant species, but methylketones are identified only in a limited number of plant species (Forney and Markovetz, 1971). Interestingly, the expression of the *ShMKS2* homolog from Arabidopsis in *E. coli* also led to methylketone production (Chapter 3), indicating that the thioesterase activity towards *β*-ketoacyl-ACPs might be an ancestral activity of ShMKS2-like proteins. However, high-level expression of *ShMKS2* in plant tissues disrupted fatty acid biosynthesis and caused deleterious effects for plants, as was shown by the deleterious effects exhibited by the transgenic Arabidopsis, tobacco and tomato plants expressing *ShMKS2* (Chapter 4). Thus, plants with high-level accumulation of ShMKS2-like proteins are probably negatively selected in nature. In the case of *S. habrochaites* PI126449, high-level expression of *ShMKS2* is confined to the glandular trichomes (Antonious, 2001). Therefore methylketones are synthesized only in the trichomes, which prevents the depletion of fatty acids and the accumulation of the potential toxic 3-keto acids in vegetative tissues. An ancestral ShMKS1 possessing low decarboxylation activity might have to evolve together with ShMKS2 for optimal methylketone production. Since methylketones provide pesticidal properties, *S. habrochaites* with trichome-specific co-expression of both ShMSK1 and ShMKS2 would be positively selected over time. In summary, the substrate preference of ShMKS2 and the requirement for a second protein to co-evolve have resulted in methylketone biosynthesis occurring only sporadically in the plant kingdom.

6.4 TRICHOMAL EXPRESSION OF *SHMKS1* AND *SHMKS2* TOGETHER IS INSUFFICIENT FOR METHYLKETONE PRODUCTION IN CULTIVATED TOMATO

Tomato breeders have tried to get cultivated tomatoes with methylketone producing properties for many years. However, the interspecific crosses between *S. habrochaites* and *S. lycopersicum* did not generate a commercializable tomato variety (Hartman and St. Clair, 1998). Since the expression levels of *ShMKS1* and *ShMKS2* were 800-fold and 300-fold higher in the trichomes of *S. habrochaites* than in those of *S. lycopersicum*, with other methylketone-producing associated genes having < 8-fold difference, simply increasing the levels of expression of the two methylketone synthase genes in cultivated tomato might lead to increased methylketone production. In this work both genes were expressed under the control of tomato trichome-specific promoters to prevent the deleterious effects caused by constitutive expression of *ShMKS2* in vegetative tissues (Chapter 4). However, no methylketones were detected in cultivated tomato expressing both ShMKS2 and ShMKS1 in trichomes (Chapter 4), indicating that trichomal expression of *ShMKS1* and *ShMKS2* is not sufficient for methylketone production in cultivated tomato. Since several fatty acid biosynthetic gene loci were shown to be involved in the high levels of methylketone accumulation in *S. habrochaites* (Chapter 2), successful genetic engineering of cultivated tomato for methylketone production might require introducing additional fatty acid biosynthesis related genes to cultivated tomato in addition to trichomal expression of *ShMKS1* and *ShMKS2*.

6.5 HETEROLOGOUS, CONSTITUTIVE CO-EXPRESSION OF *SHMKS1* AND *SHMKS2* LEADS TO METHYLKETONES PRODUCTION IN TRANSGENIC PLANTS

Methylketones were detected in transgenic Arabidopsis and tobacco plants expressing *ShMKS1* and *ShMKS1* together under the general promoters of p35S and pRbcs (Chapter 4), suggesting that co-expression of *ShMKS1* and *ShMKS2* in the plants' leaves is sufficient to produce methylketones. However, the plants expressing these two genes or even *ShMKS2* alone displayed severe growth retardation and developmental malfunctions (Chapter 5), which were similar to those of the Arabidopsis plants deficient in *de novo* fatty acid biosynthesis (Mou et al., 2000; Bonaventure et al., 2003; Wu and Xue, 2010). The phenotypes displayed by plants expressing *ShMKS2* are probably caused by *β*-ketoacyl-ACPs and myristoyl-ACP being hydrolyzed by ShMKS2 before they reach 16 and/or 18 carbons long, which results in insufficient fatty acid biosynthesis *in planta*. However, many plants possessing *ShMKS2*-like genes do not display phenotypes of defective in fatty acid biosynthesis, because these plants, such as Arabidopsis and cultivated tomatoes, express these *ShMKS2*-like genes only at very low levels.

Tobacco plants expressing *ShMKS2* alone also produced some methylketones at a rate of 2.46 μmol/g FLW/5 h, which was 5-fold less than that of methylketone production by tobacco plants expressing both *ShMKS1* and *ShMKS2* (Chapter 4). On the contrary, Arabidopsis or tomato plants expressing *ShMKS2* alone did not produce any detectable methylketones (Chapter 4). It appears that *ShMKS1* is not absolutely required for methylketone production *in planta*. *E. coli* expressing only *ShMKS2* also led to some methylketone production as well as 3-keto acid accumulation in the spent medium

(Chapter 2). Since 3-keto acids constantly undergo spontaneous decarboxylation (Kornberg et al., 1948), the methylketones detected in the culture of *E. coli* expressing *ShMKS2* likely result from non-enzymatic degradation of 3-keto acids, and ditto for tobacco. However, co-expression of *ShMK1* and *ShMKS2* in *E. coli* led to five times more methylketone production than the same bacteria strain expressing only *ShMKS2* (Park et al., 2012), consistent with different methylketone productivity of expressing *ShMKS2* with vs. without *ShMKS1* in transgenic tobacco plants. Thus expression of *ShMKS1* results in increased yields of methylketones in plants and *E. coli*, however is not an absolute requirement for methylketone production.

6.6 *SOLANUM* TRICHOMAL GLAND CELL-SPECIFIC PROMOTERS MAINTAIN THEIR SPECIFICITY ACROSS SPECIES

The activity of the promoters of *SlAT2* and *ShMKS1* was also investigated in this study by fusing each promoter with GFP (Chapter 5). The promoter of *SlAT2* was found active in the gland cells of tomato type I/IV trichomes as well as the gland cells of tobacco long glandular trichomes (Chapter 5), indicating that these two morphologically different trichomes in these two species might share a common developmental origin. The promoter of *ShMKS1* is specific for the gland cells of type VI trichomes of both *S. habrochaites* and *S. lycopersicum*, but is not active in either of the two glandular trichomes of tobacco, neither of which is morphological similar to type VI glandular trichome of tomato (Chapter 5). In addition, neither of the two glandular trichome promoters tested in this study was found active in the non-glandular trichomes of Arabidopsis (Chapter 5). These results suggest that the promoters of *SlAT2* and *ShMKS1* are strictly specific for the gland cells of type I/IV or type VI trichomes, respectively,

across species. Also, the long glandular trichomes of tobacco and tomato type VI glandular trichome might have evolved from an ancestral trichome type that existed before the separation of the two species.

6.7 PLASTIDS ARE PRESENT IN THE GLAND CELLS OF TOMATO TYPE I/IV AND TYPE VI TRICHOMES

The fluorescence signal of GFP with a plastid-targeting signal peptide co-localized with the red-fluorescence signal in discrete dots, which was presumably from the auto-fluorescence of chlorophyll (Chapter 5), demonstrating that measuring the fluorescence of tpGFP could serve as a good method to indicate the localization of plastids. Both promoters of *SlAT2* and *ShMKS1* led to tpGFP expression specifically in the gland cells of type I/IV or type VI trichomes, respectively, and the pattern of these tpGFP was in the form of some discrete dots (Chapter 5), indicating that the tomato gland cells of type I/IV and type VI trichomes have chlorophyll-containing plastids. However, the auto-fluorescence signal was weak in the trichome gland cells (Chapter 5), therefore, the chlorophyll content is likely to be limited in the trichomal plastids.

6.8 FUTURE DIRECTIONS

Since trichomal co-expression of *ShMKS1* and *ShMKS2* is insufficient for methylketone production in cultivated tomato (Chapter 4), an immediate follow-up question is to investigate what are the other genes that are necessary. Fatty acid biosynthetic genes (Chapter 2) are the first candidates to examine. For example, the genes encoding acyl carrier protein (ACP), acetyl-CoA carboxylase (ACC) and malonyl-CoA:ACP transacylase have higher expression in wild tomato than in cultivated tomato (Fridman et al., 2005), so introducing each of them individually or in combination into

tomato plants expressing *ShMKS1* and *ShMKS2* in trichomes may help improve methylketone production in trichomes by increasing the fatty acid flux. However, tomato trichomes are not connected to the vascular system (Schilmiller et al., 2008) and have limited photosynthetic activity due to the low chlorophyll levels inside the gland cells (Chapter 5), so the primary carbon and energy source for trichomes was proposed to be sucrose for its easy transportability in plants (Schilmiller et al., 2008). Thus, it is also worthwhile to attempt to increase the sucrose supply for the glandular trichomes in tomato. The partition of carbon source to different pathways may also need to be examined. The wild tomato *S. habrochaites* PI126449 that synthesizes large amounts of methylketones at the cost of producing limited quantities of terpenes (Fridman et al., 2005) contrasts sharply with the cultivated tomato that mainly synthesizes terpenes in its trichomes, indicating a competition for the carbon source between the fatty acid and terpene biosynthetic pathways. Thus, repression of the terpene biosynthesis in trichomes may help to redirect the carbon source to the fatty acid biosynthetic pathway. Finally, the cultivated tomatoes have a lower trichome density than the wild tomato does, so even if some methylketones were produced in the transgenic tomato trichomes, the overall amount may be too little. In order to produce detectable amounts of methylketones on the leaf surface of tomatoes, the trichome formation and determination factors will have to be discovered and further studied.

In addition to the engineering efforts mentioned above, microarray analysis on the bulked segregating F2 progeny of *S. habrochaites* X *S. lycopersicum* revealed a number of genes whose transcription levels are either higher or lower in the high-methylketone-producing lines than in the low-methylketone-producing lines (Chapter 2 and Table 2-1).

At present, attempts to identify the functions of many of these differentially expressed genes could only be based on their sequence homology to sequences annotated in the protein database. Therefore, an examination of the biological function of these genes may help to identify new steps involved in methylketone biosynthesis in tomato trichomes, which will not only deepen our understanding but also facilitate tomato trichome engineering.

Co-expression of *ShMKS1* and *ShMKS2* is sufficient for methylketone production in plant leaf tissue but also affects the normal growth and development of the transformed plants. Therefore, expression of these two genes driven by some inducible promoters, like the pest inducible promoter of *win3* gene from poplar (Hollick and Gordon, 1995), may help to alleviate the lesion phenotypes caused by the genes' constitutive expression. In addition, in order to avoid the depletion of fatty acids by the high expression these two genes, a negative- feedback loop could be further built into the wound inducible gene expression system. For example, the same pest inducible promoter used for *ShMKS2* could be placed upstream of a transcription factor, like the promoter of *XVE* that activates the transcription of the *OlexA-TATA* promoter (Brand et al., 2006), and the transcripts from *OlexA-TATA* promoter can be a siRNA for *ShMKS2*, inhibiting the expression of *ShMKS2* shortly after its transcription. Thus, the expression of *ShMKS2* in leaves will be first induced by pest wounding, which will result in methylketone production with a concurrent expression of *ShMKS1*, then its expression will be silenced by the siRNA transcribed from the *OlexA-TATA* promoter.

Fatty acid profile analysis of plants expressing *ShMKS2* indicates that ShMKS2 also hydrolyzes myristoyl-ACP in plants (Chapter 4), so ShMKS2 has substrate acceptance towards both ketoacyl-ACPs and saturated acyl-ACPs. Thioesterases having different forms of acyl-ACPs or acyl-CoAs are also found in bacterial and fungal species. Such as the *Bdellovibrio bacteriovorus* thioesterase (GenBank: CAE80300) that produces both free fatty acids and 3-keto acids (Jing et al., 2011); and the acyl-ACP thioesterase from *Lactobacillus plantarum* (GenBank: CAD63310) that produces 10:1, 12:1, 14:1, and 16:1 fatty acids when expressed in *E. coli* (Jing et al., 2011). Thus, characterization of thioesterase-like enzymes may require examining their substrates specificity towards all forms of acyl-ACPS and acyl-CoAs found in both the fatty acid biosynthetic and degradation pathways. Currently, there is an increased interest in obtaining thioesterases which could produce short to medium chain fatty acids (C4-C14), as those fatty acids have wide industrial applications (fuels, lubricants, cosmetics and detergents, *etc*.). Surveying the enzymatic activities of thioesterases from phylogenetically distantly related organisms may help to identify some thioesterase with such catalytic activities (Jing et al., 2011).

3-Keto acids could not be detected in leaf tissues of transgenic tobacco plants expressing *ShMKS2* alone (Chapter 4), suggesting that the resulted 3-keto acids are catabolized or converted back to ketoacyl-ACP at a much faster rate than their degradation rate. There are no reports studying how the 3-keto acids are catabolized or converted in plants, so investigation into 3-keto acids catabolic pathway may reveal some new information about plant metabolism, like identifying an enzyme that can reduce one

of the carbonyl group or ligate the 3-keto acid directly to an ACP. Also, whether the presence of 3-keto acids causes stresses to the plants needs further analyses.

The trichome-specific promoters presented in this study can serve as tools to manipulate gene expression in tomato-trichomal gland cells (Chapter 5). Also, the promoter of *SlAT2* could be used to drive gene expression specifically in the gland cells of tobacco long glandular trichomes. The *cis*-elements of the promoters and the *trans*-factors that interact with these promoters still need to be identified to further understand plant trichome gene expression. Furthermore, c*is*-elements can be mapped by systematic mutagenesis of specific regions inside these promoters to pinpoint what is the function of each region. Knowledge of these *cis*-elements will help to identify other potential trichomal promoters and design new promoters with higher efficiency and specificity. On the other hand, *trans*-factors can be potentially discovered by affinity purification using DNA fragments of these promoters against protein extracts of trichomes (Chodosh and Buratowski, 2001). The identified *trans*-factors may provide insight into the gene expression network in trichomes.

References

Adedeji O, Ajuwon OY, Babawale OO (2007) Foliar Epidermal Studies, Organographic Distribution and Taxonomic Importance of Trichomes in the Family Solanaceae. International Journal of Botany **3:** 276-282

Akhtar TA, Matsuba Y, Schauvinhold I, Yu G, Lees HA, Klein SE, Pichersky E (2013) The tomato *cis*–prenyltransferase gene family. The Plant Journal **73:** 640-652

Alford J, Smith J, Lilly H (1971) Relationship of microbial activity to changes in lipids of foods. J. appl. Bacteriol **34:** 133-146

Antonious GF (2001) Production and quantification of methyl ketones in wild tomato accessions. J Environ Sci Health B **36:** 835-848

Antonious GF, Dahlman DL, Hawkins LM (2003) Insecticidal and Acaricidal Performance of Methyl Ketones in Wild Tomato Leaves. Bulletin of Environmental Contamination and Toxicology **71:** 0400-0407

Auldridge ME, Guo Y, Austin MB, Ramsey J, Fridman E, Pichersky E, Noel JP (2012) Emergent decarboxylase activity and attenuation of alpha/beta-hydrolase activity during the evolution of methylketone biosynthesis in tomato. Plant Cell **24:** 1596-1607

Battraw M, Hall T (1990) Histochemical analysis of CaMV 35S promoter-*β*-glucuronidase gene expression in transgenic rice plants. Plant Mol Biol **15:** 527-538

Ben-Israel I, Yu G, Austin MB, Bhuiyan N, Auldridge M, Nguyen T, Schauvinhold I, Noel JP, Pichersky E, Fridman E (2009) Multiple biochemical and morphological factors underlie the production of methylketones in tomato trichomes. Plant Physiol **151:** 1952-1964

Benning MM, Wesenberg G, Liu R, Taylor KL, Dunaway-Mariano D, Holden HM (1998) The Three-dimensional Structure of 4-Hydroxybenzoyl-CoA Thioesterase from *Pseudomonas sp.* Strain CBS-3. Journal of Biological Chemistry **273:** 33572-33579

Bird D, Beisson F, Brigham A, Shin J, Greer S, Jetter R, Kunst L, Wu X, Yephremov A, Samuels L (2007) Characterization of Arabidopsis ABCG11/WBC11, an ATP binding cassette (ABC) transporter that is required for cuticular lipid secretion. Plant J **52:** 485-498

Bleeker PM, Mirabella R, Diergaarde PJ, VanDoorn A, Tissier A, Kant MR, Prins M, de Vos M, Haring MA, Schuurink RC (2012) Improved herbivore resistance in cultivated tomato with the sesquiterpene biosynthetic pathway from a wild relative. Proc Natl Acad Sci U S A **109:** 20124-20129

Blum MS, Warter SL (1966) Chemical Releasers of Social Behavior. VII. The Isolation of 2-Heptanone from *Conomyrma pyramica* (Hymenoptera: Formicidae: Dolichoderinae) and Its Modus Operandi as a Releaser of Alarm and Digging Behavior. Annals of the Entomological Society of America **59:** 774-779

Bonaventure G, Salas JJ, Pollard MR, Ohlrogge JB (2003) Disruption of the *FATB* gene in Arabidopsis demonstrates an essential role of saturated fatty acids in plant growth. Plant Cell **15:** 1020-1033

Brand L, Hörler M, Nüesch E, Vassalli S, Barrell P, Yang W, Jefferson RA, Grossniklaus U, Curtis MD (2006) A Versatile and Reliable Two-Component System for Tissue-Specific Gene Induction in Arabidopsis. Plant Physiology **141:** 1194-1204

Broun P, Gettner S, Somerville C (1999) Genetic engineering of plant lipids. Annual Review of Nutrition **19:** 197-216

Browse J, McCourt PJ, Somerville CR (1986) Fatty acid composition of leaf lipids determined after combined digestion and fatty acid methyl ester formation from fresh tissue. Anal Biochem **152:** 141-145

Buchanan BB, Gruissem W, Jones RL (2000) Biochemistry & molecular biology of plants, Vol 40. American Society of Plant Physiologists Rockville

Buttery RG, McFadden WH, Teranishi R, Kealy MP, Mon TR (1963) Constituents of Hop Oil. Nature **200:** 435-436

Campbell MK, Farrell SO (2006) Biochemistry. Cengage Learning Brooks/cole

Cavill GWK, Hinterberger H (1960) The chemistry of ants. IV. Terpenoid constituents of some *Dolichoderus* and *Iridomyrmex* species. Aust. J. Chem. **13:** 514-519

Chodosh LA, Buratowski S (2001) Purification of DNA-Binding Proteins Using Biotin/Streptavidin Affinity Systems. *In* Current Protocols in Protein Science. John Wiley & Sons, Inc.

Choi YE, Lim S, Kim HJ, Han JY, Lee MH, Yang Y, Kim JA, Kim YS (2012) Tobacco NtLTP1, a glandular-specific lipid transfer protein, is required for lipid secretion from glandular trichomes. Plant J **70:** 480-491

Clough SJ, Bent AF (1998) Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. Plant J **16:** 735-743

Cui H, Zhang S-T, Yang H-J, Ji H, Wang X-J (2011) Gene expression profile analysis of tobacco leaf trichomes. BMC Plant Biology **11:** 76

Dillon SC, Bateman A (2004) The Hotdog fold: wrapping up a superfamily of thioesterases and dehydratases. BMC Bioinformatics **5:** 109-109

Efremova N, Schreiber L, Bar S, Heidmann I, Huijser P, Wellesen K, Schwarz-Sommer Z, Saedler H, Yephremov A (2004) Functional conservation and maintenance of expression pattern of FIDDLEHEAD-like genes in *Arabidopsis* and *Antirrhinum*. Plant Mol Biol **56:** 821-837

Ehleringer J (1984) Ecology and ecophysiology of leaf pubescence in North American desert plants. *In* E Rodriguez, PL Healey, I Mehta, eds, Biology and Chemistry of Plant Trichomes. Plenum Press, New York, NY, USA, pp 113-132

Ehrlich PR, Raven PH (1964) Butterfliers and plants - a study in coevolution. Evolution **18:** 586-608

Eisner T, Eisner M, Hoebeke ER (1998) When defense backfires: Detrimental effect of a plant's protective trichomes on an insect beneficial to the plant. Proceedings of the National Academy of Sciences **95:** 4410-4414

Ennajdaoui H, Vachon G, Giacalone C, Besse I, Sallaud C, Herzog M, Tissier A (2010) Trichome specific expression of the tobacco (*Nicotiana sylvestris*) cembratrien-ol synthase genes is controlled by both activating and repressing cis-regions. Plant Mol Biol **73:** 673-685

Eshed Y, Zamir D (1995) An introgression line population of *Lycopersicon pennellii* in the cultivated tomato enables the identification and fine mapping of yield-associated QTL. Genetics **141:** 1147-1162

Falara V, Akhtar TA, Nguyen TT, Spyropoulou EA, Bleeker PM, Schauvinhold I, Matsuba Y, Bonini ME, Schilmiller AL, Last RL, Schuurink RC, Pichersky E (2011) The tomato terpene synthase gene family. Plant Physiology **157:** 770-789

Faul F, Erdfelder E, Lang A-G, Buchner A (2007) G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods **39:** 175-191

Fery RL, Kennedy GG (1987) Genetic analysis of 2-tridecanone concentration, leaf trichome characteristics and tobacco hornworm resistance in tomato. J Amer Soc Hort Sci **112:** 886-891

Forney FW, Markovetz AJ (1971) The biology of methyl ketones. J Lipid Res **12:** 383-395

Forouhar F, Yang Y, Kumar D, Chen Y, Fridman E, Park SW, Chiang Y, Acton TB, Montelione GT, Pichersky E, Klessig DF, Tong L (2005) Structural and biochemical studies identify tobacco SABP2 as a methyl salicylate esterase and implicate it in plant innate immunity. Proc Natl Acad Sci U S A **102:** 1773-1778

Fredricks KM (1967) Products of the oxidation of n-decane by *Pseudomonas aeruginosa* and *Mycobacterium rhodochrous*. Antonie Van Leeuwenhoek **33:** 41-48

Fridman E, Wang J, Iijima Y, Froehlich JE, Gang DR, Ohlrogge J, Pichersky E (2005) Metabolic, genomic, and biochemical analyses of glandular trichomes from the wild tomato species *Lycopersicon hirsutum* identify a key enzyme in the biosynthesis of methylketones. Plant Cell **17:** 1252-1267

Fridmanab E, Koezuka T, Auldridge M, Austin MB, Noel JP, Pichersky E (2006) Chapter five Tomato glandular trichomes as a model system for exploring evolution of specialized metabolism in a single cell. *In* TR John, ed, Recent Advances in Phytochemistry, Vol Volume 40. Elsevier, pp 115-130

Gang DR, Wang J, Dudareva N, Nam KH, Simon JE, Lewinsohn E, Pichersky E (2001) An investigation of the storage and biosynthesis of phenylpropenes in sweet basil. Plant Physiol **125:** 539-555

Gehrig RF, Knight SG (1958) Formation of Ketones from Fatty Acids by Spores of *Penicillium roqueforti*. Nature **182:** 1237-1237

Gershenzon J, Dudareva N (2007) The function of terpene natural products in the natural world. Nat Chem Biol **3:** 408-414

Gershenzon J, McCaskill D, Rajaonarivony JIM, Mihaliak C, Karp F, Croteau R (1992) Isolation of secretory-cells from plant glandular trichomes and their use in biosynthetic-studies of monoterpenes and other gland products. Analytical Biochemistry **200:** 130-138

Ghangas GS, Steffens JC (1993) UDPglucose: fatty acid transglucosylation and transacylation in triacylglucose biosynthesis. Proc Natl Acad Sci U S A **90:** 9911-9915

Glas JJ, Schimmel BC, Alba JM, Escobar-Bravo R, Schuurink RC, Kant MR (2012) Plant glandular trichomes as targets for breeding or engineering of resistance to herbivores. Int J Mol Sci **13:** 17077-17103

Gonçalves MIF, Maluf WR, Gomes LAA, Barbosa LV (1998) Variation of 2-Tridecanone level in tomato plant leaflets and resistance to two mite species (*Tetranychus sp*.). Euphytica **104:** 33-38

Gutiérrez-Alcalá G, Calo L, Gros F, Caissard J-C, Gotor C, Romero LC (2005) A versatile promoter for the expression of proteins in glandular and non-glandular trichomes from a variety of plants. Journal of Experimental Botany **56:** 2487-2494

Hartman JB, St. Clair DA (1998) Variation for Insect Resistance and Horticultural Traits in Tomato Inbred Backcross Populations Derived from *Lycopersicon pennellii*. Crop Sci. **38:** 1501-1508

Henricsson S, Westerholm R, Nilsson S, Berggren B (1996) Chemical characterisation of extractable compounds found in the coating of birch (Betula) pollen. Grana **35:** 179-184

Hollick JB, Gordon MP (1995) Transgenic analysis of a hybrid poplar wound-inducible promoter reveals developmental patterns of expression similar to that of storage protein genes. Plant Physiol **109:** 73-85

Hotelier T, Renault L, Cousin X, Negre V, Marchot P, Chatonnet A (2004) ESTHER, the database of the alpha/beta-hydrolase fold superfamily of proteins. Nucleic Acids Res **32:** D145-147

Hsu C-Y, Creech RG, Jenkins JN, Ma D-P (1999) Analysis of promoter activity of cotton lipid transfer protein gene LTP6 in transgenic tobacco plants. Plant Science **143:** 63-70

Hull R, Covey SN, Dale P (2000) Genetically modified plants and the 35S promoter: assessing the risks and enhancing the debate. Microbial Ecology in Health and Disease **12:** 1-5

Jasperson H, Jones R (1947) Some unsaponifiable constituents of deodorization distillate of vegetable oils. J Soc Chem Ind **66:** 13-17

Jetter R, Kunst L (2008) Plant surface lipid biosynthetic pathways and their utility for metabolic engineering of waxes and hydrocarbon biofuels. Plant J **54:** 670-683

Jing F, Cantu D, Tvaruzkova J, Chipman J, Nikolau B, Yandeau-Nelson M, Reilly P (2011) Phylogenetic and experimental characterization of an acyl-ACP thioesterase family reveals significant diversity in enzymatic specificity and activity. BMC Biochemistry **12:** 1-16

Jolly RC, Kosikowski FV (1975) A New Blue Cheese Food Material from Ultrafiltrated Skim Milk and Microbial Enzyme-Mold Spore Reacted Fat. Journal of dairy science **58:** 1272-1275

Kalapos MP (2003) On the mammalian acetone metabolism: from chemistry to clinical implications. Biochimica et Biophysica Acta (BBA) - General Subjects **1621:** 122-139

Keene CK, Wagner GJ (1985) Direct demonstration of duvatrienediol biosynthesis in glandular heads of tobacco trichomes. Plant Physiol **79:** 1026-1032

Kim J, Kang K, Gonzales-Vigil E, Shi F, Jones AD, Barry CS, Last RL (2012) Striking Natural Diversity in Glandular Trichome Acylsugar Composition Is Shaped by Variation at the Acyltransferase2 Locus in the Wild Tomato *Solanum habrochaites*. Plant Physiology **160:** 1854-1870

Koeduka T, Louie GV, Orlova I, Kish CM, Ibdah M, Wilkerson CG, Bowman ME, Baiga TJ, Noel JP, Dudareva N, Pichersky E (2008) The multiple phenylpropene synthases in both *Clarkia breweri* and *Petunia hybrida* represent two distinct protein lineages. Plant J **54:** 362-374

Koeduka T, Orlova I, Baiga TJ, Noel JP, Dudareva N, Pichersky E (2009) The lack of floral synthesis and emission of isoeugenol in *Petunia axillaris* subsp. *parodii* is due to a mutation in the isoeugenol synthase gene. Plant J **58:** 961-969

Kornberg A, Ochoa S, Mehler AH (1948) Spectrophotometric studies on the decarboxylation of *β*-keto acids. Journal of Biological Chemistry **174:** 159-172

Kraulis PJ (1991) Molscript - a program to produce both detailed and schematic plots of protein structures. Journal of Applied Crystallography **24:** 946-950

Kroumova AB, Wagner GJ (2003) Different elongation pathways in the biosynthesis of acyl groups of trichome exudate sugar esters from various solanaceous plants. Planta **216:** 1013-1021

Lange BM, Turner GW (2013) Terpenoid biosynthesis in trichomes--current status and future opportunities. Plant Biotechnol J **11:** 2-22

Leadbetter ER, Foster JW (1960) Bacterial oxidation of gaseous alkanes. Archiv für Mikrobiologie **35:** 92-104

Liu HC, Creech RG, Jenkins JN, Ma DP (2000) Cloning and promoter analysis of the cotton lipid transfer protein gene *Ltp3*(1). Biochim Biophys Acta **1487:** 106-111

Liu Y, Zamir D (1999) Second generation *L. pennellii* introgression lines and the concept of bin mapping. Tomato Genet Coop **49:** 26-30

Luckwill LC (1943) The genus *Lycopersicon*. The University press, Aberdeen

Lukins HB, Foster JW (1963) Methyl Ketone Metabolism in Hydrocarbon-Utilizing *Mycobacteria*. J Bacteriol **85:** 1074-1087

Maluf WR, Barbosa LV, Costa Santa-Cecília LV (1997) 2-Tridecanone-mediated mechanisms of resistance to the South American tomato pinworm *Scrobipalpuloides absoluta* (Meyrick, 1917) (Lepidoptera-Gelechiidae) in *Lycopersicon spp*. Euphytica **93:** 189-194

Mathur J, Chua NH (2000) Microtubule stabilization leads to growth reorientation in Arabidopsis trichomes. Plant Cell **12:** 465-477

Matiasek MG, Choudhury K, Nemecek-Marshall M, Fall R (2001) Volatile ketone formation in bacteria: release of 3-oxopentanoate by soil pseudomonads during growth on heptanoate. Curr Microbiol **42:** 276-281

Mayer KM, Shanklin J (2005) A structural model of the plant acyl-acyl carrier protein thioesterase FatB comprises two helix/4-stranded sheet domains, the N-terminal domain containing residues that affect specificity and the C-terminal domain containing catalytic residues. J Biol Chem **280:** 3621-3627

McCaskill D, Croteau R (1999) Strategies for bioengineering the development and metabolism of glandular tissues in plants. Nat Biotech **17:** 31-36

McCormick S (1991) Transformation of tomato with *Agrobacterium tumefaciens*. Plant Tissue Culture Manual **B6:** 1-9

McDowell ET, Kapteyn J, Schmidt A, Li C, Kang J-H, Descour A, Shi F, Larson M, Schilmiller A, An L, Jones AD, Pichersky E, Soderlund CA, Gang DR (2011) Comparative Functional Genomic Analysis of *Solanum* Glandular Trichome Types. Plant Physiology **155:** 524-539

Michelmore RW, Paran I, Kesseli RV (1991) Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations. Proc Natl Acad Sci U S A **88:** 9828-9832

Mou Z, He Y, Dai Y, Liu X, Li J (2000) Deficiency in fatty acid synthase leads to premature cell death and dramatic alterations in plant morphology. Plant Cell **12:** 405-418

Nagegowda DA, Gutensohn M, Wilkerson CG, Dudareva N (2008) Two nearly identical terpene synthases catalyze the formation of nerolidol and linalool in snapdragon flowers. Plant J **55:** 224-239

Nagel J, Culley LK, Lu Y, Liu E, Matthews PD, Stevens JF, Page JE (2008) EST Analysis of Hop Glandular Trichomes Identifies an O-Methyltransferase That Catalyzes the Biosynthesis of Xanthohumol. The Plant Cell Online **20:** 186-200

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

Ohlrogge J, Browse J (1995) Lipid biosynthesis. Plant Cell **7:** 957-970

Papachristoforou A, Kagiava A, Papaefthimiou C, Termentzi A, Fokialakis N, Skaltsounis A-L, Watkins M, Arnold G, Theophilidis G (2012) The Bite of the Honeybee: 2-Heptanone Secreted from Honeybee Mandibles during a Bite Acts as a Local Anaesthetic in Insects and Mammals. PLoS ONE **7:** e47432

Park J, Rodríguez-Moyá M, Li M, Pichersky E, San K-Y, Gonzalez R (2012) Synthesis of methyl ketones by metabolically engineered *Escherichia coli*. Journal of Industrial Microbiology & Biotechnology **39:** 1703-1712

Payne WW (1978) A Glossary of Plant Hair Terminology. Brittonia **30:** 239-255

Regnier FE, Wilson EO (1968) The alarm-defence system of the ant *Acanthomyops claviger*. Journal of Insect Physiology **14:** 955-970

Reis Cd, Sajo MdG, Stehmann JR (2002) Leaf Structure and Taxonomy of *Petunia* and *Calibrachoa* (Solanaceae). Brazilian Archives of Biology and Technology **45:** 59-66

Salas JJ, Ohlrogge JB (2002) Characterization of substrate specificity of plant FatA and FatB acyl-ACP thioesterases. Arch Biochem Biophys **403:** 25-34

Sali A, Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol **234:** 779-815

Schilmiller A, Shi F, Kim J, Charbonneau AL, Holmes D, Jones AD, Last RL (2010a) Mass spectrometry screening reveals widespread diversity in trichome specialized metabolites of tomato chromosomal substitution lines. Plant Journal **62:** 391-403

Schilmiller AL, Charbonneau AL, Last RL (2012a) Identification of a BAHD acetyltransferase that produces protective acyl sugars in tomato trichomes. Proc Natl Acad Sci U S A **109:** 16377-16382

Schilmiller AL, Last RL, Pichersky E (2008) Harnessing plant trichome biochemistry for the production of useful compounds. The Plant Journal **54:** 702-711

Schilmiller AL, Miner DP, Larson M, McDowell E, Gang DR, Wilkerson C, Last RL (2010b) Studies of a biochemical factory: tomato trichome deep expressed sequence tag sequencing and proteomics. Plant Physiol **153:** 1212-1223

Schilmiller AL, Pichersky E, Last RL (2012b) Taming the hydra of specialized metabolism: how systems biology and comparative approaches are revolutionizing plant biochemistry. Current Opinion in Plant Biology **15:** 338-344

Schilmiller AL, Schauvinhold I, Larson M, Xu R, Charbonneau AL, Schmidt A, Wilkerson C, Last RL, Pichersky E (2009) Monoterpenes in the glandular trichomes of tomato are synthesized from a neryl diphosphate precursor rather than geranyl diphosphate. Proceedings of the National Academy of Sciences **106:** 10865-10870

Schmidt A, Li C, Shi F, Jones AD, Pichersky E (2011) Polymethylated Myricetin in Trichomes of the Wild Tomato Species *Solanum habrochaites* and Characterization of Trichome-Specific 3′/5′- and 7/4′-Myricetin O-Methyltransferases. Plant Physiology **155:** 1999-2009

Shangguan XX, Xu B, Yu ZX, Wang LJ, Chen XY (2008) Promoter of a cotton fibre *MYB* gene functional in trichomes of Arabidopsis and glandular trichomes of tobacco. J Exp Bot **59:** 3533-3542

Shearer DA, Boch R (1965) 2-Heptanone in the Mandibular Gland Secretion of the Honey-bee. Nature **206:** 530

Shepherd RW, Bass WT, Houtz RL, Wagner GJ (2005) Phylloplanins of Tobacco Are Defensive Proteins Deployed on Aerial Surfaces by Short Glandular Trichomes. The Plant Cell Online **17:** 1851-1861

Sinha N, Lynch M (1998) Fused organs in the adherent1 mutation in maize show altered epidermal walls with no perturbations in tissue identities. Planta **206:** 184-195

Slocombe SP, Schauvinhold I, McQuinn RP, Besser K, Welsby NA, Harper A, Aziz N, Li Y, Larson TR, Giovannoni J, Dixon RA, Broun P (2008) Transcriptomic and Reverse Genetic Analysesof Branched-Chain Fatty Acid and Acyl Sugar Production in *Solanum pennellii* and *Nicotiana benthamiana*. Plant Physiology **148:** 1830-1846

Snyder JC, Carter CD (1985) Trichomes on leaves of *lycopersicon-hirsutum, lycopersicon-esculentum* and their hybrids. Euphytica **34:** 53-64

Sunilkumar G, Mohr L, Lopata-Finch E, Emani C, Rathore K (2002) Developmental and tissue-specific expression of CaMV 35S promoter in cotton as revealed by GFP. Plant Mol Biol **50:** 463-479

Szymanski DB, Marks MD, Wick SM (1999) Organized *F-Actin* Is Essential for Normal Trichome Morphogenesis in Arabidopsis. The Plant Cell Online **11:** 2331-2347

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

Tissier A (2012) Glandular trichomes: what comes after expressed sequence tags? The Plant Journal **70:** 51-68

Treutter D (2005) Significance of flavonoids in plant resistance and enhancement of their biosynthesis. Plant Biol (Stuttg) **7:** 581-591

Turner G, Gershenzon J, Nielson EE, Froehlich JE, Croteau R (1999) Limonene synthase, the enzyme responsible for monoterpene biosynthesis in peppermint, is localized to leucoplasts of oil gland secretory cells. Plant Physiol **120:** 879-886

Tzfira T, Tian GW, Lacroix B, Vyas S, Li J, Leitner-Dagan Y, Krichevsky A, Taylor T, Vainstein A, Citovsky V (2005) pSAT vectors: a modular series of plasmids for autofluorescent protein tagging and expression of multiple genes in plants. Plant Mol Biol **57:** 503-516

van der Hoeven RS, Monforte AJ, Breeden D, Tanksley SD, Steffens JC (2000) Genetic Control and Evolution of Sesquiterpene Biosynthesis in *Lycopersicon esculentum* and *L. hirsutum*. The Plant Cell Online **12:** 2283-2294

van Schie CCN, Haring MA, Schuurink RC (2007) Tomato linalool synthase is induced in trichomes by jasmonic acid. Plant Mol Biol **64:** 251-263

Wagner GJ (1991) Secreting glandular trichomes: more than just hairs. Plant Physiol **96:** 675-679

Wagner GJ, Wang E, Shepherd RW (2004) New Approaches for Studying and Exploiting an Old Protuberance, the Plant Trichome. Annals of Botany **93:** 3-11

Wang E, Gan S, Wagner GJ (2002) Isolation and characterization of the *CYP71D16* trichome-specific promoter from *Nicotiana tabacum* L. Journal of Experimental Botany **53:** 1891-1897

Wang E, Wagner G (2003) Elucidation of the functions of genes central to diterpene metabolism in tobacco trichomes using posttranscriptional gene silencing. Planta **216:** 686-691

Watts F (1886) XXXII.-On the essential oil of lime leaves (*Citrus limetta*). Preliminary notice. Journal of the Chemical Society, Transactions **49:** 316-317

Werker E (2000) Trichome diversity and development. *In* Advances in Botanical Research, Vol Volume 31. Academic Press, pp 1-35

Williams CG (1857) On the Constitution of the Essential Oil of Rue. Proceedings of the Royal Society of London **9:** 167-169

Williams WG, Kennedy GG, Yamamoto RT, Thacker JD, Bordner J (1980) 2-Tridecanone: A Naturally Occurring Insecticide from the Wild Tomato *Lycopersicon hirsutum* f.*glabratum*. Science **207:** 888-889

Wu GZ, Xue HW (2010) Arabidopsis beta-ketoacyl-[acyl carrier protein] synthase i is crucial for fatty acid synthesis and plays a role in chloroplast division and embryo development. Plant Cell **22:** 3726-3744

Xie Z, Kapteyn J, Gang DR (2008) A systems biology investigation of the MEP/terpenoid and shikimate/phenylpropanoid pathways points to multiple levels of metabolic control in sweet basil glandular trichomes. Plant J **54:** 349-361

Yang Y, Xu R, Ma C-J, Vlot AC, Klessig DF, Pichersky E (2008) Inactive methyl indole-3-acetic acid ester can be hydrolyzed and activated by several esterases belonging to the AtMES esterase family of Arabidopsis. Plant Physiol **147:** 1034-1045

Yang YN, Li RG, Qi M (2000) *In vivo* analysis of plant promoters and transcription factors by agroinfiltration of tobacco leaves. Plant Journal **22:** 543-551

Yu G, Nguyen TT, Guo Y, Schauvinhold I, Auldridge ME, Bhuiyan N, Ben-Israel I, Iijima Y, Fridman E, Noel JP, Pichersky E (2010) Enzymatic functions of wild tomato methylketone synthases 1 and 2. Plant Physiol. **154:** 67-77

Yu G, Schilmiller AL, Last RL, Pichersky E (Submitted) Specificity of two trichomal gland cell promoters from Solanum across species boundaries and evidence for the presence of plastids in gland cells.

Zamir D, Ben-David TS, Rudich J, Juvik JA (1984) Frequency distributions and linkage relationships of 2-tridecanone in interspecific segregating generations of tomato. Euphytica **33:** 481-488

Zuriaga E, Blanca J, Nuez F (2009) Classification and phylogenetic relationships in *Solanum* section *Lycopersicon* based on AFLP and two nuclear gene sequences. Genetic Resources and Crop Evolution **56:** 663-678