2015

ENGINEERING TRITERPENE METABOLISM IN TOBACCO

Zuodong Jiang
University of Kentucky, zuodong.j@uky.edu

STUDENT AGREEMENT:

I represent that my thesis or dissertation and abstract are my original work. Proper attribution has been given to all outside sources. I understand that I am solely responsible for obtaining any needed copyright permissions. I have obtained needed written permission statement(s) from the owner(s) of each third-party copyrighted matter to be included in my work, allowing electronic distribution (if such use is not permitted by the fair use doctrine) which will be submitted to UKnowledge as Additional File.

I hereby grant to The University of Kentucky and its agents the irrevocable, non-exclusive, and royalty-free license to archive and make accessible my work in whole or in part in all forms of media, now or hereafter known. I agree that the document mentioned above may be made available immediately for worldwide access unless an embargo applies.

I retain all other ownership rights to the copyright of my work. I also retain the right to use in future works (such as articles or books) all or part of my work. I understand that I am free to register the copyright to my work.

REVIEW, APPROVAL AND ACCEPTANCE

The document mentioned above has been reviewed and accepted by the student's advisor, on behalf of the advisory committee, and by the Director of Graduate Studies (DGS), on behalf of the program; we verify that this is the final, approved version of the student's thesis including all changes required by the advisory committee. The undersigned agree to abide by the statements above.

Zuodong Jiang, Student

Dr. Joe Chappell, Major Professor

Dr. Arthur G Hunt, Director of Graduate Studies

ENGINEERING TRITERPENE METABOLISM IN TOBACCO

DISSERTATION

A dissertation submitted in partial fulfillment of the
requirements for the degree of Doctor of Philosophy in the
College of Agriculture, Food and Environment
at the University of Kentucky

By
Zuodong Jiang

Lexington, Kentucky

Director: Dr. Joe Chappell, Professor of Plant Physiology

Lexington, Kentucky

2015

ABSTRACT OF DISSERTATION

ENGINEERING TRITERPENE METABOLISM IN TOBACCO

Terpenes comprise a large diverse class of natural products and many of them attract interest because of their physiological function, therapeutic and industrial values. Triterpene oils including squalene (C30), botrycococcene (C30) and their methylated derivatives (C31-C37) generated by the green algae *Botryococcus braunii* race B, which have recently received significant attention because of their utility for advanced biofuels. However, the slow growth habit of *B. braunii* makes it impractical as a robust biofuel production system. In this thesis, we firstly evaluated the potential of generating high levels of triterpene (C30) production in tobacco plants by diverting carbon flux from cytosolic MVA pathway or plastidic MEP pathway by overexpressing avian farnesyl diphosphate synthase along with triterpene synthase targeted to the cytoplasm or the chloroplast of cells. Up to 1,000 µg/g fresh weight of squalene and 544 µg/g fresh weight of botryococcene was achieved in our transgenic plants with this metabolism direct to the chloroplasts, which is about approximately 100-times greater than that accumulating in the plants engineered for cytosolic production. To test if methylated triterpenes can be produced in tobacco, we also engineered triterpene methyltransferases (TMTs) into wild type plants and transgenic tobacco plants selected for high level triterpene accumulation. We observed that up to 91% of the total triterpene content was converted to methylated forms (C31, C32) by targeting the TMTs to the chloroplasts of transgenic plants, whereas only 4-14% of total triterpenes were methylated when TMTs were directed to the cytoplasm. Select transgenic lines were growing in field studies from 2011 to 2014 to evaluate their physiological performance under field conditions. Surprisingly, the field studies suggested that the growth and agronomic performance of the transgenic lines accumulating squalene were not compromised, while those accumulating high levels of botryococcene were only 72%-76% as tall, had about 59%-75% of the leaf area, and about 55%-75% of the biomass as wild type plants. Yet, these transgenic plants had photosynthetic capacity equal to the wild type plants

KEYWORDS: squalene, botryococcene, methylated triterpene, tobacco, biofuel

ZUODONG JIANG

Student's Signature

01.22.2015

Date

ENGINEERING TRITERPENE METABOLISM IN TOBACCO

By

Zuodong Jiang

Joe Chappell

Director of Dissertation

Arthur G Hunt

Director of Graduate Studies

01.22.2015

ACKNOWLEDGMENTS

This thesis would not have come about without the support from my mentor, Joe Chappell. He provided valuable training, advice and long-term encouragement throughout my graduate career. Without his encouragement, I would not have been able to accomplish all the work reported in this thesis. I will forever be grateful to him for helping me become who I am today. I also thank my committee members Arthur Hunt, Ling Yuan, Steven Van Lanen and my outside examiner Erich Grotewold for their support and critical comments. Next, I wish to thank our former lab member Dr. Xun Zhuang who gave important suggestions on experimental design and performed the initial work of diterpene engineering in yeast. My achievements would also not have been possible without the help and encouragement of all the Chappell laboratory members. Last but not least, I also thank my wife Yi Shu and family for their continual support of my life and career.

TABLE OF CONTENTS

List of Tables

List of Figures

Chapter 1: Background and Introduction

1.1 Terpene metabolism in plants

Terpenes and terpenoids comprise a large diverse class of natural products (Buckingham, 2003) and many of them have attracted interest because of their therapeutic and industrial value (Dewick, 2009). For example, terpenes extracted from plants are used as anti-cancer and anti-malarial drugs (Cragg, 1998; Dhingra et al., 2000). These valuable compounds are commonly isolated from plants, microbes and marine organisms. However, most of these compounds are produced in very small amounts by their natural host and often not as single compounds but as complex mixtures. Hence, there have been many efforts to engineer terpene metabolism into model organisms such as bacteria, yeast and plants, in order to produce large quantities for chemical identification and functional characterization (Zook et al., 1996; Wallaart et al., 2001; Martin et al., 2003; Kirby et al., 2008).

In plants, terpene biosynthesis occurs via mevalonate (MVA) pathway operating in the cytoplasm and the methyl-erythritol phosphate (MEP) pathway operating in the chloroplast compartment (Figure 1.1). By two different routes, both pathways can produce isopentenyl diphosphate (IPP) and its isomer dimethylallyl diphosphate (DMAPP), which are C5 building blocks for specific class of terpenes. The cytosolic pathway (MVA) is mainly responsible for the generation of terpenes like sterols and sesquiterpenes, compounds consisting of 30 (C30) or 15 (C15) carbons, respectively. Monoterpenes (C10), diterpenes (C20) and carotenoids (C40) are synthesized via the plastidic pathway (MEP). Terpenes play important roles in plants, including providing essential metabolites for general growth and development (Clouse, 2011; Kohlen et al., 2011) as well as molecules mediating interactions between plants and their biotic and abiotic environments (Keeling and Bohlmann, 2006; Kegge and Pierik, 2010; Huffaker et al., 2011).

1.2 Strategies for metabolic engineering terpene metabolism in plants

Engineering terpene metabolism in plant system is attractive because plants rely on photosynthesis for growth instead of an exogenous carbon source. However, because of complex innate regulation of terpene metabolism, limited success has been achieved in generating high-level terpene products in plants. In fact, any production platform can not

be considered economically viable until the value of the compound produced exceeds the production costs. Obvious means for addressing such limitation are to increase production levels per plant and to generate high-valued products. In a general conceptual sense, producing compounds equivalent to 1% of the dry weight of the plant is often mentioned as a minimal target level (Horsch, 1993; Snell and Peoples, 2009).

However, plants are more complex multicellular organisms compared to microbial system, which make them more difficult to genetically engineer. Although efforts have indicated that metabolically engineering of plants is feasible, there are still important considerations, technologies, and strategies that are limiting successful development of strategies for plant production platforms. These important considerations include macro-scale issues such a selection of an appropriate plant host to engineer to more detailed considerations such as tissue-specific and cell-specific targeting of the engineered metabolism to assure success. Micro-scale and molecular details also need to be considered. Overall, there are three essential components that underwrite any engineering strategy: 1. manipulation of gene expression (e.g. using heterologous genes from a species unrelated to the host under a promoter that drives the desired temporal and/or spatial expression patterns, or altering endogenous gene expression to change flux towards the desired pathway); 2. design to avoid endogenous regulation (*e.g.* using specific hosts and targeting the introduced pathway to specific tissues and subcellular compartments); and 3. combining these two considerations in unique and novel ways with the available technologies (e.g. using metabolic models, synthetic biology to ensure enzyme cooperation, use of transporters to sequester accumulated compounds and avoid toxicity problems). Besides the choice of plant host, these variables can be controlled to large extent by construct design and are best summarized in Figure 1.2.

1.3 Engineering isoprenoid metabolism in plant organelles

As noted earlier, metabolic flux in isoprenoid biosynthesis revolves around a complex network that involves multiple subcellular compartments within the plant cell: the cytosolic MVA pathway, the plastid-localized MEP pathway, localization of many sequential isoprenoid biosynthetic enzymes arrayed on the ER membrane or targeted to mitochondria and microbodies, and the possible participation of the vacuole as a storage/sequestration organelle (Figure 1.2). Hence, targeting of enzymes and biosynthetic capacity to an appropriate subcellular location is a prerequisite for

successful metabolic engineering in plants, which must overcome several layers of complicated regulation (Heinig et al., 2013a; Lange and Turner, 2013). Early metabolic engineering attempts achieved less success in elevating terpene production by introducing enzymes into the cytoplasm without considering the regulatory complexity of the MVA pathway (Wu et al., 2006). Now, engineering terpene metabolism into organelles to overcome these innate regulatory constraints has become an important strategy, as illustrated by the following two examples.

It is well established that the biosynthesis of distinct classes of isoprenoids take place in distinct organelles (Figure 1). For example, the major steps of monoterpenes, diterpenes and carotenoids biosynthesis are known to operate in higher plant chloroplasts, so most of the engineering efforts to enhance biosynthesis of these compounds have used heterologous enzymes fused with a plastid-targeting signal peptide to direct them to their endogenous site of biosynthesis (or by introducing the transgene into the plastid genome directly). This type of “straight-forward” genetic engineering has been conducted in various species of higher plants, extensively reviewed (Fraser et al., 2009; Misawa, 2009; Beyer, 2010; Bai et al., 2011; Farré et al., 2011; Wurtzel et al., 2012; Lange and Turner, 2013; Morandini, 2013) and will not be reiterated in this section other than to note that this type of engineering utilizes pre-existing substrate biosynthesis, which is captured and diverted to produce compounds that may or may not be normally present within a particular organelle. Hence, this type of strategy does not completely avoid endogenous regulatory elements that may respond directly to the synthesized compound or indirectly through changes in intermediates and large amounts of target compound accumulation. It is also necessary to be aware of any inherent regulatory mechanisms (*e.g.* allosteric sites) of the introduced enzymes that may subject these engineered activities to non-native regulation.

In order to avoid this type of regulation, targeting introduced enzymes to particular organelles to divert available substrate(s) for the formation of a novel pathway that is foreign to the organelle has become a commonplace strategy. In an early study, a strawberry linalool/nerolidol synthase *FaNES1*, was targeted to mitochondria by fusing a mitochondrial targeting signal sequence to the amino terminus of the *FaNES1* protein, leading to generation of two novel sesquiterpenes in transgenic *Arabidopsis*, (3S)-(E)-nerolidol and (*E*)-DMNT, neither of which are found in WT lines (Kappers et al., 2005). In another recent study (Farhi et al., 2011), amorphadiene-4,11-diene synthase (ADS)

targeted to the mitochondria was co-expressed with a cytochrome P450 (CYP71AV1), a cytochrome P450 reductase (CPR), an artemisinic aldehyde reductase (DBR2) from A. annua, and a truncated and deregulated HMGR from yeast in *N. tabacum*. This resulted in artemisinin accumulation of 5-7 mg/g DW, which is 8-fold more than when ADS was targeted to the cytoplasm (Farhi et al., 2011). These results provide strong evidence that mitochondrial targeted FaNES1 and ADS have access to FPP in the mitochondria and convert this intermediate into (*3S*)-(*E*)-nerolidol and amorpha-4,11-diene, respectively. These reaction intermediates were then were further converted by native (and unknown enzymes in *Arabidopsis*) or engineered enzymes (CYP71AV1, CPR, DBR2) present in cytosol to the final products (*E*)-DMNT and artemisinin, respectively. Although it is not clear how the intermediate product (*3S*)-(*E*)-nerolidol and amorpha-4,11-diene are shuttled between the mitochondria and cytosol, the results show that the heterologous terpene synthases can efficiently use the FPP pool in the mitochondria for novel sesquiterpene production. In contrast, free pools of FPP in chloroplasts and the cytoplasm must be less available in these species because simply introducing a FPP-dependent synthase does not yield appreciable novel terpene accumulation (Aharoni et al., 2003; Kappers et al., 2005; Wu et al., 2006; Farhi et al., 2011)

Wu et al. (2006) furthered this approach by applying it to chloroplasts. Accumulation greater than ~25 µg/g FW of the non-native sesquiterpenes patchoulol and amorpha-4,11-diene, was achieved in transgenic *N. tabacum* when an avian FPS gene and non-native sesquiterpene synthases, ADS or patchoulol synthase (PTS), genes were transformed into the nuclear genome with plastid targeting signal sequences appended to the amino-terminus of encoded proteins. Using the same strategy, up to ~500 µg/g FW of the triterpenes squalene and botryococcene were achieved when an avian FPS gene was co-expressed with either a yeast squalene synthase (SQS) or an engineered algal botryococcene synthase (BS) targeted to the chloroplast compartment (Wu et al., 2012 and Jiang et al., unpublished). Conceptually, the plastid targeted FPS diverts the IPP and DMAPP intermediates from the MEP pathway towards the accumulation of free FPP, the substrate for sesquiterpene and triterpene production. The biosynthesis of FPP, sesquiterpenes, and triterpenes are foreign to the chloroplast, and introduction of these two-step biochemical pathways allows for elevated non-native terpene production in the plastid compartment, sheltering these non-native biosynthetic pathways from any native regulatory mechanisms and hence allowing for an unlimited flux of carbon to a desired

terpene. Kumar et al., (2012) corroborated this notion in an independent study where the entire yeast MVA pathway (a total of six enzymes) was introduced into the chloroplast genome of tobacco.The resulting homoplasmic transgenic lines accumulated multiple isoprenoid products, including mevalonate, carotenoids, sterols, squalene, and interestingly, triacylglycerides.

When engineering novel biosynthetic capacity into the chloroplast, there is an important choice to be made between plastidic transformation versus nuclear transformation. Plastidic transformation offers several advantages over nuclear transformation: homologous recombination methodology exists, expression of transgene operons could improve coordinated gene expression, transgene inheritance should only pass from maternal tissue, nuclear epigenetic affects should not present any difficulties, and the translation of expressed transgenes into protein is likely to be higher compared to nuclear transgene expression (Daniell et al., 2005; Daniell, 2006; Kumar et al., 2012). However, there are no any direct comparisons of isoprenoid production by plastid genome transformations versus nuclear genome transformations expressing the same enzymes within or targeted to the chloroplast. Moreover, the higher level of protein expression does not always positively correlate with higher level of terpene production and reflects the need to consider many factors, such as: protein (enzyme) activity, substrate availability, flux control within the pathway, and other regulatory complexities which may exist in the plastid. Thus, the best transformation strategy, nuclear versus plastidic, will probably vary on a case-by-case basis.

1.4 Altering expression patterns of endogenous genes

1.4.1 Down-regulating gene expression

Down-regulation or knockout of endogenous biosynthetic genes is another important strategy commonly used to regulate/re-direct metabolic pathway flux. These efforts attempt to suppress or abolish gene expression of a possible competing enzyme, thus flux can be redirected into desired enzyme/pathway. However, this technique has not been used substantially and effectively in plant metabolic engineering compared with efforts in microbial systems. This is, in part, because of the difficulties in obtaining specific plant mutants and the low efficiency in obtaining appropriate amounts of down-regulation of gene expression in plants (due mainly to large unknown genomes and/or

redundant genes). Nevertheless, a number of studies have reported that anti-sense RNA and RNAi (RNA interference) techniques have been successfully used in manipulating plant terpene metabolism to increase terpenoid production.

Monoterpene essential oil production was been elevated (61% yield increase over WT plants) in transgenic peppermint expressing peppermint antisense (+)-menthofuran synthase (MFS) with simultaneous overexpression of DXR (Lange et al., 2011a). Down-regulation of MFS alone was shown to decrease the level of side-product (+)-menthofuran, and redirected carbon flux to desirable monoterpene oil production, leading to an increased oil yield by roughly 35% (Mahmoud and Croteau, 2001).

Tuber-specific expression of antisense fragments for either lycopene cyclase (*LCY-e*) or *ß-carotene hydroxylase*, the genes encoding the enzymes which compete for lycopene and further metabolism of ß-carotene, respectively, lead to significantly increased levels of ß-carotene (up to 14-fold and 38-fold, respectively) and total carotenoids (up to 2.5-fold and 4.5-fold, respectively) in potato tuber (Diretto et al., 2006; Diretto et al., 2007). When *LCY-e* was suppressed by an RNAi approach, increased carotenoid content in *B. napus* seeds was reported (Yu et al., 2008). RNAi was also used in several studies to increase the content of artemisinin in A. annua by down-regulating SQS and ß-caryophyllene synthase, both enzymes that compete for FPP, which is a key intermediate in the artemisinin biosynthetic pathway (Feng et al., 2009; Zhang et al., 2009; Chen et al., 2011). When the *Catharanthus roseus* gene encoding 7-deoxyloganic acid 7-hydroxylase (*CrDL7H*), which is involved in secologanin biosynthesis, were suppressed by virus-induced gene silencing the accumulation of secologanin was reduced by at least 70%. Critically, the accumulation of the intermediate, 7-deoxyloganic acid (the substrate for *CrDL7H*), was 4 mg/g FW in silenced plants while this compound is normally undetectable in WT plants (Salim et al., 2013).

Transgenic oranges with reduced levels of limonene caused by an antisense down-regulation of the (+)-limonene synthase gene, were shown to be resistant to economically important pathogens (Rodríguez et al., 2011). Exactly how the reduced level of limonene in fruits activates a defense response has yet to be determined, but reduced limonene accumulation does correlate with increased levels of *GGPS* which could provide substrate for the formation of diterpene antimicrobial compounds that inhibit pathogen infection. The authors of this study also suggest that a link between

limonene accumulation and pathogen attack could play an ecological role in facilitating seed dispersal by allowing herbivores easier access to the fruit pulp (Rodríguez et al., 2014).

1.4.2 Transcription factors

Many specialized metabolites in plants accumulate when plants respond to acute developmental or environmental signals (Chappell and Nable, 1987). Therefore, it seems reasonable that their biosynthetic genes would be regulated in a coordinated manner by transcriptional factors. Although transcriptional regulation of the isoprenoid biosynthetic pathways is not well characterized, metabolic engineering of certain regulatory genes (the transcriptional factors) provides a novel approach to enhance terpene production in plants (Patra et al., 2013).

Recent studies have reported transcription factors in A. annua that appear to regulate artemisinin biosynthesis. *AaWRKY1* was characterized as a transcription factor that regulates the native ADS gene in *A. annua*. Transient expression of *AaWRKY1* also led to increased transcript accumulation of the majority of artemisinin biosynthetic genes (Ma et al., 2009). Two jasmonate-responsive (*AaERF1* and *AaERF2*) and a trichome-specific (*AaORA*) AP2/ERF transcription factors were also characterized as positive regulators for artemisinin biosynthesis in *A. annua*. Overexpressing either transcription factor resulted in increased accumulation of artemisinin and artemisinic acid (Yu et al., 2012; Lu et al., 2013). In contrast and contrary to expectations, constitutive expression of an *Arabidopsis* blue light receptor, *CRY1*, gene in *A. annua* increased the transcript abundance for *FPS*, *ADS*, and *CYP71AV1*, three important enzymes in artemisinin biosynthesis, and lead to 30~40% increases in the artemisinin and anthocyanins (Hong et al., 2009).

Catharanthine accumulation was improved up to 6.5-fold higher than WT in *C. roseus* hairy roots by co-expression of the *ORCA3* transcription factor and the gene encoding for geraniol 10-hydroxylase, an enzyme involved in the terpenoid indole alkaloid (TIA) biosynthetic pathway (Wang et al., 2010). Likewise, overexpression of the *Arabidopsis* transcription factor Agamous-like 12 in *C. roseus* suspension cells promoted enhanced accumulation of ajmalicine, a TIA with antihypertensive properties (Montiel et al., 2007). *Arabidopsis* transcription factors MYC2 and MYB21 have also been reported to regulate

the expression of sesquiterpene synthase genes *TPS11* and *TPS21*. When mutated, the *Arabidopsis myc2* and *myb21* mutants emit less sesquiterpene volatiles from their flowers than the WT plants, which was correlated with reduced levels of the *TPS11* and *TPS21* mRNAs (Hong et al., 2012; Reeves et al., 2012).

1.5 Engineering terpene metabolism by trichome specific gene expression

Trichomes encompass a group of specialized cells that originate from the epidermis of plant tissues and are differentiated on the basis of their biochemical capabilities. A recent review by Lange and Turner (2013) summarizes the current knowledge of isoprenoid biosynthesis in trichomes and touches upon how our current understanding in trichome biology might be harnessed for use in metabolic engineering efforts. Briefly, there are several types of trichomes but glandular trichomes (GTs) are the most important in terms of specialized metabolism capacity due to their seemingly dedicated role as metabolite production factories. The presence, number, and type of trichomes vary between plant species. Thus, the capacity to engineer them does not exist for every engineering project (although, this could be an important factor in choosing a particular production host). Equally important, glandular trichomes can secrete their products onto the leaf surface, facilitating collection, or the glandular head synthesized compounds may accumulate in cavities/sacs associated with the metabolically active cells. Trichomes secreting hydrophobic compounds like isoprenoids generally have interesting intracellular features like an extensive smooth ER network that maintains contact with non-pigmented leucoplasts. The leucoplasts appear to have a non-uniform shape which could be implicated in increasing the contact surface area with the smooth ER (Lange and Turner, 2013). These extensive connections between intracellular membrane networks may be critical for transport of large amounts of hydrophobic compounds. A recent study has also implicated a possible role of a lipid transfer protein in the export of isoprenoids from *Nicotiana* tall GTs (Choi et al., 2012).

The species in which isoprenoid biology of the trichomes has been best studied is the pathway leading to the production of (-)-menthol in *Mentha x piperita* (peppermint), which occurs exclusively in this plants' trichomes. The biosynthesis and subcellular organization of (-)-menthol production has been well described (Croteau et al., 2005). The production of this monoterpene through the MEP pathway illustrates the robust carbon flux through trichome plastids. Further support for a high carbon flux through the

MEP pathway in trichomes is the high level of Δ^9-tetrahydrocannabinolic acid accumulation in *Cannabis sativa*. Tetrahydrocannabinolic acid contains an integrated GPP molecule attached to a phenolic precursor which allows for further carbon- and lactone- cyclization yielding the final product (Lange and Turner, 2013).Trichome-specific promoters have also been identified which enable trichome specific transgene targeting and novel biotechnology applications (summarized in Tissier, 2012). The natural capacity for trichomes to function as specialized biochemical factories, in fact, makes them intriguing targets for the redirection of flux into heterologous, high-valued compounds (Wu et al., 2012).

1.6 The unique triterpene and methylated triterpene biosynthesis in *Botryococcus Braunii* race B

Botryococcus braunii, an ancient green algae, accumulates 30-86% of its dry weight as hydrocarbon oils. Three races of *Botryococcus braunii* have been identified race A, B and L. Triterpenes in *Botryococcus braunii* race B are of particular value because this algae has directly contributed to existing oil and coal shale deposits found on Earth, accounting for up to 1.4% of total hydrocarbon content in oil shales (Moldowan and Seifert, 1980). However, because *B. braunii* is such a slow growing alga, the use of it for large-scale cultivation and oil production seems interesting but unrealistic. Therefore characterization of genes for botryococcene and triterpene methylation is important and will make it possible for engineering triterpene metabolism into other robust systems for renewable petrochemcals and biofuels production (Banerjee et al., 2002).

Squalene and botryocene are the two triterpene species accumulating in *B. braunii* race B, and they share similarity not only in their structure but also in their biosynthesis pathway. Both compounds are derived from two 15-carbon farnesyl (FPP) residues: squalene accurs from an intial head to head condensation of two FPPs into the stable intermediate presqualene diphosphate (PSPP), followed by a reductive rearrangement to form 1-1’ linkage final product. Botryococcene has the same first half reaction, but the intermediate PSPP is converted via a different reductive rearrangement to form 1’-3 linkage final product (Fig. 4A) (Okada et al., 2000).

Squalene synthase has been extensively studied, in part because it is found in all eukaryotes. Based on the assumption that squalene synthase and botrycocene synthase

would share similarity in both their sequence and catalytic function, the *B. braunii* squalene synthase cDNA was used to screen *B. braunii* cDNA library under low stringency hybridization conditions, a unique squalene synthase-like gene (SSL-1) was isolated. Other two additional squalene synthase-like (SSL) genes, SSL-2, and SSL-3 were discovered by exhaustive sequencing assessment of *B. braunii* transcriptomic date (Niehaus et al., 2011). Functional analysis of SSL-1 demonstrated it to catalyze PSPP formation from two FPPs in a NADPH-dependent manner. SSL-2 itself catalyzed the conversion of FPP to 90% bisfarnesyl ether and 10% squalene. But unexpectedly, when SSL-2 were co-expressed with SSL-1 or provided with PSPP, high levels of squalene biosynthesis and accumulation were observed. Similarly, SSL-3 was unable to utilized FPP as a substrate, but when co-expressed with SSL-1 or provided with PSPP, high-levels of botryococcene accumulated (Figure 1.3).

To improve the efficiency of botryococcene biosynthesis, different configurations of SSL-1 and SSL-3 genes was tested. Fusion of SSL-1 and SSL-3 with a triplet repeat linker of GGSG improved production efficiency 2 to 3-fold. Further enhancement was found when the SSL-1 and SSL-3 enzymes were appended with the carboxy-terminal amino acids domains of B. braunii squalene synthase to C-terminal of SSL-1 and SSL-3 respectively (SSL1M+3M) (Figure 1.4). Further experiment shows the best configuration is fusion enzyme appending with only one membrane associated domain at its carboxyl-terminal (SSL1+3M) (data not shown). The function of these terminal amino acids would serve to direct SSL-1 and 3 to the yeast's endo-membrane system, which might give rise to the enzymes greater access to endogenous FPP pools (Niehaus et al., 2011)

The triterpene oils content of *B. braunii* race B is primarily of botryococcenes, of which less than 1% is in the non-methylated C30 form, while the majority is in methylated forms (C32-C34). The proportions of different methylated triterpenes varied by algae strains algae and their growth conditions (Metzger, 1985; Metzger et al., 1988). Compared with non-methylated forms, methylated botryococcene and squalene are much more efficiently hydrocracked to fuel components processes (Figure 1.3). Hence, characterization of the specific mechanism for how these triterpenes are methylated could provide significant mechanistic insights to the catalytic capacity of these methyltransferases.

Triterpene methyltransferase TMT-1, TMT-2 and TMT-3 were identified by blast searches of the *B. braunii* 454 transcriptomic dataset using the C24 sterol methyl transferase as the search query (Niehaus et al., 2012). TMT-1 and TMT-2 were subsequently shown to exhibit strong methylating activity squalene. Co-expression of triterpene methyl-transferase-1 (TMT-1) with *B. braunii* squalene synthase (BSS) in yeast resulted in 63% of the total squalene being methylated, of which 43% accumulated as dimethyl-squalene. When TMT-2 was co-expressed with BSS, 40% of the squalene was methylated, of which 31% accumulating as monomethyl-squalene and the rest as dimethylated squalene. In contrast, TMT-3 showed much lower activity for methylating squalene, but readily methylated botryococcene. When TMT-3 co-expression with BSS or SSL1+3m (bytrococcene synthase), 18% of the total squalene was methylated, whereas 33% of the botrycoccene was methylated, of which more than half was dimethylated (Figure 1.4). The function of these three enzymes were corroborated by in vitro enzyme assay using ^{3}H-SAM and botryococcene or squalene as the substrates (Niehaus et al., 2012).

Through NMR analysis of these unique methylated compounds, methylation on mono-methyl squalene was only found at C-3, and methylation on dimethyl squalene was at C-3 and C-22. In contrast, mono-methylated botryococcene produced by TMT-3 has methylation at two positions, C-20 or C-3. The methylation on dimethylated-botryococcenes occurs at both C-3 and C-20. Although botryococcene accumulates in *B. braunii* predominantly in a tetra-methylated form, further work will be needed to find the enzymes responsible for the specific methylation at C-33 and C-34 (Niehaus et al., 2012).

1.7 Methylation, methyltransferase and SAM in plants

Methylation is one of the most common reactions occurring in all organisms. It is involved in wide range of biological processes, including cell signaling, regulation of gene expression and protein function, and biosynthesis and metabolism of primary and secondary metabolites. Especially DNA and protein methylation have been extensively studied for their critical roles in epigenetic inheritance (Martin and McMillan, 2002; Schubert et al., 2003; Kozbial and Mushegian, 2005).

The majority of methylation reactions are catalyzed by methyltransferases, which are a group of enzymes that share a diverse and limited sequence similarity, but broadly conserved for AdoMet binding domain (Kagan and Clarke, 1994). These enzymes can be divided into several classes based on their substrate specificity (*C-, O-, N-, S*-, or halide methyltransferases) or their structural similarity (Noel et al., 2003; Schubert et al., 2003; Zubieta et al., 2003; Liscombe et al., 2010). In plants, these enzymes are responsible for methylation of proteins, nucleic acids, lipids cell wall polymers, as well as secondary metabolites, which plays important roles for plant growth and development (Huang et al., 2012; Sauter et al., 2013).

Most of these methyltransferases utilize *S*-Adenosylmethionine (SAM) as the methyl donor (Liscombe et al., 2010; Scheer et al., 2011), one of the most abundant co-factors in plant metabolism (Fontecave et al., 2004; Sauter et al., 2013). SAM is synthesized exclusively in the cytosol by using methionine (Met) as the building block (Ravanel et al., 2004), accounting for 80% of Met metabolism (Giovanelli et al., 1985). Although 90% of SAM will be used in vivo as the methyl donor (Giovanelli et al., 1985), the remaining SAM can also serve as primary precursors for the biosynthesis of the plant hormone ethylene (Wang et al., 2002a), the growth stimulating polyamines (Takahashi et al., 2003a; Kusano et al., 2008), the iron-chelating nicotinamine (Takahashi et al., 2003b) and cyclopropane fatty acids (Figure 1.5)(Bao et al., 2002; Bao et al., 2003).

SAM also plays a critical role in the methylation and metabolism occurring in the chloroplast, where SAM was strictly imported from the cytosol by specific SAM/SAH exchangers exist on the outer membranes of plastids (Ravanel et al., 2004; Bouvier et al., 2006). The imported SAM is utilized for in the biogenesis of aspartate-derived amino acids, such as Lys (lysine), Ile (isoleucine) and Thr (threonine) (Curien et al., 1998; Jander and Joshi, 2009; Sauter et al., 2013). SAM also serves as the methyl donor for methylation of primary, and secondary metabolites in the chloroplast, such as plastid DNA (Nishiyama et al., 2002; Ahlert et al., 2009a) and proteins (Houtz et al., 1989; Niemi et al., 1990; Ying et al., 1999; Trievel et al., 2003; Alban et al., 2014); or specific metabolites prenyllipids (e.g.,chlorophylls, plastoquinone, tocopherol, and phylloquinone) (Bouvier et al., 2005; DellaPenna, 2005; Bouvier et al., 2006) and the diterpene antioxidant carnosic acid (Munné-Bosch and Alegre, 2001) (Figure 1.5).

Figures of chapter 1

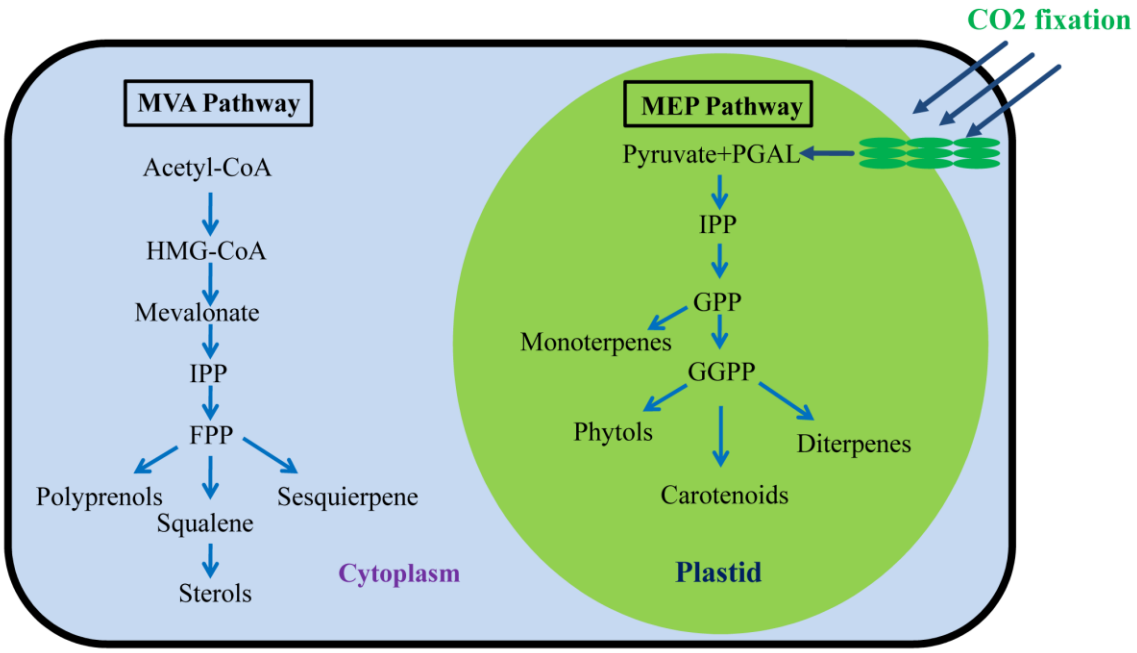


Figure 1.1 Outline of the two terpene biosynthetic pathways that operate generally in plants, the mevalonate (MVA) pathway in the cytoplasm and the methyl erythritol phosphate (MEP) pathway in the plastidic compartment. MVA, Mavalonate; MEP, Methylerythritol phosphate; HMG-CoA, 3-hydroxy-3-methylglutaryl-coenzyme A; IPP Isopentenyl diphosphate; DMAPP, Dimethylallyl diphosphate; FPP, Farnesyl diphosphate; PGAL, Glyceraldehyde 3-phosphate; GPP, Geranyl pyrophosphate; GGPP, Geranylgeranyl pyrophosphate.

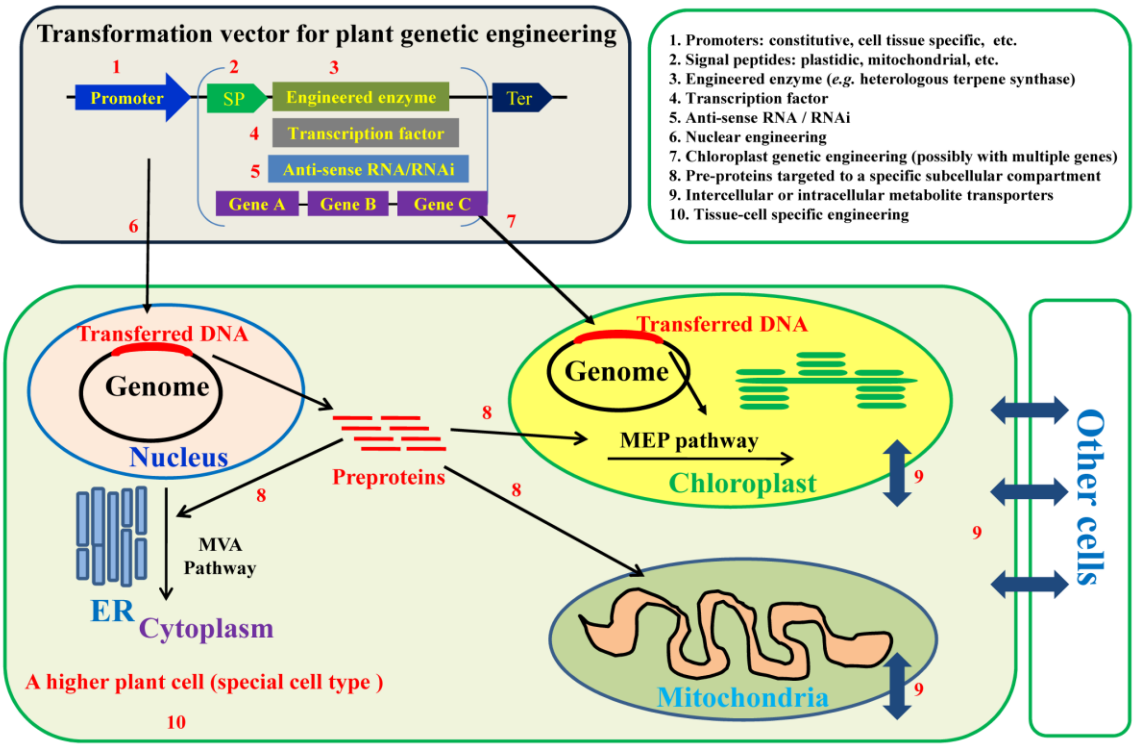


Figure 1.2 A schematic blueprint illustrating important variables and approaches to metabolic engineering in plants. This figure shows a typical transgene design that can capitalize upon various permutations to yield the desired flux of carbon to the biosynthesis of unique end-products. The process(es) controlled by each arrow is defined within the inset figure legend.

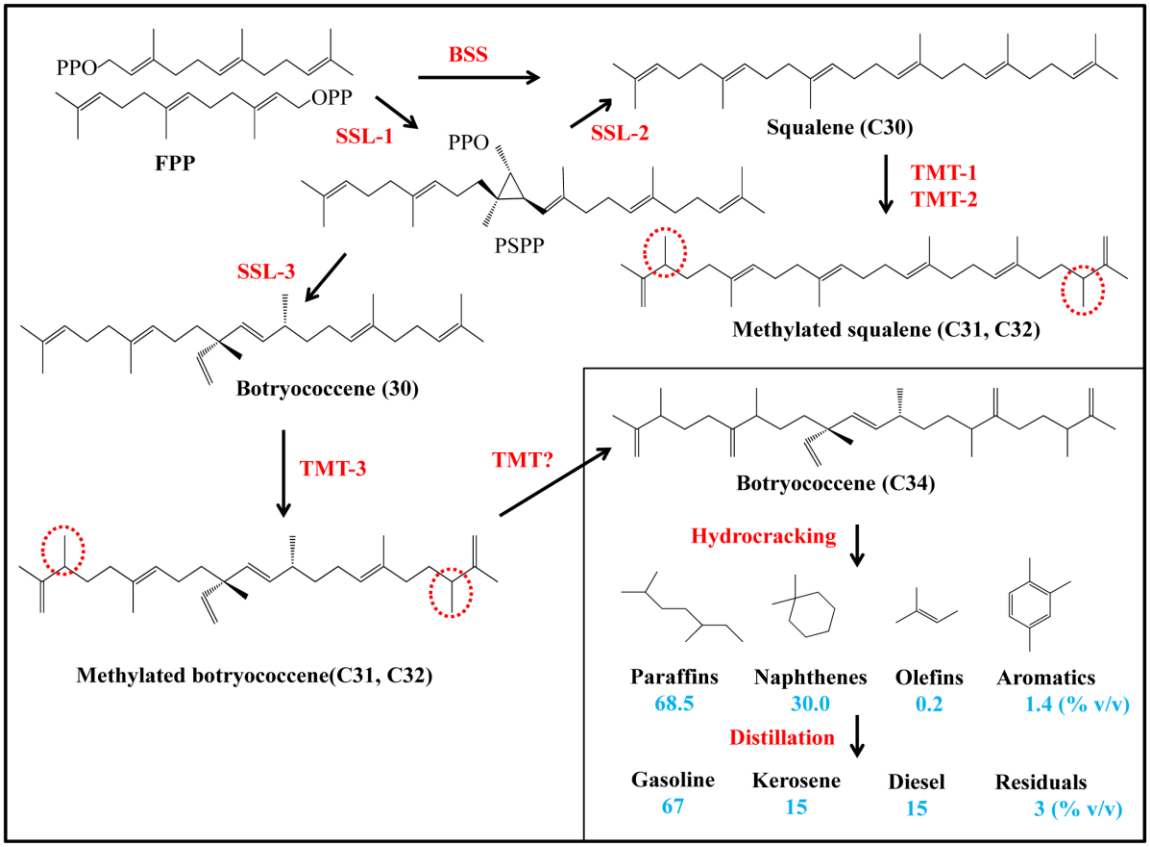


Figure 1.3 A depiction of the catalytic roles of the squalene synthase-like enzymes in *Botryococcus braunii* race B and their putative contributions to the triterpene constituents (Niehaus et al., 2011). SSL-1 converts two farnesyl diphosphate molecules (FPP) to pre-squalene diphosphate (PSPP), which is converted to either squalene or botryococcene by SSL-2 or SSL-3, respectively. TMT-1 and TMT-2 catalyze the transfer the methyl donor group from AdoMet (SAM) to squalene to form mono- or dimethylated squalene, whereas TMT-3 acts on botryococcene to form mono- or dimethylated botryococcene. TMT-1 and TMT-2 can transfer a methyl group from SAM to squalene form mono- or dimethylated squalene, while TMT-3 favors on botryococcene to form mono- or dimethylated-botryococcene (Niehaus et al., 2012). Catalytic hydrocracking of tetramethylated botryococcene isolated from *B. braunii* race B yields petroleum distillate-like products that can be used directly for industrial chemical manufacturing, or can be distilled in high yields to give all classes of combustible fuels, including gasoline (67%), aviation fuels (15%) and diesel (15%) (Niehaus et al., 2011).

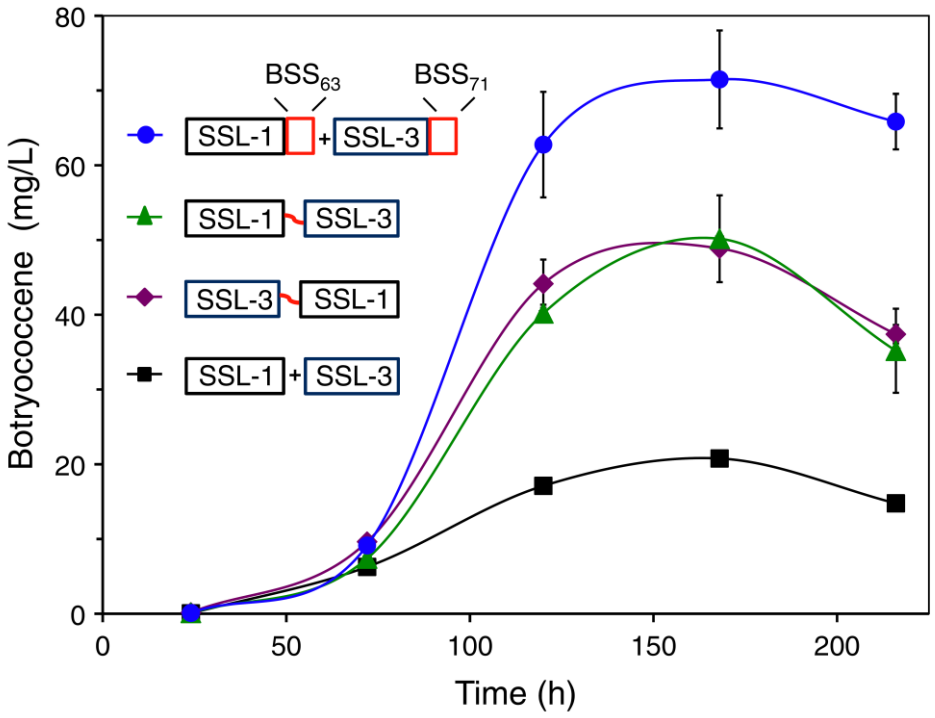


Figure 1.4 Comparison of botryococcene production in yeast engineered with different configurations of SSL-1 and SSL-3 (Niehaus et al., 2011). Yeast line TN7 was engineered with the SSL-1 and SSL-3 genes on separate plasmids (squares); SSL-1 fused to SSL-3 via a triplet repeat of GGSG (triangles); the carboxyl terminus of *Botryococcus* squalene synthase appended to C-terminal of the SSL-1 and SSL-3 enzymes, respectively.

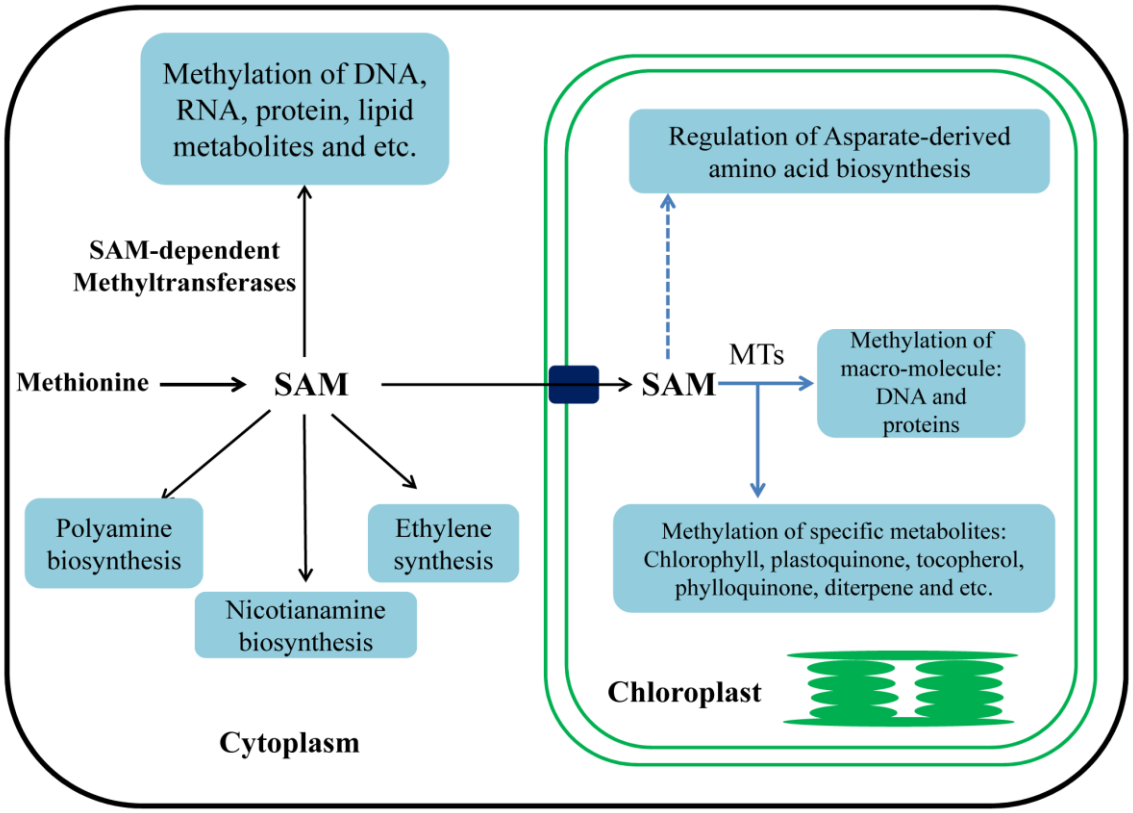


Figure 1.5 Model of subcellular compartmentation of SAM metabolism and its multiple roles in plant cells. SAM is exclusively synthesized from methionine in the cytosol, where it serves as a building block for biosynthesis of polyamines, ethylene, nicotianamne and is essential as a methyl donor for the methylation of many macromolecules catalyzed by methyltrasferases. SAM biosynthesized in the cytoplasm is also imported into the chloroplast by specific transporters that located on the inner membrane of the plastid envelope, and utilized in the biosynthesis of a variety of metabolites, amino acids and macromolecules as illustrated.

Chapter 2: Engineering squalene metabolism in tobacco

2.1 Summary

Terpenes comprise a distinct class of natural products that serve a diverse range of physiological functions, provide for interactions between plants and their environment, and represent a resource for many kinds of practical applications. To better appreciate the importance of terpenes to overall growth and development, and to create a production capacity for specific terpenes of industrial interest, we have pioneered the development of strategies for diverting carbon flow from the native terpene biosynthetic pathways operating in the cytosol and plastid compartments of tobacco for the generation of specific classes of terpenes. In the current work, we demonstrate how difficult it is to divert the 5-carbon intermediates DMAPP and IPP from the mevalonate pathway operating in the cytoplasm for triterpene biosynthesis, yet diversion of the same intermediates from the methylerythritol phosphate pathway operating in the plastid compartment leads to the accumulation of very high levels of the triterpene squalene. This was assessed by the co-expression of an avian farnesyl diphosphate synthase and yeast squalene synthase genes targeting metabolism in the cytoplasm or chloroplast. We also evaluated the possibility of directing this metabolism to the secretory trichomes of tobacco by comparing the effects of trichome-specific gene promoters to strong, constitutive viral promoters. Surprisingly, when transgene expression was directed to trichomes, high-level squalene accumulation was observed, but overall plant growth and physiology were reduced up to 80% of the non-transgenic controls. Our results support the notion that the biosynthesis of a desired terpene can be dramatically improved by directing that metabolism to a non-native cellular compartment, thus avoiding regulatory mechanisms that might attenuate carbon flux within an engineered pathway.

2.2 Introduction

Terpenes are a structurally diverse class of compounds in plants that contribute to an equally diverse array of physiological and ecological functions. The structural diversity is most readily recognized in the classification of terpene families with repeating units of 5-carbon building blocks, like sesquiterpenes with 15 carbons and triterpenes derived from a 30-carbon scaffold. Terpene chemical diversity, however, extends much beyond polymer size or linear versus cyclized forms to the substituent decorations like

hydroxylation, acylation, aroylation, methylation and glycosylation. Given such chemical richness, it isn't too surprising how recent efforts have uncovered unique roles for terpenes in general growth and developmental processes. For instance, the essentiality of brassinolides (30 carbon triterpenes derivatives) for overall plant growth (Clouse, 2011) and strigolactones (15 carbon sesquiterpenes arising from the breakdown of 40 carbon carotenoids) for the control of axillary bud dormancy and root architecture (Kohlen et al., 2011) are two such examples. Our appreciation for the specialized roles of terpenes in mediating ecological interactions between plants with other plants (Kegge and Pierik, 2010), insects (Keeling and Bohlmann, 2006) and microbes (Huffaker et al., 2011) has also grown in parallel with our understanding for the structural diversity of plant biosynthesized terpenes. Plant derived terpenes have also played a large role in various industrial applications ranging from flavors and fragrances (Schwab et al., 2008) to medicinals (Shelar and Shirote 2011) to the more recent attention to their utility for biofuels (Niehaus et al., 2011).

With the increased recognition of terpene contributions to physiological functions and evolving industrial uses, a parallel effort has been to manipulate the biosynthesis and accumulation of these compounds in plants for a variety of reasons. One, genetic and molecular genetic technologies to abolish or ectopically produce specific terpenes have been used to identify the biochemical and physiological function of genes, thus providing a gene annotation capability (Tholl et al., 2005). This has also been important for testing our understanding for biochemical processes in general and understanding the genetic and biochemical components associated with terpene biosynthetic enzymes (Mandel et al., 1996). Second, generating transgenic plants accumulating altered amounts of a specific terpene or suite of terpenes can be evaluated for their health promoting properties (Sawai and Saito, 2011) or could provide new means for the sustainable production of high-value chemicals for industrial uses (Krings and Berger, 2010).

Engineering terpene metabolism in microbial hosts has advanced significantly in the recent past with much of the emphasis on providing a higher yield and recovery of high valued terpenes. Much of the early success took advantage of the innate biosynthetic machinery in *E. coli*, the methylerythritol phosphate or MEP pathway, and introduced a limited number of carotenoid biosynthetic genes to yield visibly distinct lines (Schmidt-Dannert et al., 2000). Additional efforts to up-regulate putative rate-limiting steps early in the MEP have also improved carotenoid yields (Kim and Keasling, 2001). Significant

gains in the production of sesquiterpenes was dramatically realized when the complement of the eukaryotic mevalonate, MVA, pathway from yeast was mobilized into *E. coli* (Martin et al., 2003). Yields improved from µg/l to mg/l for the sesquiterpene hydrocarbon amorphadiene upon heterologous expression of these yeast genes in E. coli. Complementary to these prokaryotic studies, investigators also desired production platforms for more highly decorated terpenes and especially hydroxylated forms. Because the eukaryotic enzymes for terpene hydroxylation and their associated cofactors like cytochrome P450 reductases require internal membrane systems unique to eukaryotic cells, development of terpene production in yeast has also been advanced. In contrast to *E. coli*, yeast only possess the MVA pathway, which directs a significant amount of carbon down this cytosolic pathway to ergosterol biosynthesis, the dominant sterol required for normal growth of yeast. Introducing additional mutations in yeast allowing them to utilize exogenous ergosterol under aerobic conditions frees up intermediates that can be redirected in desired ways. Ro et al. (2006) and Takahashi et al., (2007), for instance, demonstrated that such a strategy allowed for the development of yeast strains producing greater than 50 mg/l of oxygenated sesquiterpenes.

The manipulation of terpene metabolism in plants as a means for investigating key biochemical processes as well as for developing plants as production platforms for high value terpenes has also been advanced significantly. Notable examples include molecular breeding efforts to enhance carotenoid (Harjes et al., 2008) and artemisinin (Graham et al., 2010) metabolism in maize and Artemisia annua, respectively. Other investigators have focused on augmenting terpene metabolism by the ectopic expression of terpene biosynthetic activities in different cellular compartments. Monoterpene biosynthesis naturally occurs in the chloroplast compartment of plant cells, but several investigations have documented that over-expression of a single monoterpene synthase to either the chloroplast or cytoplasm compartment resulted in the accumulation of new monoterpenes and in some cases novel derivatives like glyco-conjugates (Lewinsohn et al., 2001; Lücker et al., 2001; Aharoni et al., 2003; Ohara et al., 2003). Re-directing the biosynthesis of sesquiterpene metabolism to the mitochondria and plastid compartment has had an equal, if not greater, impact on overall terpene metabolism. Kappers et al. (2005) demonstrated the ability of plants to synthesize unusual sesquiterpenes in the mitochondrial compartment led to plants able to attract predatory insects as a biocontrol mechanism. Likewise, Wu et al. (2006) demonstrated

that engineering a more robust sesquiterpene metabolism normally associated with the cytoplasm to the chloroplast compartment over-rode any innate regulatory mechanisms and yielded robust production of industrially valued sesquiterpenes. More recently, Kumar et al. (2012) extended such ectopic engineering strategies by inserting genes coding for the normal cytoplasmic MVA pathway into the chloroplast genome to affect high-level expression of the enzymes leading up to mevalonate. These transgenic plants lines accumulated mevalonate while no such accumulation in control plants was evident, and exhibited a 2-fold increase in their main sterol level and a 10-fold increase in squalene, but only a 20% increase in ß-carotene content. The results of Kumar et al. (2012) suggested that a mevalonate biosynthetic pathway engineered into the chloroplast was able to complement and augment overall terpene biosynthetic processes occurring both within and outside the chloroplast compartment.

In the present effort, our objective was to determine if our previous strategy for engineering high level sesquiterpene accumulation was applicable to larger terpenes, and in particular to the triterpene class of compounds. We also aimed to evaluate the possibility of targeting this metabolism to trichomes such that the biosynthesis of a target molecule might be secreted as per the suggestions of Wang et al. (2004) and Ennajdaoui et al. (2010).

2.3 Results

2.3.1 Experimental approach

Terpene metabolism in plant cells is divided between the mevalonate (MVA) pathway operating in the cytoplasm and the methylerythritol phosphate (MEP) pathway occurring in the chloroplast (Figure 2.1). Interestingly, a convenient division of labor between these two pathways have been established with the MVA pathway largely dedicated to sesquiterpene, triterpenes and polyprenol biosynthesis in association with the ER endomembrane system, while the MEP pathway is responsible for monoterpenes, diterpenes, carotenoids (tetraterpenes) and long-chain phytol biosynthesis occurring in the chloroplast stroma. Given this sort of organizational complexity, we reasoned it would be best to evaluate several different strategies for engineering triterpene metabolism and specifically squalene accumulation.

The first approach was to compare squalene accumulation in transgenic plants expressing a heterologous squalene synthase directed to the cytosolic compartment versus the chloroplast compartment (Figure 2.2). The squalene synthase gene used in all these constructs is from yeast and has a 3' truncation of the DNA sequence coding for a carboxy-terminal, membrane-spanning domain, hence yielding a functionally soluble squalene synthase enzyme activity (ySS) (Zhang et al., 1993). This deletion was important to assure catalytic activity of the squalene synthase vectorially imported into the chloroplast compartment via an amino terminal targeting signal sequence (tp) from the Rubisco small subunit gene from *Arabidopsis* (Lee et al., 2006). The yeast squalene synthase gene was also chosen because it was assumed to be devoid of sequences subject to transcriptional to post-translational regulation that another plant squalene synthase gene/enzyme might be. Expression of these initial constructs were directed by trichome-specific promoters as described by (Ennajdaoui et al., 2010) and (Wang et al., 2002a) to potentially provide for the secretion of the trichome synthesized squalene onto the leaf surface, as well as the fairly conventional constitutive promoters from caulimoviruses (Benfey et al., 1990; Verdaguer et al., 1998).

These first constructs assumed that an introduced squalene synthase could compete for any available FPP in the cytoplasm, or FPP that could arise in the chloroplast compartment as an intermediate released from the MEP pathway or imported from the cytoplasm. Cytosolic FPP levels are, however, generally low, and FPP is thought to serve a regulatory role in controlling carbon flux into the MVA pathway (Closa et al., 2010). There is also little evidence, if any, for all trans-FPP being formed in chloroplasts (Sallaud et al., 2009). The second construct iterations thus included a chicken gene encoding for a well-characterized farnesyl diphosphate synthase (FPS) (Tarshis et al., 1994) in addition to the yeast squalene synthase (Figure 2.2). Expression of these constructs varied by using either strong constitutive promoters or trichome-specific promoters, plus/minus amino terminal sequences targeting the SS and FPS enzymes to the cytoplasm or chloroplast compartments.

2.3.2 Screening of the T0 and T1 transgenic lines

The constructs of Figure 2.2 were then used to generate approximately 20 independent transgenic lines per construct. The particular cultivar of tobacco, TI 1068, used for these experiments is an accession line identified for its high density of secretory trichomes

(Nielsen et al. 1982). The various transgenic lines were first evaluated for squalene accumulation while still in their final stages of plantlet develop in tissue cultures, then re-screened multiple times as the T0 generation was propagated in the greenhouse (Table 2.1).

Plants expressing only the squalene synthase gene tended to accumulate only marginally higher levels of squalene than observed in the control, non-transgenic lines, approximately 2-fold. Nonetheless, whether the squalene synthase was targeted to the cytoplasm or chloroplast, 1 to 2 transgenic plants within each group accumulated squalene levels 4 to 5 times the maximum level observed for the control plants. Surprisingly, plants engineered with both SQS and FPS targeted to the cytoplasm and directed by either strong constitutive or trichome specific promoters also did not accumulate squalene much beyond those lines engineered with the SQS gene by itself.

In contrast, plants engineered with SQS and FPS targeted to the chloroplast compartment demonstrated an overall average accumulation of squalene 20 to 30-fold greater than the non-transgenic control lines. More impressive, individual lines accumulated 200 to 600 µg of squalene per g fresh weight, 27 to 90 times greater than the non-transgenic controls. Nonetheless, abnormal growth characteristics were observed for several of these lines. For the constitutive expressed forms of SQS and FPS, 30% of the lines exhibited a noticeable dwarfing phenotype, whereas greater than 85% of the regenerated lines with the trichome specific expression cassette demonstrated some degree of dwarfing (35%), chlorosis (17%), or some combination of both (30%)(Figure 2.3).

Similar trends in squalene accumulation were noted for individual plants examined in the T1 generation (Table 2.2). The T1 screen also attempted to correlate squalene accumulation with plant development, hence samples representing young, developing and mature stages of leaf developmental were evaluated. Plant lines engineered for SQS targeted to the cytoplasm exhibited little or no difference in their ability to accumulate squalene relative the non-transgenic control. Cytosolic targeting of SQS and FPS, regardless of the expression promoters used, also did not accumulate significantly more squalene than found in the control plant. Transgenic plants with SQS and FPS targeted to the chloroplast compartment, however, demonstrated a very significant accumulation of squalene in a developmental dependent manner. Levels of squalene

accumulation were 300 to 1,000 fold greater than those levels measured in control plant lines.

While squalene is a relatively stable compound, it could be subject to both secondary metabolisms occurring in planta, as well as environmental induced changes like oxidation. To examine the integrity of the squalene accumulating by plants grown in the greenhouse, 60 g of leaves from a plant constitutively expressing plastid targeted SQS and FPS were extracted with hexane, the putative squalene compound purified by successive rounds of silica chromatography, and the isolated compound then analyzed by GC-MS and NMR analyses. When evaluated by GC-MS, the squalene purified from tobacco leaves exhibited an identical retention time and mass spectrum to an authentic squalene standard, as did its ^{1}H-NMR and ^{13}C-NMR spectra (Figure 2.4, figure 2.5).

Transgenic plantlets generated with the indicated construct (expression promoter, gene(s) and intracellular targeting information) were propagated under sterile conditions until they established root systems and were ready for growth in the greenhouse. The greenhouse grown plants were screened several independent times for their squalene content. In the data shown, plants were from 40 to 60 cm tall and the first fully expanded leaf was sampled. Data represents the average from all the independent plant lines sampled twice, as well as the minimum (min) and maximum (max) observed. Plants were scored as dwarf if their height was 25% less than their sibling plants. Plants were scored as chlorotic if there was obvious yellowing within 3 or more leaves.

Individual, second generation (T1) plants were propagated in the greenhouse and identified on the basis of preliminary chemical profiling screens. On average, young leaves were 5-7.5 cm in length, developing leaves were 10 -15 cm in length, and mature leaves were greater than 20 cm long. Distinct phenotypes are noted for the plants harboring the constructs for trichome specifically expression of SQS and FPS targeted to the plastid compartment (see Figure 2.3 for examples).

2.3.3 Assessment of field grown plants

To gain a better appreciation for the robustness of the squalene accumulation trait and its impact on overall growth performance, segregating populations of transgenic lines expressing plastid targeted SQS and FPS enzymes under the direction of the trichome-

specific promoters (line #32) or the constitutive promoters (line #42) were propagated in outdoor field conditions. These lines were chosen because the squalene levels determined for these lines during the initial T0 generation screens appeared more typical for this class of transgenic plants, rather than representing an extreme. In the T0 greenhouse screens, the line expressing SQS and FPS targeted to the plastid compartment with trichome specific promoters (line 32) accumulated 192 µg squalene/g fresh weight, whereas 150 µg squalene/g fresh weight was recorded for the constitutive expressing line 42.

T1 seeds for both lines were germinated without any selection for the transgenes, hence representing a segregating population, and grown for approximately 6 weeks in a greenhouse before transplanting the plantlets in the field. Two replicate rows of each line were grown with standard plant and row spacing, along with independent rows of the non-transgenic parental line. Plantlets were randomly selected from the greenhouse propagation trays for planting, watered once to twice a week for a couple of weeks to support their initial establishment, then allowed to grow without any additional treatments (*i.e.* no fertilizer or pesticide treatments) for a 60-day growing period. Agronomic performance characteristics and chemical profiles were measured twice for each plant and data for plants accumulating squalene (presumably homozygous and heterozygous for the transgenes) as well as those not accumulating squalene were clustered for quantitative comparisons. The data in Table 2.3 were obtained from plants at the end of field growth cycle with photosynthetic measurements and squalene accumulation determined for the uppermost, fully expanded leaf. For line 32, 6 plants accumulated squalene and 18 plants did not. Of the 22 plants of line 42 examined, 14 plants accumulated squalene.

Of the non-squalene accumulating plants segregating out of lines 32 and 42, these plants performed directly comparable to the wild type check plants with regards to any of the agronomic indicators (height, biomass accumulation, leaf area) or photosynthetic measurements (CO_2 fixation rates, transpiration and internal CO_2 levels). The only modest impact on performance was for their total leaf area measurements with the non-squalene accumulators within line 42 having 84% and those within line 32 having only 72% of that for the control plants.

In contrast, for the squalene accumulating plants of line 32 (SQS and FPS targeted to the chloroplast and expression directed by the trichome specific promoters), their overall agronomic performance was only 20, 21 and 37% that of their non-squalene accumulating siblings for biomass accumulation, leaf area and height, respectively. The squalene accumulating plants within line 32 also only exhibited about 32% the photosynthetic rate of their non-squalene accumulating siblings. Growth was also impacted for the squalene accumulating plants within transgenic line 42, though the average squalene accumulation within these plants exceed that observed for line 32 (112 versus 80 µg/g fresh weight). Leaf area, total biomass accumulation and plant height of the squalene accumulators were 66, 54 and 68% of that for their non-squalene accumulating siblings. However, their photosynthesis rates, stomatal conductance and their internal CO_2 levels were almost identical to their non-squalene accumulating siblings and the wild type control plants.

Given the relative normal growth appearance of the squalene accumulating plants within line 42 (they appear to grow slower and hence appear smaller than the control check plants), we also examined the squalene accumulation in leaves of different developmental stages in several of these plants. The importance of this measurement became obvious when we noted that the squalene level of these plants after about 4 weeks in the field was 26.5 µg/g fresh weight and over 230 µg/g fresh weight after an additional 4 more weeks. As shown in Table 2.4, squalene levels in the very young leaf tissues was quite low, almost below detection limits for the just emerging leaves, but exhibited a significant increase in accumulation with leaf maturation. In lower, more senescent leaves, the amount of extractable squalene was significantly lower than in the mature leaves. Also evident in the data of Table 2.4 is the variability between individual plants and absolute leaf position, some of which might be related to the zygosity of the particular plant as well as difficulty in attaining an absolute standardization of leaf development between plants.

Segregating (T1 generation) seed for transgenic lines targeting SQS and FPS to the chloroplast under the direction of trichome specific promoters (line 32) or constitutive promoters (line 42) were grown under greenhouse conditions for approximately 6 weeks prior to transplanting them to the field. Plants were chosen randomly for field planting and the plants grown for a total of 60 days, which is approximately two-thirds of a full growth cycle. The plants were screened several times throughout the growing season,

but the data shown is for the final data collection at the end of the field season. For squalene determinations, two leaf discs of 2 cm diameter collected from the first fully expanded leaf from the top of each plant were extracted with organic solvent and their squalene content determined by GC-MS. For line 32, 6 plants accumulated significant squalene (denoted as +), while 18 others did not (-). For line 42, 14 plants accumulated squalene (+) and 8 did not (-). Six control plants were evaluated. No squalene detected below the detection limit of ~0.5 µg/g fw are denoted as such (nd). Measurements of photosynthetic gas exchanges were conducted between 10AM and 12PM on a cloudless day with light intensity of 1,500 µmol/m^2•sec with a LI-COR 6400 portable instrument

Leaves at various positions of two plants within line 42 were sampled for their squalene levels. Plants were grown for 60 days in the field and appeared visually comparable to one another. Not detected, nd, refers to levels below detection limit of 0.5 µg/g fw.

2.4 Discussion

The current work extends earlier efforts to engineer terpene metabolism in plants in several significant ways. Investigators, including ourselves, have successfully engineered relatively robust monoterpene (Lewinsohn et al., 2001; Lücker et al., 2001; Aharoni et al., 2003; Ohara et al., 2003) and sesquiterpene (Kappers et al., 2005; Wu et al., 2006) biosynthesis in transgenic plants, and more modest manipulations in the level of diterpenes (Besumbes et al., 2004) and triterpenes (Seo et al., 2005; Lee et al., 2006; Kumar et al., 2012). Interestingly, many of the early efforts were focused on introducing enzymes to effectively compete for substrates or intermediates in distinct cellular compartments where this metabolism occurs naturally, or were attempts to overcome prospective rate-limiting steps by over-expressing a gene coding for the suspected limiting enzyme. We, and other investigators, have demonstrated that a more successful strategy is to divert carbon flux at earlier upstream intermediates to build particular terpene compounds in compartments were this metabolism does not normally occur. That was indeed the case here in the successfully engineering of triterpene metabolism. For example, when SQS was targeted to the cytoplasm and thus potentially accessing FPP synthesized by the MVA pathway, only a 2-fold increase in squalene levels relative to the wild type controls was observed. When the SQS was targeted to the chloroplast where FPP biosynthesis is not known to occur, the levels of squalene observed were again, on average, 2-fold greater than the control, non-transgenic plants. There were, of

course, a few exceptions where squalene accumulation in transgenic lines having SQS targeted to the cytoplasm or chloroplast were more than 4-fold greater than the levels in the control plants, but these were single transgenic events and not general observations.

Explanations for why such a narrow window in the enhancement of squalene accumulation is observed when attempting to directly divert the normally produced FPP include possible channeling of this intermediate within metabolons or stringent regulation imposed upon carbon flux down these pathways by metabolite feedback regulatory networks (Gardner and Hampton, 1999; Masferrer et al., 2002; Manzano et al., 2004; Muñoz-Bertomeu et al., 2007; Sawai and Saito, 2011), thus limiting the availability of FPP. More difficult to understand is how chloroplast targeted SQS by itself would even come into contact with FPP because all trans-FPP is not known to be synthesized in the chloroplast. Perhaps some cytoplasmic produced FPP can diffuse or be transported into the chloroplast, or some of the chloroplast targeted SQS gains access to the cytosolic FPP during its movement from its site of synthesis in the cytosol to the chloroplast. Equally possible, at least for the plastid targeted SQS, is that some of the targeted enzyme may not actually be making its way to the chloroplast compartment and is simply diverting FPP formed in the cytoplasmic compartment. But then, one would expect cytosolic targeted SQS to have an equal effect and it doesn't. Nonetheless, when FPS and SQS are co-expressed and targeted to the chloroplast compartment, the FPS appears able to divert significant DMAPP and IPP (the substrates for FPS) from the MEP pathway for the biosynthesis of FPP, which in turn is available to SQS for the novel and very robust biosynthesis of squalene in the chloroplast compartment.

Interestingly, the same is not true for FPS and SQS co-expressed and targeted to the cytosolic compartment where the MVA pathway operates. This result suggests that DMAPP and IPP are not as readily available in the cytoplasmic compartment, or that metabolic flux down the MVA pathway is much more regulated by unknown factors than is observed for these metabolites in the chloroplast. Similar conclusions were reached by Wu et al. (2006) when FPS was co-expressed with sesquiterpene synthases targeted to either the chloroplast or the cytoplasmic compartments.

An equally surprising observation was that targeted expression of the FPS and SQS using trichome-specific promoters resulted in physiologically impaired plants, while use of strong, constitutive promoters were much less so. The first indication of this effect was

evident in the T0 generation plants. Greater than 80% of the transgenic lines utilizing the trichome-specific promoters were either dwarf, chlorotic or a combination of both. In comparison, only 23% of the T0 transgenic lines using the constitutive viral promoters showed some signs of dwarfism, but no chlorosis. These differences were even more accentuated when the plants were grown in field conditions. While the trichome-specific expressing plants accumulated relatively modest levels of squalene, overall growth and physiological functioning were reduced 60 to 80%. The impact of constitutively expressing FPS and SQS targeted to the chloroplast was a growth reduction of 30 to 40% without any adverse effects on photosynthetic parameters, even though these plants accumulated greater amounts of squalene than the trichome-specific expressing plant line. Why the trichome-specific expressing transgenic plants are more impacted than the constitutively expressing plants is currently unknown. It may be that trichome-directed expression simply distorts biochemical processes in such a manner that trichome derived cues or signals alter metabolism occurring elsewhere in the plant. Or, the enhancer elements used to improve trichome expression may cause ectopic expression in meristematic or progenitor cells distorting their contributions to normal physiological growth features. We are not aware of any other reports of similar growth distortion when attempting to engineer trichome metabolism, although the number of such engineering efforts are rather limited at this time. Additional screening of more independent transgenic lines created with these and other trichome-specific constructs, and unbiased metabolomics profiling might help resolve these issues.

Finally, comparison of the developmental squalene accumulation profiles by the field grown plants and separate transgenic lines grown under greenhouse conditions suggests there is certainly variability by leaf position, which is expected. If the novel squalene biosynthetic machinery is produced constitutively over developmental time, then squalene accumulation should continue as leaves develop and expand over time. Importantly, the levels of squalene accumulating in the field grown plants harboring the constitutively expressing FPS and SQS targeted to the chloroplast were equal to or higher than those levels measured in greenhouse grown plants, an indication of the robustness of this engineering approach for plants grown under a variety of conditions and the utility of such a platform for high value tritepene production.

2.5 Materials and methods

2.5.1 Expression vector construction and plant transformation

Construct design and assembly were based on the work previously described by Wu et al. (2006) using standard molecular methodologies (see Figure 2.2). Gene constructs consisted of a truncated form of the yeast squalene synthase (ySQS) gene (ERG9, GenBank accession NM 001179321) (Zhang et al., 1993) and the avian farnesyl diphosphate synthase (FPS) gene (P08836) (Tarshis et al., 1994). The truncated SQS was created by PCR amplifying the yeast SQS mRNA from its start codon to nucleotide 1260, thus deleting the DNA encoding for the carboxy-terminal 24 amino acids. These carboxy-terminal amino acids are predicted to tether the SQS protein to endomembrane systems in vivo. Hence, deletion of these amino acids creates a functionally soluble enzyme (Zhang et al., 1993). The ySQS and FPS genes were inserted downstream of strong constitutive promoters ((Pca, 35S cauliflower mosaic viral promoter (Benfey et al., 1990); Pcv, cassava vein mosaic viral promoter (Verdaguer et al., 1996)), or trichome-specific promoters ((Pcbt, the cembratriene-ol synthase promoter (Ennajdaoui et al., 2010) or the Pcyp16, cembratriene-ol hydroxylase promoter (Wang et al., 2002b)). The Pcbt and Pcyp16 promoters were further augmented with duplicated CAMV 35S enhancer elements (Benfey et al., 1990) fused to the 5' termini of the promoters. Where indicated, a plastid targeting signal sequence (tp) encoding for the first 58 amino acids of the *Arabidopsis* Rubisco small subunit gene (NM23202) (Lee et al., 2006) was fused onto the 5' end of the respective genes.

The DNA sequences were assembled together using standard molecular biology methods and the various elements verified by DNA sequencing. The expression cassettes were then introduced into pBDON (Wu et al., 2006), a modified pBI101 Ti plasmid vector harboring a hygromycin selection marker and a recombination cloning cassette. In some cases, simple substitution cloning of the desired DNA elements into previously constructed intermediate helper vectors were performed as described by Wu et al. (2006). The engineered Ti plasmid vectors were then introduced into *Agrobacterium tumefaciens* GV3850, and the resulting *Agrobacterium* lines used to genetically engineer *Nicotiana tabacum* (tobacco) TI accession 1068 (Nielsen et al. 1982) as previously described previously by Wu et al. (2006).

Leaf explants were transformed with the respective gene constructs and the resulting calli selected for hygromycin resistance (15 µg/ml) under tissue culture conditions to regenerate plantlets. The selected T0 plantlets were then propagated in the greenhouse and assessed for squalene accumulation by GC-MS analyses.

2.5.2 Plant propagation and field tests

All the T0 plantlets selected for hygromycin resistance were grown in common commercial vermiculite/soil blends in the greenhouse and fertilized weekly with a commercially available high nitrogen, phosphorus, potassium fertilizer. Insect control was performed on as needed basis. The T0 plants were allowed to flower in the greenhouse and the T1 seed collected for subsequent cycles of propagation. Segregation of the hygromycin resistance trait in the T1 seed lines was also evaluated by germinating sterilized seeds on 50 µg/ml hygromycin in T- tissue culture media (4.2 g MS salts (Phytotechnology Laboratories, Overland Park, KS), 0.112 g B5 vitamins (Phytotechnology Laboratories), 30 g sucrose, 8 g agar). For field evaluation, T1 seeds were sown directly in propagation trays in the greenhouse 6 weeks prior to transplanting in the field and were not pre-selected for antibiotic resistance. To determine squalene accumulation, leaf discs of 2 cm diameter were collected from the upper most, fully expanded leaves. Photosynthetic gas exchange measurements of first fully expanded leaves were determined at atmospheric concentrations of CO_2 and a saturating irradiance of 1,500 micromoles photons m^{-2} s^{-1} using a LI-COR 6400 portable photosynthesis system according to Salvucci and Crafts-Brandner (2004). At the time of harvest, plant height was taken, the plants cut at the soil interface, weighed, and all the stripped leaves combined for leaf area determinations. All transgenic work was done in accordance with regulations and permits provided by the APHIS Division of the USDA.

2.5.3 Squalene determinations

One hundred to 500 mg of transgenic leaf material were collected for chemical analyses using a 2 cm diameter cork borer tool to obtain leaf discs of approximately 100 mg each. Each sample was ground in liquid nitrogen, then extracted with 2-3 ml of a hexane:ethyl acetate mixture (v/v 85:15) containing 200 ng of α-cedrene as an external standard for quantification and calculations of recovery. The extracts were carefully concentrated to 500 µl under a nitrogen steam without drying the sample. The concentrated extracts

were then partially purified by passing through a silica column (500 mg, prepared in glass wool plugged glass pipette) and further eluted with 1 ml of the hexane solvent.

After concentration of the combined eluate under a stream of nitrogen, aliquots were injected onto a GC-MS equipped with a Rtx-5 capillary column (30 m X 0.32 mm, 0.25 µm phase thickness) with the following temperature program of 70^{o}C for 1 min, followed by a 4^{o}C per min gradient to 250^{o}C. Mass spectra were recorded at 70 eV, scanning from 35 to 500 atomic mass units, and experimental samples were compared to authentic standards of squalene for verification.

The structure of purified squalene from tobacco was determined by ^{1}H-NMR and ^{13}C-NMR spectral analyses. Squalene was extracted from greenhouse grown plants as described above, except additional purification was afforded by a silica HPLC methodology. Essentially, 60 g leaf material of homozygous line #5 expressing plastid target SQS and FPS under the direction of the constitutive promoters was ground in liquid nitrogen, then extracted with 1.2 l of hexane:ethyl acetate (85:15), the extract concentrated to 5 ml and the extract fractionated on a silica column with 5 ml aliquots of hexane as the eluting solvent. Fractions were monitored by TLC (silica plates, hexane solvent, iodine vapor stain) and GC for the desired triterpene compound. Enriched fractions were pooled, concentrated under nitrogen and the entire sample processed by silica HPLC-PDA using hexane as the eluting solvent (Niehaus et al. 2012). Recovery of 4 mg of purified squalene sample with a 50% yield was obtained. ^{1}H-NMR and ^{13}C-NMR spectra were recorded on a 500 MHz Varian J-NMR spectrometer at 300 K, and chemical shifts were referenced relative to solvent peaks, namely δH 7.24 and δC 77.0 for CDCl3.

Tables and Figures of chapter 2

Table 2.1 Screen of T0 transgenic lines for their squalene content

Construct	# of lines evaluated	Ave (µg/g fw)	Min	Max	Dwarf	Chlorotic	Dwarf & Chlorotic
Wild type (control)	15	3.5	0.5	7.4	0	0	0
Trichome SQS only cytosolic	24	7.6	1.0	34.2	1	0	0
Trichome SQS only plastidic	18	6.7	0.6	30.7	0	0	0
Constitutive SQS+FPS cytosolic	29	5.5	0.8	38.7	5	0	0
Trichome SQS +FPS cytosolic	16	8.1	1	20.1	1	0	0
Constitutive SQS+FPS plastidic	26	63.9	1.4	659.7	6	0	0
Trichome SQS+FPS plastidic	17	101.8	3.6	203.5	6	3	5

Transgenic plantlets generated with the indicated construct (expression promoter, gene(s) and intracellular targeting information) were propagated under sterile conditions until they established root systems and were ready for growth in the greenhouse. The greenhouse grown plants were screened several independent times for their squalene content. In the data shown, plants were from 40 to 60 cm tall and the first fully expanded leaf was sampled. Data represents the average from all the independent plant lines sampled twice, as well as the minimum (min) and maximum (max) observed. Plants were scored as dwarf if their height was 25% less than their sibling plants. Plants were scored as chlorotic if there was obvious yellowing within 3 or more leaves.

Table 2.2 Developmental accumulation of squalene in T1 greenhouse grown plants

Construct	Line designation	Leaf development	Squalene (µg/g fw)
Wild type (control)	# 21	young	1.9
		developing	3.4
		mature	2.3
Trichome SQS cytosolic	# 21	young	1.3
		developing	4.8
		mature	9.8
	# 44	young	3.6
		developing	4.1
		mature	2.3
Trichome SQS plastidic	# 39	young	7.1
		developing	16.7
		mature	26.3
Constitutive SQS+FPS cytosolic	# 16	young	6.4
		developing	5.5
		mature	7.4
	# 204	young	2.6
		developing	1.9
		mature	5.6
Trichome SQS+FPS cytosolic	# 27	young	5.7
		developing	8.4
		mature	5.3
Constitutive SQS+FPS plastidic	#7	young	30.1
		developing	121.15
		mature	147.4
	#15	young	329.3
		developing	450.4
		mature	667.5
Trichome SQS+FPS plastidic	#21 dwarf	young	90.0
		developing	74.3
		mature	256.7
	#31 mosaic	young	527.6
		developing	594.7
		mature	1760.2

Individual, second generation (T1) plants were propagated in the greenhouse and identified on the basis of preliminary chemical profiling screens. On average, young leaves were 5-7.5 cm in length, developing leaves were 10 -15 cm in length, and mature

leaves were greater than 20 cm long. Distinct phenotypes are noted for the plants harboring the constructs for trichome specifically expression of SQS and FPS targeted to the plastid compartment (see Figure 2.3 for examples).

Table 2.3 Field performance and squalene accumulation by select, segregating transgenic lines

Plant line	**Height** (cm)	**Weight** (kg)	**Leaf area** (cm^2)	**Photosynthesis** (µmol CO_2/m^2•sec)	**Conductance** (mol H_20/m^2•sec)	**Ci** (µmol CO_2/mol air)	**Squalene** (µg/g fw)
Wild type (control)	46.8 ± 13.6	1.0 ± 0.4	13887 ± 4586	23.4 ± 2.2	0.6 ± 0.1	254.0 ± 8.6	nd
32-	50.2 ± 12.9	1.0 ± 0.3	11313 ± 3356	21.5 ± 3.1	0.5 ± 0.1	249.3 ± 19.0	nd
32+	18.9 ± 5.8	0.2 ± 0.1	2398 ± 1128	6.9 ± 2.2	0.6 ± 0.2	327.8 ± 11.2	79.9 ± 36.6
42-	58.4 ± 8.1	1.1 ± 0.3	11704 ± 2956	21.3 ± 2.1	0.5 ± 0.1	249.6± 15.6	nd
42+	39.5 ± 6.3	0.6 ± 0.1	7776 ± 1650	21.4 ± 2.5	0.4 ± 0.1	243.2 ± 20.5	112.0 ± 27.8

Segregating (T1 generation) seed for transgenic lines targeting SQS and FPS to the chloroplast under the direction of trichome specific promoters (line 32) or constitutive promoters (line 42) were grown under greenhouse conditions for approximately 6 weeks prior to transplanting them to the field. Plants were chosen randomly for field planting and the plants grown for a total of 60 days, which is approximately two-thirds of a full growth cycle. The plants were screened several times throughout the growing season, but the data shown is for the final data collection at the end of the field season. For squalene determinations, two leaf discs of 2 cm diameter collected from the first fully expanded leaf from the top of each plant were extracted with organic solvent and their squalene content determined by GC-MS. For line 32, 6 plants accumulated significant squalene (denoted as +), while 18 others did not (-). For line 42, 14 plants accumulated squalene (+) and 8 did not (-). Six control plants were evaluated. nd, no squalene detected, below the detection limit of ~0.5 µg/g fw. Measurements of photosynthetic gas exchanges were conducted between 10AM and 12PM on a cloudless day with light intensity of 1,500 µmol/m^2•sec with a LI-COR 6400 portable instrument.

Table 2.4 Developmental accumulation of squalene in field-grown plants

construct	Plant	Leaf development	**Squalene** (µg/g fw)
Constitutive SQS + FPS plastidic	# 10	Young (4th)	91.8
		Developing (7th)	235.1
		Mature (9th)	349.4
		Senescing (11th)	51.1
	# 21	Young (2nd)	nd
		Developing(5th)	83.7
		Maturing (8th)	163.3
		Mature (11th)	219.5
		Senescing	263.2

Leaves at various positions of two plants within line 42 were sampled for their squalene levels. Plants were grown for 60 days in the field and appeared visually comparable to one another. nd, not detected, below detection limit of 0.5 µg/g fw.

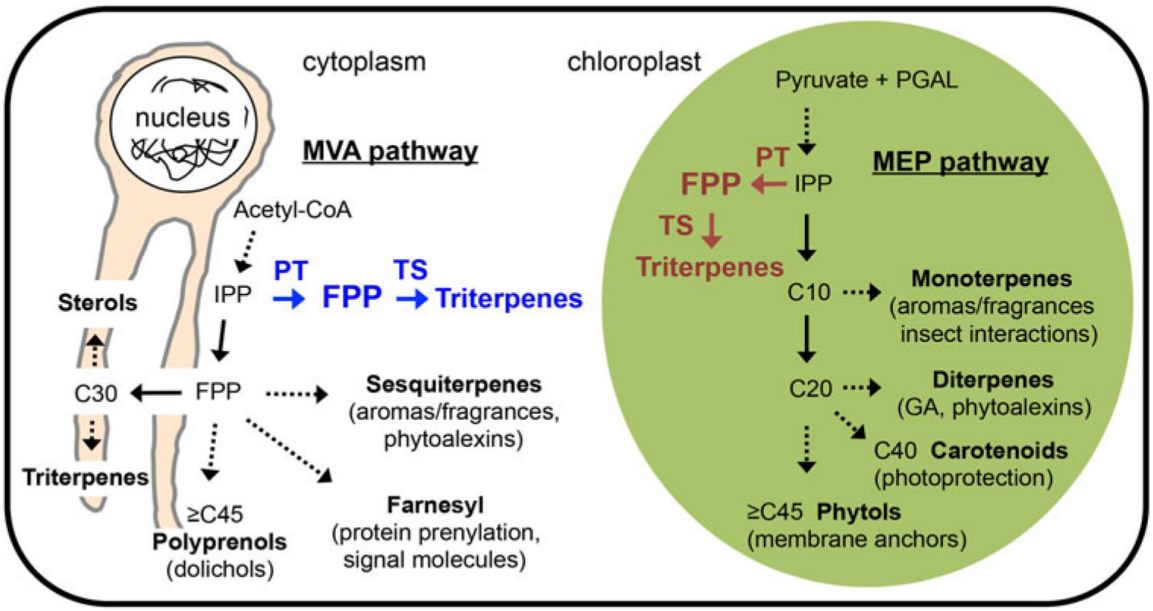


Figure 2.1 A depiction of the mevalonate (MVA) and methylerythritol phosphate (MEP) pathways operating natively in plants and their contributions to the biosynthesis of particular classes of terpenes (black), along with a conceptualization for targeting novel triterpene metabolism to the cytoplasm (blue) or to the chloroplast (red) compartments. PT, prenyl transferase (*i.e.* farnesyl diphosphate synthase); TS, triterpene synthase (*i.e.* squalene synthase).

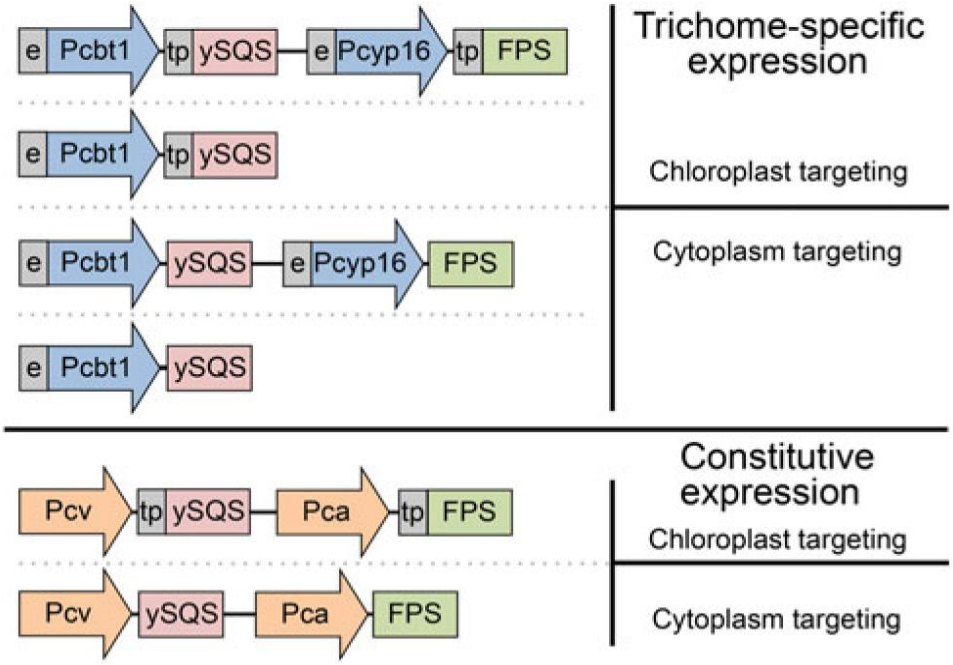


Figure 2.2 Gene constructs used to introduce squalene synthase (SQS) and farnesyl diphosphate synthase (FPS) genes into the genome of transgenic plants, and to target the encoded catalytic activities to the cytoplasm or chloroplast (tp) compartments. Promoters: cbt1, cembredienol synthase (Ennajdaoui et al., 2010); cyp16, cytochrome P450 71D16 (Wang et al., 2002a); e, duplicated transcriptional enhancer sequence found in the CaMV 35S promoter (Benfey et al., 1990); cv, 35S cassava mosaic virus (Verdaguer et al., 1998); ca, 35S cauliflower mosaic virus (Benfey et al., 1990). Genes; ySQS, the yeast squalene synthase with a truncation of 168 bp at the 3' end (Zhang et al., 1993); FPS, the chicken farnesyl diphosphate synthase (Tarshis et al., 1994); tp, the chloroplast targeting signal sequence from the *Arabidopsis* Rubisco small subunit gene (Lee et al., 2006).

Figure 2.3 Example phenotypes of plants engineered with genes encoding for SQS and FPS targeted to the chloroplast and directed by trichome specific promoter (left hand pictures) or constitutive viral promoters (center pictures), relative to wild type plants (right hand pictures). T2 seed from the trichome promoter line 31 and constitutive promoter line 5 were germinated in the presence of hygromycin, while the wild type control seed was germinated on medium without antibiotic. Six week old plants were transferred to the greenhouse and grown for several months before representative plants were chosen for these pictures.

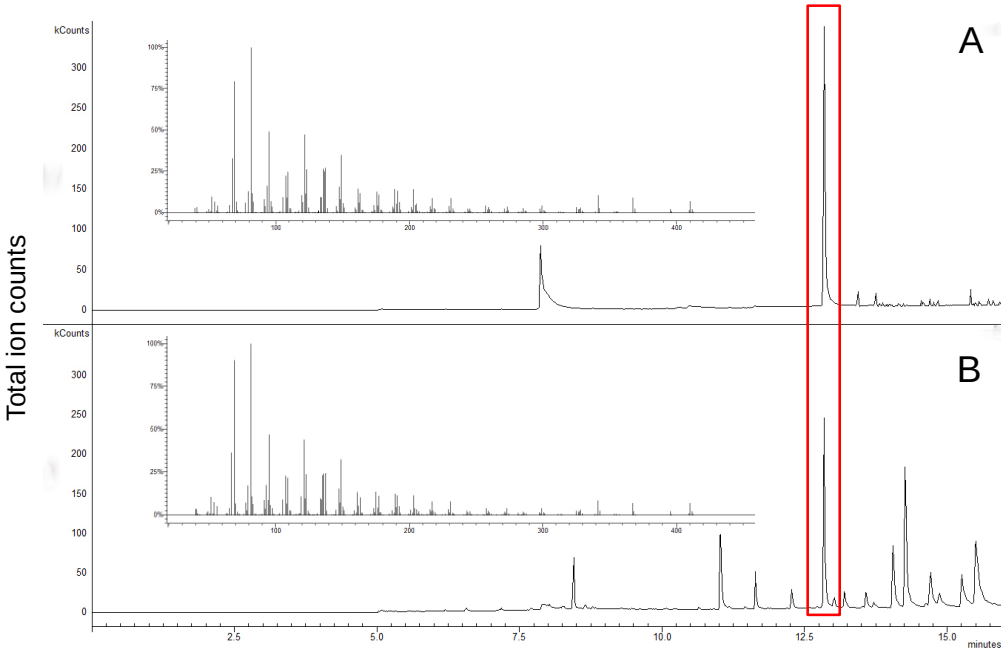


Figure 2.4 GC-MS comparison of transgenic leaf hexane extract (B) to authentic squalene (A). Leaf material of homozygous line #5 expressing plastid target SQS and FPS under the direction of the constitutive promoters was ground in liquid nitrogen, extracted with hexane:ethyl acetate (85:15), the extract concentrated under nitrogen, then fractionated on a silica column. An aliquot of the flow through fraction was then analyzed by GC-MS (B) in comparison to a squalene standard (B). The MS for the 12.85 min peak in each sample is shown in the inset.

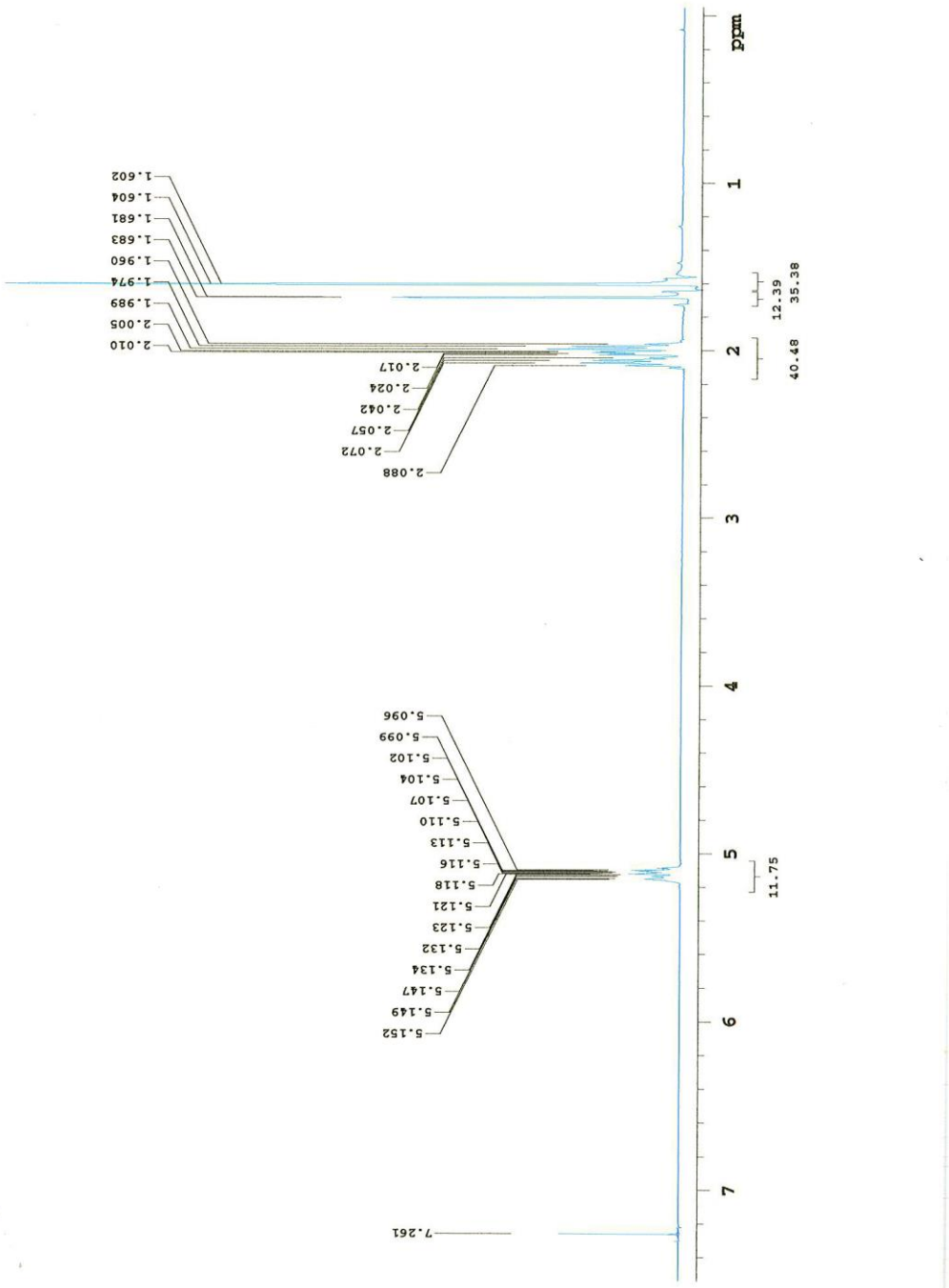


Figure 2.5A. ^{1}H-NMR spectrum of isolated squalene produced in planta. (500 MHz, CDCl3).

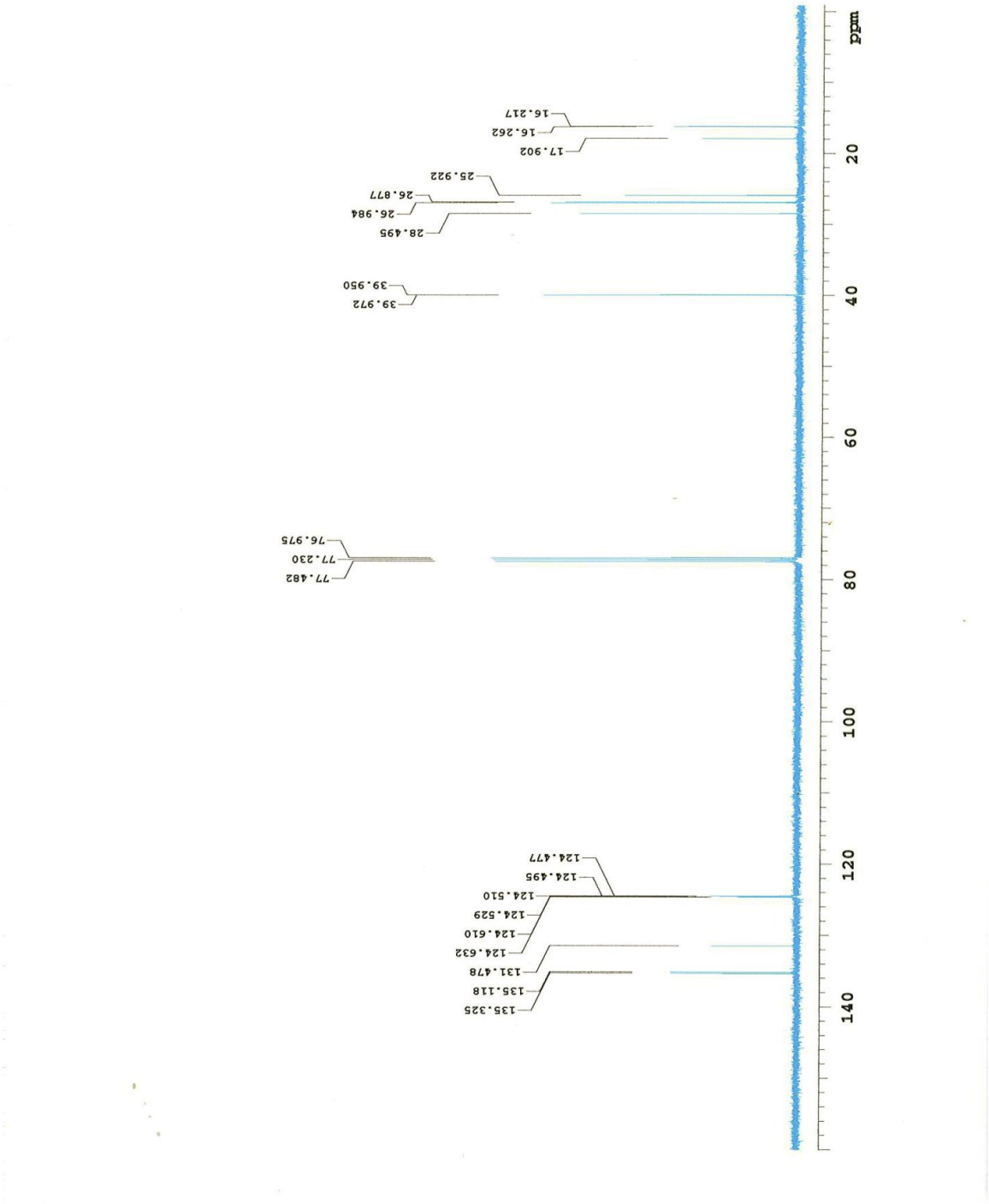


Figure 2.5B ^{13}C-NMR spectrum of isolated squalene produced in planta (500 MHz, CDCl3).

Figure 2.5 Structure identification of squalene. The structure of purified squalene from tobacco was determined ^{1}H-NMR and ^{13}C-NMR spectral analyses. ^{1}H-NMR and ^{13}C-NMR spectra were recorded on a 500 MHz Varian J-NMR spectrometer at 300 K. Chemical shifts were referenced to solvent peaks, namely δH 7.24 and δC 77.0 for CDCl3. (6E,10E,14E,18E)-squalene. Colorless oil. GC-MS mass: 410.5 amu (M+). ^{1}H-NMR (500 MHz, CDCl3) δH 1.60 (s, R-CH3, 18H), δH 1.68 (s, R-CH3, 6H), δH 1.99-2.09 (m, R-CH2-R', 20H), δH 5.10-5.15 (m, R=CH, 6H). ^{13}C-NMR (125 MHz, CDCl3) δC 15.9 (=CH-CH3, 2C), δC 16.0 (=CH-CH3, 2C), δC 17.7 (=CH-CH3, 2C), δC 25.9 (=CH-CH3, 2C), δC 26.88 (=CH-CH2-R, 2C), δC 26.98 (=CH-CH2-CH2, 2C), δC 28.5 (=CH-CH2-CH2, 2C), δC 39.95 (=CH-CH2-CH2, 2C), δC 39.97 (=CH-CH2-CH2, 2C), δC 124.5 (=CH, 2C), δC 124.6 (=CH, 2C), δC 124.63 (=CH, 2C), δC 131.5 (=CH, 2C), δC 135.1 (=CH, 2C), δC 135.3 (=CH, 2C).

Chapter 3: Engineering botryococcene and methylated triterpene production in tobacco

3.1 Summary

Terpenes comprise a large and diverse class of natural products, and many of them have attracted interest because of their physiological functions, therapeutic and industrial values. Triterpene oils including squalene (C30), botrycococcene (C30) and their methylated derivatives (C31-C37) generated by the green algae *Botryococcus braunii* Race B, which have recently received significant attention because of their utility for advanced biofuels. However, the slow growth habit of *B. braunii* makes it impractical as a robust biofuel production system. In this study, we evaluated the potential of generating high levels of botryococcene (C30) production in tobacco plants by diverting carbon flux from the cytosolic MVA pathway or the plastidic MEP pathway by overexpressing an avian farnesyl diphosphate synthase along with two versions of botryococcene synthases targeted to the cytoplasm (MVA pathway) or the chloroplast of (MEP pathway) cells. Up to 544 µg/g fresh weight of botryococcene was achieved in our transgenic plants when this metabolism was directed to the chloroplasts, which is approximately 90-times greater than that accumulating in the plants engineered for cytosolic production. To test if methylated triterpenes could be produced in tobacco, we also engineered triterpene methyltransferases (TMTs) from *B. braunii* into wild type plants and transgenic tobacco plants selected for high level of triterpene accumulation. We observed that up to 91% of the total triterpenes content was converted to methylated forms (C31, C32) by targeting the TMTs to the chloroplasts of transgenic plants, whereas only 4-14% of total triterpenes were methylated when TMTs were directed to the cytoplasm. When the TMTs were over-expressed in the cytoplasm of wild type plants without engineering triterpene biosynthesis, up to 72% of the total squalene was methylated. Interestingly, the level of total triterpene (C30+C31+C32) was elevated up to 52µg/g, a 7-fold increase relative to endogenous squalene accumulated in wild type plants. Interestingly, botryococcene accumulating lines, both with and without corresponding TMT expression, exhibited a unique phenotype that was not observed in squalene (with and without TMT expression) accumulating lines.

3.2 Introduction

Terpenes and terpenoids represent a distinct class of natural products (Buckingham, 2003) that are derived from two universal 5-carbon precursors: isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP). In eukaryotes, IPP and DMAPP are synthesized via the mevalonate (MVA) pathway whereas in prokaryotes they are synthesized via methylerythritol phosphate (MEP) pathway. In higher plants, both pathways are present in separate compartments and are believed to operate independently: the MVA pathway in the cytoplasm, is predominately responsible for sesquiterpene (C15), triterpene (C30), and polyprenol (>45) biosynthesis and is known to be associated with the endoplasmic reticulum (ER) system. The MEP pathway resides in plastids and is dedicated to monoterpenes (C10), diterpenes (C20), carotenoids (C40) and long-chain phytol biosynthesis (Figure 2A). These compounds are usually produced in plants and microbes, and play a variety of physiological (i.e. hormones, aliphatic membrane anchors, maintaining membrane structure) and ecological roles (i.e. defense compounds, insect/animal attractants) in general growth and developmental process (Kempinski et al., 2015). Terpenes are also important products in various industrial applications ranging from flavors and fragrances (Schwab et al., 2008) to medicines (Shelar et al., 2011). Many of them have attracted interests because of their therapeutic uses, industrial value and potential for advanced biofuels (Dewick, 2009; Niehaus et al., 2011; Wu et al., 2012).

The significant potential to use terpenes as biofuel feedstocks has received much attention recently. Reported isoprenoid-derived biofuels includes farnesane (Rude et al., 2009; Renninger et al., 2008), bisabolane (Peralta-Yahya et al., 2011), pinene dimers (Harvey et al., 2010), isopentenal (Withers et al., 2007), and botryococcene (Glikson et al., 1989; Hillen et al., 1982; Mastalerz et al., 1996; Moldowan et al., 1980). The richness of rings and branches within these hydrocarbon scaffolds usually correlate with a high-energy content, which enables them to have similar properties and be utilized as suitable alternatives to crude petroleum to generate gasoline, diesel and jet fuel (Peralta-Yahya et al., 2010). Indeed, some of them are already found in the components of petroleum-based fuels. One of the best examples of this is the triterpene hydrocarbon oils accumulating in the green algae *Botryococcus braunni* race B, which is considered as a major progenitor to oil and coal shale deposits on Earth (Moldowan et al., 1980). This alga has been well studied and the major constituents of its prodigious hydrocarbon oils

are a group of triterpenes including squalene (C30), and the organismal-specific, botryococcene (C30), methylated squalene (C31-C34) and methylated botryococcene (C31-C37) (Huang et al., 1989; Metzger et al., 1988; Okada et al., 1995), which can be readily converted into all classes of combustible fuels under hydrocracking conditions (Hillen et al., 1982b).

The unique biosynthetic mechanism for these triterpenes in *Botryococcus* was recently described by Niehaus et al. (2011), in which a series of novel squalene synthase-like genes were identified. In short, squalene synthase-like enzyme (SSL-1) performs a head-to-head condensation of two farnesyl diphosphate (FPP) molecules into presqualene diphosphate (PSPP), followed by a reductive rearrangement to yield squalene (C30) by enzyme SSL-2, or converted by SSL-3 to form botryococcene through a different reductive rearrangement (Figure 3.1) (Niehaus et al., 2011). Methylated derivatives are the dominant triterpene species generated by *Botryococcus brauni* race B in variable amounts under natural and artificial growth conditions (Metzger et al., 1988; Metzger et al., 1985) and these derivatives are known to yield better quality fuels due to their higher energy content by virtue of having more hydrocarbon branches. Triterpene methyltransferases (TMTs) that can methylate squalene and botryococcene were successfully characterized by Niehaus et al. (2012). Briefly, triterpene methyltransferase 1 (TMT-1) and triterpene methyltransferase 2 (TMT2) prefers squalene C30 as substrate for production of mono- (C31) or dimethylated (C32) squalene, while TMT-3 prefers botryococcene as substrate for the biosynthesis of mono- (C31) or dimethylated (C32) botryococcene (Figure 3.1) (Niehaus et al., 2012).

Like the majority of identified methyltransferases, these TMTs utilize S-adenosylmethionine (SAM) as the methyl donor, which is ubiquitously present in both prokaryotes and eukaryotes (Liscombe et al., 2012; Scheer et al., 2011). In plants, SAM is one of the most abundant co-factors (Fontecave et al., 2004; Sauter et al., 2013), and is exclusively synthesized in the cytosol. While it is mainly used as a methyl donor in methylation reaction (Ravanel et al., 2004) It also serves as the primary precursor for the biosynthesis of the ethylene (Wang et al., 2002), polyamines (Kusano et al., 2008), and nicotianamine (Takahashi et al., 2003a), which play a variety of important roles for plant growth and development (Huang et al., 2012; Sauter et al., 2013). The SAM present in the chloroplast is strictly imported from the cytosol by specific SAM/S-adenosylhomocysteine (SAH) exchangers that reside on the envelope membranes of

plastids (Ravanel et al., 2004; Bouvier et al., 2006). The imported SAM is involved in the biogenesis of aspartate-derived amino acids (Curien et al., 1998; Jander et al., 2009; Sauter et al., 2013) and serves as the methyl donor for the methylation of macromolecules, such as plastid DNA (Ahlert et al., 2009; Nishiyama et al., 2002) and proteins (Alban et al., 2014; Houtz et al., 1989; Niemi et al., 1990; Trievel et al., 2003; Ying et al., 1999), and small molecule metabolites such as prenyllipids (e.g. plastoquinone, tocopherol, chlorophylls and phylloquinone (Bouvier et al., 2006; Bouvier et al, 2005; DellaPenna, 2005).

Although plants and microbes are the major natural sources for useful terpenes, most of them are produced in a very small amount and often as complex mixtures. *Botryococcus braunii*, produces large quantities of triterpenes, but its slow growth makes it undesirable as a possible biofuel production platform (Niehaus et al., 2011). Nevertheless, metabolic engineering and synthetic biology offer strategies to manipulate terpene metabolism in various biological systems through genetic modification, in order to engineer production of highly-valued terpene with high yield and high fidelity for particular practical applications (Nielsen et al., 2011). Many successes have been achieved in engineering valuable terpenes in heterotrophic microbes, such as *Escherichia coli* (Nishiyama et al., 2002; Martin et al., 2003; Ajikumar et al., 2010), and *Saccharomyces cerevisiae* (Ro et al., 2006; Westfall et al., 2012). The strategies that have been developed in these efforts, usually take advantage of specific microbe strains whose innate biosynthetic machinery are genetically modified to accumulate certain prenyldiphosphate terpene precursors (*e.g.* isopentenyl diphosphate [IPP] or farnesyl diphosphate [FPP]), which can be utilized by further engineered heterologous terpene synthase(s) for production of the desired terpene(s). An example of this for isoprenoid-derived biofuel production, is >900 mg/L of bisabolene produced when plant bisabolene synthase genes were introduced into FPP-overproducing *E. coli* or *S. cerevisiae* strains (Peralta-Yahya et al., 2011); Farnesane production for diesel fuels was also achieved by reductive hydrogenation of its precursor farnesene, which was generated in genetically engineered yeast strain using plant farnesene synthases (Renninger et al., 2008; Ubersax et al., 2010). However, terpene production using microbial platforms is still dependent on exogenous feedstocks (*i.e.* sugars), which significantly increase the cost for production.

Compared to microbial systems, engineering terpene production in plant systems seems like an attractive target as well. This is because plants can take advantage of

photosynthesis by using atmosphere CO_2 as their carbon resource instead of relying on exogenous carbon feedstocks. Moreover, crop plants such as tobacco can generate a large amount of green tissues efficiently when grown for biomass production (Schillberg et al., 2003; Andrianov et al., 2010), which make plants a robust, sustainable and scalable platform for large-scale biofuel production. Nonetheless, compared to microbial platforms, there have only been a few examples of elevating terpene production in bioengineered plants. This is partly due to higher plants being complex multicellular organisms, in which terpene metabolism generally utilizes much more complex innate machinery which can be compartmentalized intracellularly and cell/tissue specific (Kempinski et al, 2015; Lange and Ahkami, 2013). Tremendous efforts have been made to overcome these obstacles to improve the production of valuable terpenes in plants, such as monoterpenes (Lücker et al., 2004; Ohara et al., 2010; Lange et al., 2011b), sesquiterpene (Aharoni et al., 2003; Kappers et al., 2005; Wu et al., 2006; Davidovich-Rikanati et al., 2008), diterpene (Besumbes et al., 2004; Anterola et al., 2009), and triterpene (Inagaki et al., 2011; Wu et al., 2012). Among these, engineering terpene metabolism in a heterologous organelle, where the engineered enzymes/pathway can utilize unlimited/unregulated precursors as substrates (without endogenous regulatory mechanism), appears more successful. For example, Wu et al. (2012) expressed an avian FPP synthase (FPS) with foreign sesquiterpene/triterpene synthases targeted to the plastid, to divert IPP/dimethylallyl diphosphate (DMAPP) pool from the plastidic MEP pathway to synthesize high levels of the novel sesquiterpenes, patchoulol and amorpha-4,11-diene (up to 30 µg/g fresh weight) and the triterpene, squalene (up to 1000 µg/g fresh weight)(Wu et al., 2006; Wu et al., 2012). This strategy appears to be more robust because it avoids possible endogenous regulation of sesquiterpene and triterpene biosynthesis operating in the cytoplasm, as well as utilizing the robust IPP/DMAPP pools inherent in the plastid which are generate from carbon derived from the local CO_2 fixation.

The goal of this study is to engineer unique triterpene metabolism from *Botryococcus* into tobacco, to serve as a plant platform for triterpene oil production, which could be a potential way to help alleviate the problem of the world energy crisis. In order to achieve this, we first introduced the key steps of botryococcene biosynthesis into specific subcellular compartments of tobacco cells under the direction of constitutive promoters or trichome specific promoters. The transgenic lines expressing the enzymes in the

chloroplast were found to accumulate the highest levels of botryococcene. Triterpene methyltransferases were next introduced into the same intracellular compartments of selected high accumulating lines. A high yield of methylated triterpenes was also achieved in transgenic lines when the TMTs were targeted to the chloroplast. Through careful comparison of the levels of triterpenes and the methylated triterpene products in the various transgenic lines we have also gained a deeper insight into subcellular distribution of the triterpene products in these transgenic lines, as well as a better understanding of methylation metabolism for specified metabolites in particular compartments. These findings all contribute to our understanding of regulatory elements that control the carbon flux through the innate terpene biosynthetic pathways.

3.3 Results

3.3.1 Engineering botryococene synthase genes into particular subcellular compartments in tobacco

The earlier study demonstrated that plastidic engineering of a foreign squalene synthase coupled with FPP synthase can successfully divert carbon flux from the MEP pathway to accumulate a high level of squalene in transgenic tobacco (Wu et al., 2012).That study revealed that available IPP/DMAPP precursors are adequate and strong regulatory mechanisms are absent in the chloroplast for novel triterpene (C30) production to occur. This, in turn, leads us to attempt this strategy to the metabolic engineer of botryococene (C30) biosynthesis into tobacco plants. However, botryococcene biosynthesis requires two squalene synthase-like enzymes, SSL-1 and SSL-3, to catalyze successive reactions to make the botryococcene product. This is in contrast to squalene biosynthesis which requires only a single enzyme, squalene synthase (Figure 3.1, Niehaus et al., 2011). We chose to over-express two chimeric versions of botrycococcene synthase: one is SSL1-3 (Figure 3.2B), which is a fusion of the SSL-1 and SSL-3 enzymes by a peptide linker, which exhibited a 2-fold greater accumulation of botryococcene when expressed in yeast in comparison to simple co-expression of the two enzymes separately (Niehaus et al., 2011). The second design is referred to as SSL1-3M (Figure 3.2B) in which the SSL1-3 chimeric enzyme has 71 amino acids of carboxy-terminal of *Botryococcus* squalene synthase (M) appended to its C-terminal. This construct thus contains a membrane spanning domain that was hypothesized to improve botryococcene productivity by integrating the enzyme into the ER membrane in

order to promote proximity between enzymes for substrates (Niehaus et al., 2011). The overall gene constructs thus consist of either botryococene systhase SSL1-3 or SSL1-3M directed by cassava mosaic promoter (Pcv) (Verdaguer et al., 1996), coupled with avian FPP synthase gene (FPS) (Tarshis et al., 1994), driven by the 35S-CaMV promoter (Pca) (Benfey et al., 1990). An amino terminal, plastid targeting signal sequence (tp) from the Rubisco small subunit gene of *Arabidopsis* (Lee et al., 2006) was also inserted onto the chimeric SSL1-3 constructs to target these enzymes to the chloroplast compartment, whereas constructs without the signal sequence would target the encoded proteins to the cytoplasmic compartment. The respective gene constructs (Figure 3.2B, table 3.1) were introduced into *Nicotiana tabacum* accession KY 1068 by standard agrobacterium transformation. Thirty or more T0 independent transgenic lines were generated and the leaf materials from different transgenic plants were extracted and analyzed by GC-MS and GC-FID. When evaluated by GC–MS, a unique molecule was detected in the extraction from some of the transgenic plants (Figure 3.3E) that was not evident in any of the wild type plants (Figure 3.3A). This unique chemical peak had identical mass spectrum (410 *amu*) (Figure 3.8A) as compared to botryococcene standard (Niehaus et al., 2011). This molecule was also confirmed as botryococcene by ^{1}H-NMR and ^{13}C-NMR analysis (see below).

We observed that transgenic lines engineered with the construct that directed botryococcene synthase (SSL1-3) along with FPS to the chloroplasts (tpSSL1-3+tpFPS) (Figure 3.2B) generate a high level of botryococcene (544 µg/g fw, maximum and 269 µg/g fw, average) (Figure 3.2C, Table 3.1), which is about 70-90 fold increase over the level of botryococcene (6.3 µg/g fw, maximum and 3.5 µg/g fw, average) accumulated in the lines (SSL1-3+FPS)(Figure 3.2B) with the same enzymes targeted to the cytoplasm. The results indicate that the chimeric SSL1-3 enzyme efficiently uses FPP as a substrate derived from the universal C5 precursors present in the chloroplast compartment supported by the accompanying engineered FPS. In contrast, the failure to enhance botryococcene yield by cytosolic engineering is most likely because the cytosolic FPP pool is low and highly regulated, even if avian FPS was used to over-ride potentially regulatory mechanisms in the cytoplasm (Wu et al., 2006; Wu et al., 2012). The overall production of botryococcene by plastidic engineering and its fold increase over that achieved by cytosolic engineering coincided well with what was found earlier for engineering squalene biosynthesis by Wu et al. (2012).

A similar contrast was also found in comparison of production by engineering the membrane tethered version of botryococcene synthase SSL1-3M and FPS in the chloroplasts (tpSSL1-3M) to that directed the same metabolism to the cytoplasm (SSL1-3M). A relatively high amount of botryococcene accumulation was achieved by plastidic engineering, with a maximum level of 202 µg/g fw and average level of 131 µg/g fw, which is about a 10 to 20-fold increase over that for cytosolic engineering with a maximum of 16.4 µg/g fw and average of 5.8 µg/g fw (Table 3.1, figure 3.2). The low production by cytosolic engineering of SSL1-3M suggests again a limited flux of carbon and strict endogenous regulation might be occurring in the cytoplasm, but absent in the chloroplast. The membrane domain (M) attaching to SSL1-3 were used to help the enzymes associate with ER for accessing more available substrates in the cytoplasm, which might explain why cytosolic engineering of SSL1-3M accumulated a slightly higher level of botryococcene than that was achieved by cytosolic engineering SSL1-3. Interestingly, plastidic engineering of SSL1-3M yielded only half the level of botryococcene produced by engineering a soluble form of SSL1-3 in chloroplast, which is in contrast to what was observed in a yeast system, where membrane tethered enzyme increased the yield about 5 times more than that produced by the same enzyme without the membrane-spanning domain (Niehaus et al., 2011). One possible reason for this could be the different effects of the two intracellular environments on the enzymes' activity, in that chloroplast stroma, which is known as soluble subcellular environment could be more favorable to soluble enzymes whereas in the yeast cell that has a full of endo-membrane system may be more favorable to membrane-associated enzymes. Another possibility could be attributed to the membrane domain (M) that associates the enzyme with the ER, which could reduce some of the catalytic activity when imported and integrated into the chloroplast compartment.

We also observed that botryococcene accumulation exhibited a significant developmental-dependent pattern of accumulation. The level of botryococcene accumulated in mature leaves was 2 to 4-fold higher over that in their young leaves (Table 3.1, Figure 3.2). That could be the result of the engineered enzymes being constitutively expressed having more time to synthesize products reflected in the mature plants over that in young plants. In addition, there is no known mechanism in plants or any other organisms for the catabolism of botryococcene. The integrity of the botryococcene accumulated by these transgenic lines was also verified by extracting

100 mg of leaves from a plant constitutively expressing construct tpSSL1-3+tpFPS with hexane and the putative botryococcene compound purified by successive rounds of silica chromatography. The isolated compound then was analyzed by NMR analyses (Figure 3.9). ^{1}H-NMR and ^{13}C-NMR data were recorded on a 400 MHz Varian J-NMR spectrometer at 300 K. Chemical shifts were referenced to solvent peaks, namely δH 7.24 and δC 77.0 for CDCl3. Botryococcene isolated from our transgenic tobacco had all of the expected signals as compared to published data (supporting information from Pulis and Aggarwal, 2012). Most importantly, the vinyl protons at C-26 (δ_H 5.82, 1H, dd, *J*=18, 11 Hz), C-27 (δ_H 4.95, 1H, dd, *J*= 11, 1 Hz; δ_H 4.94, 1H, dd, *J*= 18, 1 Hz), and olefinic protons of C-11 (δ_H 5.33, 1H, dd, *J*=16, 1 Hz) and C-12 (δ_H 5.20, 1H, dd, *J*=16, 8 Hz) were apparent. In the ^{13}C spectrum, the corresponding signals for C-26 (δ_C 146.97), C-27 (δ_C 111.32), C-11 (δ_C 136.03), and C-12 (δ_C 133.95) were all present. Altogether, the ^{1}H and ^{13}C-NMR analyses confirmed the identity of C30 botryococcene.

3.3.2 Trichome specific expression of botryococcene metabolism

Besides the constitutive viral promoters, two trichome specific promoters cbts and cyp16 derived from cembratrien-ol synthase and cembratriene-ol hydroxylase genes (Ennajdaoui et al., 2010; Wang et al., 2002) were also used to drive botryococcene synthase and FPS expression, respectively. Trichomes are specialized organs located on the surface of the aerial parts of plant species, which are the site of production of abundant secondary metabolites, which in certain species, such as tobacco, may represent up to 15% of the leaf dry weight (Wagner et al., 2004). This large contribution to leaf biomass relative to the actual volume of the glandular trichomes makes trichome engineering an attractive target for metabolite bioengineering (Ennajdaoui et al., 2010). In order to strengthen overall expression by the trichome promoters, the 35S double enhancer was also appending to the 5' prime end of each trichome promoters (Wu et al., 2012). Four constructs harboring SSL1-3 or SSL1-3M with FPS, plus/minus chloroplast amino terminal sequences (tp) were also introduced into wild type tobacco plants (*Nicotiana tabacum* accession 1068), which is known to have high trichome density (Nielsen et al.1982). More than 30 independent transgenic lines were generated and analyzed for botryococcene content (Table 3.1).

We found that targeting metabolism to the chloroplasts driven by trichome specific promoters achieved a moderate level of botryoccocene production in both young and

middle-aged leaves (below 30 µg/g fw average), which is also 10-20 fold increase over that accumulated in the lines with the enzymes directed to the cytoplasm. Unexpectedly, most of the lines with trichome specific expression of SSL1-3 and FPS in the chloroplast showed a strong chlorotic, white, mottling, dwalf phenotype (Figure 3.10) which may have contributed to difficulties in propagating these materials. This phenotype was more serious than anything observed with trichome specific expression of squalene biosynthesis (Wu et al., 2012). This adverse phenotype resulting from use of the trichome specific promoters to direct botryococcene biosynthesis was surprising and suggested that the double 35S enhancers may have conflicting effects when used in combination with trichome specific promoters, thus making this design unsuitable for metabolic engineering.

3.3.3 Engineering triterpene methyltransferases into particular subcellular compartments in tobacco

The success in engineering squalene and botryococcene C30 production in transgenic tobacco led us to take advantage of these high triterpene accumulating lines as platforms to generate methylated triterpenes (C31-C32) products. Our working hypothesis was that if we introduce triterpene methyltransferases into these lines, the accumulating triterpene (C30) could be converted to their methylated forms. However, this hypothesis involved several fundamental questions that remained to be considered: 1. Subcellular localization of C30 triterpene: although there was evidence that the large amounts of C30 triterpene synthesized by enzymes targeted to the chloroplast, the exact subcellular localization of this capacity was inferred, and we could not exclude the possibility that some of this biosynthesis may be occurring in the cytoplasm; 2. SAM consumption: methylation of large amounts of triterpenes would require stoichemical amounts of SAM as co-substrates, therefore would there be sufficient amounts of SAMs to power these reactions in various locales, and if so, what would be the effects on SAM availability for native metabolism; 3. Enzyme specificity: would the three TMTs exhibit the same substrate specificity in *planta* as in yeast (Niehaus et al., 2012); 4. Enzyme solubility: these methyltransferases all contain Transmembrane Domain (TMD), which might influence their solubility properties and could affect their transport and activity in any particular subcellular compartment, such as the chloroplast stoma.

To address these questions, all three of the TMTs genes were individually constructed into a plant transformation vector harboring a 35S promoter, with a kanamycin resistance gene, and separately engineered into a squalene-accumulating line constitutively expressing enzymes targeted plastid compartment (#5 of line tpSQS+tpFPS, a T2 homozygous generation), or botryococcene accumulating lines constitutively expressing enzymes targeted to the plastid compartment (#10 of line tpSSL13+tpFPS, a T1 heterozygous generation; #31 of line tpSSL13m+tpFPS, a T1 heterozygous generation). The methyltransferases were themselves also directed to either the cytoplasm or chloroplast compartments in each of these lines in order to test for the possible subcellular localization of C30 substrates *in planta*. Ten or more independent transgenic lines were generated for each construct design. The content of methylated triterpenes, including monomethylated (C31) and dimethylated (C32) squalene (Figure 3.3C and D), and monomethylated (C31) and dimethylated botryococcene (C32) (Figure 3.3G) were successfully determined in different transgenic plants, but not in the wildtype plants (Figure 3.3A) by GC-MS and GC-FID. Each unique methylated triterpene was determined by having an identical mass spectrum (Figure 3.8) as compared to published data (Niehaus et al., 2012). In order to simplify the results, three experimental data sets are presented.

In the first experimental set, the constructs consisted of one of the TMT genes with and without the chloroplast targeting signal sequence (tp) and were introduced into high squalene accumulating line (tpSQS+tpFPS), which targeted squalene biosynthesis to the plastid compartment (Wu et al., 2012). TMTs enzymes targeted to the cytoplasm are designated as TMT-1, TMT-2, TMT-3, or when directed to the chloroplasts as tpTMT-1, tpTMT-2, tpTMT-3 (Figure 3.4B). When the chemical analysis of all the transgenic lines resulting from particular construct were averaged, we observed that the transgenic lines targeting TMT-1 and TMT-2 to the chloroplast accumulated a large proportion of methylated squalene, from the highest of 91% to an average of 65% of total triterpene for TMT-1, and 82% (highest) and 51% (average) of total triterpene for TMT-2, respectively (Table 3.2). In contrast, transgenic lines targeting TMT-1 and TMT-2 to the cytoplasm accumulated 7% (highest) and 4% (average) methylated squalenes, and 6% (highest) and 4% (average) methylated squalene for construct TMT-1 and TMT-2, respectively (Table 3.2). The total triterpene amount (C30+C31+C32) among these

accumulators are approximately similar to the parental line which accumulates only non-methylated squalene (Table 3.2).

The large amounts of methylated squalene products achieved by plastidic engineering does not only reveals that both TMT-1 and TMT-2 are able to use squalene C30 as their substrates for methylation, which corroborates the preference reported earlier (Niehaus et al., 2012), but also indicates there are at least stoichemical amounts of squalene C30 substrates and SAM are available for the TMTs in the chloroplast. While squalene C30 that remain unmethylated in this case could be either newly synthesized by the engineered SQS or pre-existing, it was nevertheless unavailable to the TMTs. This could arise because of some kinetic inconsistences between the synthase and TMT enzymes, or the accumulation of methylated products that saturated or inhibited the TMTs directly. The co-existence of both umethylated (C30) and two methylated squalene products (C31, C32) was also observed by co-expressing TMTs in squalene overproduction yeast line (Niehaus et al., 2012), and compared to which, conversion of squalene (C30) to methylated squalene (C31, C32) by engineering TMTs in plant chloroplasts was even higher (Table 3.2). However, we cannot exclude the possibility that some of this unmethylated squalene C30 is because it is in a different compartment from where the engineered TMTs were localized. The small amounts of methylated squalene produced by targeting TMTs to the cytoplasm supported this possibility and provide evidence that cytosolic expressed TMTs are only able to access limited amounts of C30 squalene substrates present in the cytosol in high squalene accumulating line (tpSQS+tpFPS). This cytosolic localized squalene C30 could be either native squalene (C30) or synthesized by mistargeted engineered squalene synthase (SQS) that is not transported to the chloroplast yet remains active in the cytoplasm (see other possibilities in discussion). In addition, the proportion of C32 (41%) was slightly higher than that of C31 (25%) in the tpTMT-1 engineered lines, whereas C31 (31%) was higher than C32 (18%) in the tpTMT-2 engineered lines, which may represent catalytic limitation of the different TMTs (Table 3.2).

In contrast, only 5 % (average) of the total squalene was methylated when TMT-3 was targeted to the chloroplast, suggesting that TMT-3 exhibited weak catalytic activity with squalene, which was also observed when it was expressed in yeast line accumulating squalene (Niehaus et al., 2012). Targeted expression of TMT-3 to the cytoplasm also did not result in any methylation products accumulating, also corroborating the inability of

the TMT-3 to utilize the limited amounts of squalene produced normally in this compartment (Table 3.3).

Secondly, when the respective TMT genes constructs were introduced into botryococcene (C30) accumulating lines, we observed a large proportion of methylated botryoccocene only when TMT-3 was over-expressed and only when it was targeted to the chloroplasts. In parental line tpSSL1-3M+tpFPS, 87% of the botryococcene was maximally methylated and on average 54% was methylated, whereas in parental line tpSSL1-3+tpFPS, 66% of the botryococcene was maximally converted and more typically 35% (average) of total botryococcene was methylated when TMT-3 was also targeted to the chloroplasts (Table 3.2). Little to none of the chloroplast synthesized botryococcene was methylated when either TMT-1 or TMT-2 were targeted to the chloroplast (Table 3.3), further demonstrating the striking preference of TMT-3 for botryococcene. By comparison, only a small proportion of methylated botryococcene was formed when TMT-3 was expressed in cytoplasm of parental line tpSSL1-3M+tpFPS, 14% maximal and 6% on average of the total triterpene, and 10% maximal and 3% on average for parental line tpSSL1-3+tpFPS (Table 3.2). Like that suggested above for squalene, the low level of methylated botryococcene produced by TMT-3 targeted to the cytoplasm could arise from botrococcene (C30) produced by mistargeted SSL1-3(M) in the cytoplasm of high botryococcene accumulating lines but not from natively synthesized botryococcene, because there was no endogenous botryococcene biosynthesis occurring in the plants (Figure 3.3A) (see other possibilities in discussion). In addition, the transgenic lines expressing TMT-3 in the chloroplast of partental line tpSSL1-3M+tpFPS shows a higher percentage of total botryococcene methylation than that of lines tpSSL1-3+tpFPS (Table 3.2), which could be simply be proportional to the lesser amount of botryococcene produced by line tpSSL1-3M+tpFPS (Figure 3.2 and table 3.1).

There was no methylated botryococcene products found when TMT-1 and TMT-2 genes were overexpressed in the chloroplasts of the botryococcene accumulating line tpSSL1-3+tpFPS. While a small amount of methylated squalene, average 37% of the total squalene, was evident in these lines (Table 3.3), which could be derived from natively produced squalene that could be methylated by mistargeted TMT-1 in the cytoplasm, or a small amount of squalene synthesized by the enzyme SSL1-3 expressing in the chloroplast. This latter explanation is entirely possible given that when SSL1-3 was co-

expressed with TMT-1, a smidgeon of methylated squalene was observed although no pool of squalene was evident to non-transgenic yeast (Niehaus et al., 2011).

Finally, based on these interesting findings of methylation of endogenous squalene, the TMT-1 and TMT-2 genes with expression of each targeted to the chloroplast and cytoplasm were introduced into control, wild type tobacco cultivar 1068 and the resulting transgenic lines screened for methylated triterpenes. Interestingly, a small proportion of methylated squalene (average 41% of total squalene) was observed when TMT-1 expressed in the chloroplasts (Table 3.4), where there was no evidence for squalene being synthesized or present (Aharoni et al., 2003). In these cases, we must assume some mis-targeting of the engineered TMTs and methylation of squalene occurring in the cytoplasm. In contrast, when TMT-1 and TMT-2 expression was targeted to the cytoplasm of wild type plants, a high proportion of methylated squalene (average 72% and 67% of total squalene respectively) was found (Table 3.4, figure 3.5). But even more surprising, the level of total squalene (C31+C32+C33) in transgenic lines expressing TMT-1 was elevated up to 55 µg/g fw (maximum) and 36 µg/g fw (average) (Table 3.4), which was about 4 to 7 fold greater than the level of endogenous squalene (C30) accumulating in wild type plants (Table 3.4 and Figure 3.5).

Squalene biosynthesis is known to be one of the key steps in sterol biosynthesis, which might be regulatory for squalene accumulation (Wu et al., 2012), but the results of relatively high levels of squalene (methylated and non-methylated forms) accumulation in the TMT over-expression lines suggests that reduction of the non-methylated squalene pool triggers the biosynthesis of additional squalene, which necessarily means an enhanced flux down the MEP pathway for triterpene production. These results have not only revealed that the algal TMTs are fully functionally when expressed in the cytoplasm, but also provide a glimpse into the regulatory complexity of squalene biosynthesis, which is crucial for homoeostatic control of sterol biosynthesis in the plants.

3.3.4 Triterpene accumulation in different tissues and leaf layers

To determine if the various transgenic plants accumulated triterpenes in other tissues beside leafs, the triterpene chemical profiles across various tissues and over developmental time were determined (Figure 3.6A). Triterpene content (either squalene or botryococcene) was found in all the tissues examined, but the levels of varied

dramatically. Not surprisingly the leaf accumulates the greatest amount of triterpene, which is about 10 to 64 fold increase over the root that accumulates the least triterpene among these tissues. A low amount of triterpene (never exceeding 25 µg/g) was also determined in other tissues veins and stems. In wild type plants, squalene level in leaf is only 1.5-3 fold greater than that in other tissues. This result provides indirect evidence that the triterpene accumulation in these transgenic lines reply on the plastidic metabolism, because the leaf as the major photosynthetic organ and harbors many more chloroplasts than these other tissue types.

The developmental accumulation of triterpene in various leaf positions was previously reported (Wu et al., 2012) , which was also seen in methylated triterpene accumulating lines (Figure 3.6B). The leaves at various positions from tobacco lines grown for 4 months in the greenhouse were sampled for their triterpene content. Interestingly, the triterpene level showed a successive increase with leaf maturation on the plant. The more mature leaves in the lower leaf positions usually had more total triterpene accumulation (Figure 3.6B, right axis). However, the ratio of methylated to total triterpene (C30+C31+C32) at the various leaf positions remained essentially the same from 55% to 75% for each transgenic line (Figure 3.6B, left axis).

3.3.5 Phenotypes of triterpene accumulating plants

Over 75% of the botryococcene accumulating lines directing this metabolism to the chloroplast exhibited several distinguishing phenotypes, including some dwarfing, chlorosis and mottling (Figure 3.7A, B, C, table 3.1). These phenotypes were not observed in any transgenic lines wherein the engineered metabolism was targeted to the cytoplasm and obviously different from what was observed in any of squalene accumulating plants (Wu et al., 2012). The results thus indicated that botryococcene accumulation had unknown toxic effects on tobacco chloroplast,plant morphology and growth while squalene did not. Moreover, there were not any noticeable differences in phenotypes between triterpene accumulating plants and their respective methylated triterpene accumulating plants (Figure 3.7D, E).

3.4 Discussion

The current work successfully introduced the key steps of unique triterpene metabolism that originally occurs in the algae *Botryococcus brauni* race B into tobacco plants, leading to a high level accumulation of botryococcene and methylated triterpenes. It utilized the strategy of diverting the C5 (IPP/DMAPP) precursors in MEP pathway to form FPP, which could then be utilized for novel triterpenes (C30) biosynthesis by the co-expression of botryococcene and squalene synthases. The accumulating triterpenes (C30) in the transgenic plants could be further methylated by targeting TMTs to the chloroplasts of these transgenic plants. Therefore, the engineered enzymes FPS, triterpene synthase and TMTs created a unique metabolic channel redirecting carbon flux from MEP pathway in chloroplasts to the production of a desired triterpene (Figure 3.2A).

The strategy was successful in taking advantage of engineering terpene metabolism in the plant chloroplasts. First, chloroplasts offer an unrestricted abundance of carbon passing through the MEP pathway, and diverting an intermediate and carbon flux from this pathway does not adversely impact the biosynthetic needs in the chloroplasts, for large amounts of carotenoid and chlorophyll. The second equally important observation is that the chloroplast provides an ideal environment for heterologous terpene production, perhaps due to lax endogenous regulation of the MEP pathway in plastids as compared to the MVA pathway operating in the cytosol (Kempinski et al., 2015).

This approach has now been demonstrated to be applicable for the metabolic engineering of various types of terpene compounds including monoterpenes, sesquiterpenes, and triterpenes in tobacco plants. However, we also note that accumulation level of each type of terpene differed significantly between the respective terpene targets. Plants engineered for triterpene production accumulated 200 to 1000 µg/g fw of triterpene, whereas sesquiterpene production has not exceeded 30 µg/g fw and monoterpene accumulation is maximally in the range of 1 µg/g fw (Wu et al., 2006) Kempinski et al. 2015). Such stark differences strongly suggest that the limitation in specific terpene class accumulation lies with the engineered terpene synthase. More specifically, catalytic efficiency and durability of the engineered terpene synthase may be the most important limitations rather than carbon source for building these designer molecules. Consistent with this notion, we found that overexpression and chloroplast

targeting of the soluble form of SSL1-3 with FPS yielded similar levels of botryococcene accumulation to that for squalene achieved by plastidic engineering of a yeast squalene synthase along with the avian FPS, but two times more than expressing the SSL1-3M enzyme form with FPS targeted to the chloroplast. It suggests that chimeric enzyme SSL1-3 functions as well as the single enzyme yeast SQS in the chloroplast, and exhibits a higher catalytic capacity than SSL1-3M could.

TMTs are functionally insoluble enzymes which exhibit an unexpectedly high catalytic activity for the methylation reaction when engineered into both the chloroplast and cytoplasm compartments of the appropriate transgenic plant lines. Up to 91% of the C30 triterpenes accumulating in high triterpene C30 accumulating lines was subsequently transformed to mono- or di-methylated triterpene when one of the 3 TMT genes targeted methyltransferase activity to the chloroplast. The methylation ratio of 51%-91% by TMTs directed to the plastid compartment versus 3%-14% by TMTs targeted to the cytosol TMTs to the cytosol provide additional evidence to show that the distribution of triterpene C30 in the high triterpene accumulating transgenic lines remained in the chloroplast. This was not unexpected because the triterpene C30 was supposedly synthesized in the chloroplast and methylation in the cytosol would come about by some mechanism, either active or passive, to export the novel triterpene out of the chloroplasts to the cytoplasm.

Therefore, in order to account for the small but significant methylation of triteprenes occurring in the cytoplasm, at least four possible routes remain plausible: First, the methylated squalene produced by targeting TMT-1 and TMT-2 to the cytoplasm in the wildtype plants proves that natively synthesized squalene can be methylated by TMTs; Second, the small amount of methylated botryococcene generated in plants wherein TMT-3 was directed to the cytoplasm while high botryococcene biosynthesis was directed to the chloroplasts [tpSSL1-3(M)+tpFPS], could arise from a low level of botryococcene (C30) biosynthesized by mis-targeted SSL1-3(M). This observation validates that TMTs can methylate cytosolic triterpene (C30) produced by mis-targeted triterpene synthase; 3. Expressing the construct of tpTMT-1 in wild type plants also resulted in methylated products, which must be derived from cytosolic endogenous squalene catalyzed by mis-targeted TMT-1. This evidence supports that mis-targeted TMTs are also able to methylate cytosolic-localized triterpenes. Fourth, the cytosolic engineered TMTs finds a way to access the plastidic-localized squalene. One example to support this is the recent discovery that plastid envelope-localized substrates can be

accessed by the enzymes targeted to ER membrane through the continuity of ER and chloroplast (Mehrshahi et al., 2013). Of course, the methylation status of triterpenes could come about by some combination of these routes, which might also be variable upon plant development and growth habit. Nevertheless, these findings provide a valuable depiction of methylation of native squalene, mistargeted enzymes in the plastidic engineered plants, and associations between cytosol and chloroplast compartments.

A final issue raised during the initial phases of this work was whether there would be sufficient SAM to support formation of the methylated triterpenes. This concern arose because of an appreciation for how important SAM is to methylation of macromolecules as well as very diverse small molecules. Fortunately, concern for SAM availability seemed unfounded regardless if the methylation reactions were targeted to the chloroplasts or to the cytoplasm.

Equally interesting was the observation that plants engineered for botryococcene accumulation tended to exhibit distinct phenotypic outcomes like dwarfism, chlorosis and mottling, while plants accumulating high levels of squalene did not show any of these adverse effects. Why this might be so is currently unknown. However, if one were able to discern how the plants were able to accumulate high levels of squalene without any negative impact on growth performance, then one might be able to use this information in the engineering of advanced accumulation mechanisms for terpenes like botryococcene.

3.5 Materials and Methods

3.5.1 Expression vector construction and plant transformation

Design of gene constructs and assembly for engineering botryococene biosynthesis were based on the work previously described by Wu et al. (2006) and Wu et al., (2012) using standard molecular methodologies. Gene constructs consisted of a peptide fusion of SSL-1 (Genebank accession: HQ585058.1) and SSL-3 (Genebank accession: HQ585060.1) connected by a triplet repeat peptide linker of GGSG, with or without appending the carboxy-terminal (71 amino acids) of the *Botryococcus* squalene synthase (Genebank AF205791.1) onto the carboxy-termini of the SSL-3, and the avian

farnesyl diphosphate synthase (FPS) gene (P08836) (Tarshis et al., 1994). The chimeric SSL1-3 genes and FPS genes were inserted downstream of strong constitutive promoters Pcv, cassava vein mosaic viral promoter (Verdaguer et al., 1996) and Pca, 35S cauliflower mosaic viral promoter (Benfey et al., 1990) respectively. For trichome specific expression of triterpene biosynthesis, the trichome-specific promoters [(Pcbt, the cembratrien-ol synthase promoter (Ennajdaoui et al., 2010) or the Pcyp16, cembratriene-ol hydroxylase promoter (Wang et al., 2002a)] were fused to 5' end of botryococcene synthase genes and FPS gene respectively. The duplicated CAMV 35S enhancer elements (Benfey et al., 1990) was fused to the 5' end of the each trichome promoter. A chloroplast targeting signal sequence (tp) encoding for the first 58 amino acids of the *Arabidopsis* Rubisco small subunit gene (NM23202)(Lee et al., 2006) was fused in-frame with the 5' end of the respective terpene synthase genes. The gene cassette were assembled together in a helper vector described in Wu et al. (2012) by standard molecular biology methods and the various DNA segments were verified by DNA sequencing. The gene cassettes were then introduced into pBDON, a modified Ti plasmid vector harboring a hygromycin resistance gene by DNA recombination (Wu et al., 2006).

The triterpene methyltransferase genes TMT-1 (JN828962.1), TMT-2 (JN828963.1), and TMT-3 (JN828964.1) were inserted directly into plant transformation vector pKYLx71 (Schardl et al., 1987), harboring a 35S viral promoter and a kanamycin resistance gene. In order to target TMT genes to the chloroplast, the chloroplast targeting signal sequence (tp) noted above was then inserted in-frame with the 5' termini of the respective TMT genes.

The engineered Ti plasmid vectors were introduced into *Agrobacterium tumefaciens* GV3850 by electroporation, and the resulting *Agrobacterium* lines were used to genetically engineer Nicotiana tabacum (tobacco) TI accession 1068 (Nielsen et al. 1982) or transgenic line of tpSQS+tpFPS #5 (T2 homozygous generation) with high level of squalene, as previously described by Wu et al. (2012), or high botryococcene accumulating transgenic lines (tpSSL1-3+tpFPS-10, or tpSSL1-3M+tpFPS-31, T1 heterzygous generation) generated in this study. Leaf explants were transformed with the respective gene constructs and the resulting calli were selected on tissue culture media with hygromycin (50 mg/l) for engineering botryococcene biosynthesis and with both hygromycin (50 mg/l) and kanamycin (250 mg/l) for engineering methylated

triterpene biosynthesis. The culture media (1L) contained 4.2 g MS salts (Phytotechnology Laboratories, Overland Park, KS), 0.112 g B5 vitamins (Phytotechnology Laboratories), 30 g sucrose, 9 g agar, 1 mg IAA and 2.5 mg Benzylaminopurine (Sigma). The selected calli were grown under sterile tissue culture conditions to regenerate plantlets. The selected T0 plantlets were then propagated in the greenhouse and assessed for triterpene content by GC–MS or GC-FID analyses.

3.5.2 Plant propagation and segregation selection

All the T0 plantlets after hygromycin or kanamycin selection were grown in common commercial vermiculite/soil blends in a greenhouse and fertilized weekly with water soluble fertilizer (20-20-20 for nitrogen, phosphorus, and potassium). Insect control was performed monthly. The T0 plants were allowed to flower in the greenhouse and the T1 seed collected for subsequent cycles of propagation. Segregation of the hygromycin and kanamycin resistance trait in the T1 seed lines was also evaluated by germinating sterilized seeds on 50 mg/l hygromycin and 250 mg/l kanamycin in T-tissue culture media (4.2 g MS salts, 0.112 g B5 vitamins, 30 g sucrose, and 9 g agar in 1 liter medium).

3.5.3 Triterpene (squalene, botryococcene, methylated squalene and botryococcene) determinations

Fifty to one hundred and fifty mg of transgenic leaf material were collected from the upper most, fully expanded leaves of tobacco plants grown in greenhouse condition. The other plant tissues roots, stem, and veins were collected from plants grown in the tissue culture condition for chemical analysis. The terpene content for each sample was determined by the methods previously described in Wu et al. (2012). Each plant sample was ground in liquid nitrogen, then extracted with 2-4 ml of a hexane:ethyl acetate mixture (v/v 85:15) containing 200 ng of α-cedrene as an external standard for quantification and calculations of recovery. The extracts were concentrated to 500 µl under a nitrogen stream without drying the sample. The concentrated extracts were then partially purified by passing through a silica column (500 mg, prepared in glass wool plugged glass pipette) and further eluted with 1 ml of the hexane solvent. After concentration of the combined eluate under a stream of nitrogen, aliquots were injected onto a GC–MS equipped with a Rtx-5 capillary column (30 m × 0.32 mm, 0.25 µm phase

thickness) with the following temperature program of 70°C for 1 min, followed by a 4 °C per min gradient to 250 °C. Mass spectra were recorded at 70 eV, scanning from 35 to 500 atomic mass units, and experimental samples were compared with standards that were previously used in earlier studies (Wu et al., 2012, Niehaus et al., 2011 and 2012) for verification.

The structure of purified botryococcene from tobacco was determined by ^{1}H-NMR and ^{13}C-NMR spectral analyses, which were also described in an earlier study (Wu et al., 2012). Botryococcene was extracted from leaf material of transgenic line (tpSSL1-3+tpFPS-10) #10 targeting the chimeric botryococcene synthase SSL1-3 and FPS to the plastid compartment under the direction of the constitutive promoters.

One hundred g leaf materials were ground in liquid nitrogen, then extracted with 1.2 l of hexane:ethyl acetate (85:15), the extract concentrated to 5 ml, and the extract fractionated on a silica column with 5 ml aliquots of hexane as the eluting solvent. Fractions were monitored by GC-MS for the desired triterpene compound. Enriched fractions were pooled, concentrated under nitrogen, and the entire sample processed by silica HPLC–PDA using hexane as the eluting solvent (Niehaus et al. 2012, Wu et al., 2012). Alternatively, the crude extract was resuspended in hexane and fractionated via silica gel chromatography, a final purification step provided by HPLC. Recovery of 6 mg of purified botryococcene sample with a 50% yield was obtained. ^{1}H-NMR and ^{13}C-NMR spectra were recorded on a 400-MHz Varian J-NMR spectrometer at 300 K, and chemical shifts were referenced relative to solvent peaks, namely dH 7.24 and dC 77.0 for CDCl3.

Tables and Figures of chapter 3

Table 3.1 Chemical assessments of T0 transgenic lines for their botryococcene content

Constructs	Expression type	# of lines evaluated	# of lines having botryococcene	Young (µg/g)			Mature (µg/g)			# of lines exhibited phenotype
				Min	Max	Ave	Min	Max	Ave	
tpSSL1-3+tpFPS	Constitutive	60	34	0.2	275	91.4	0.9	544	269	25
tpSSL1-3M+tpFPS	Constitutive	75	24	5	110	48	1.1	202	131	19
SSL1-3M+FPS	Constitutive	38	17	0.5	5.4	1.4	1.0	16.4	5.8	0
SSL1-3+FPS	Constitutive	20	6	0.5	7.1	2.6	0.5	6.3	3.5	0
e2tpSSL1-3M+e2tpFPS	Trichome	62	19	0.6	53.8	16.2	1.1	105	32.8	15
e2tpSSL1-3+e2tpFPS	Trichome	37	9	0.9	75	22	3.2	8.5	5.3	6
e2SSL1-3M+e2FPS	Trichome	60	18	0.7	2.3	1.7	0.5	8.1	3.1	0
e2SSL1-3+e2FPS	Trichome	29	5	1.0	6.7	2.1	0.5	14.7	4.1	0

Wild type tobacco (Nicotiana tabacum accession 1068) was transformed with each indicated constructs consisting of chimeric botryococcene synthase gene (SSL1-3 or SSL1-3M) and the avian farnesyl diphosphate (FPS) gene, inserted downstream of constitutive promoters (cassava vein mosaic viral promoter, Pcv; cauliflower mosaic viral promoter, Pca, respectively), or enhanced trichome specific promoters (two 35S enhancers [e2] fused to cembreinene synthase and hydroxlase promoters, respectively). More than 20 independent lines for each indicated construct were generated and grown in the greenhouse condition. The first fully expanded leaf from each plant was sampled for botryococcene content at their young age (1 month old) and mature age (6 month old). The lines were analyzed by GC-FID and the average (ave), as well as the minimum (min) and maximum (max), of those lines accumulating botryococcene are reported. The number of plant lines exhibiting stunting, chlorosis or other developmental abnormalities are reported as the number of lines showing an unusual phenotype.

Table 3.2 Chemical assessment of T0 transgenic lines harboring TMT genes targeted to the chloroplast or the cytoplasm in high triteprene (C30) accumulating lines for their methylated squalene content.

Constructs	Parental line	# of lines evaluated	# of lines having methylated triterpene	triterpene/total (ave)			Conversion: (C31+C32)/total					Total triterpene ave (µg/g)
				C30	C31	C32	# <10%	# 10%-50%	# >50%	highest	ave	
tpTMT-1	tpSQS+tpFPS #5	75	47	34%	25%	41%	2	26	19	91%	65%	222
tpTMT-2	tpSQS+tpFPS #5	69	36	50%	31%	18%	2	17	17	82%	51%	301
TMT-1	tpSQS+tpFPS #5	36	11	96%	0	4%	11	0	0	7%	4%	251
TMT-2	tpSQS+tpFPS #5	33	12	96%	0	4%	12	0	0	6%	4%	241
tpTMT-3	tpSSI1-3M+tpFPS #31	76	18	46%	17%	37%	3	3	12	87%	54%	131
TMT-3	tpSSI1-3M+tpFPS #31	40	19	94%	3%	3%	15	4	0	14%	6%	126
tpTMT-3	tpSSL1-3+tpFPS #10	69	29	65%	10%	25%	0	23	6	66%	35%	283
TMT-3	tpSSL1-3+tpFPS #10	40	17	97%	1%	2%	16	1	0	10%	3%	294
	tpSQS+tpFPS #5	4	0	100%	0	0	4	0	0	0	0	234
	tpSSI1-3+tpFPS #10	3	0	100%	0	0	3	0	0	0	0	237
	tpSSL1-3M+tpFPS #31	3	0	100%	0	0	3	0	0	0	0	125

More than 30 independent lines were generated by each transformation construct consisting one of the three TMT genes targeted to the chloroplast (with tp) or the cytoplasm (without tp) of indicated parental lines: T1 homozygous squalene accumulating lines (tpSQS+tpFPS #5), or homozygous botryococcene accumulating lines (tpSSL1-3+tpFPS #10, or tpSSL1-3M+tpFPS #31) and their triterpene and methylated triterpene content was evaluated in 4 months old plants by GC-FID and GC-MS. The plants accumulating methylated squalene or methylated botryococcene were averaged (ave) based on the percentage of the mono (C31) or dimethyl (C32) methylated triterpene to that of the total triterpene (C30+C31+C32). Percentage of methylated triterpene (C31+C32) to total triterpene was also denoted as conversion. The number of plants with different conversion amounts (less than 10% [<10%], between 10% and 50% [10%- 50%], and above 50% [>50%]) for each engineering effort was analyzed and the highest and average conversions were noted. Average of total triterpene content for each group of lines was also noted. Three or four plants for each parental line accumulating only non-methylated triterpene (C30) were determined for their triterpene content at the same growth stage as the other analyzed transgenic lines.

Table 3.3 Chemical assessment of T0 transgenic lines expressing TMT genes targeting methyltransferase activity to the chloroplast or the cytoplasm of in plants engineered for high-level biosynthesis and accumulation of squalene or botryococcene to the chloroplast compartment.

Constructs	Parental line	# of lines evaluated	# of lines having methyl botryococene	# of lines having methyl squalene	triterpene/total			Conversion					Total squalene ave
					C30	C31	C32	# <10%	# 10%-50%	# >50%	highest	ave	
tpTMT-3	tpSQS+tpFPS #5	45	0	12	95%	4%	1%	12	0	0	20%	5%	122
TMT-3	tpSQS+tpFPS #5	13	0	0	100%	0	0	13	0	0	0	0	173
tpTMT-1	tpSSL13+tpFPS #10	19	0	5	63%	16%	21%	0	4	1	64%	37%	6.5
tpTMT-2	tpSSL13+tpFPS #10	11	0	0	100%	0	0	0	0	0	0	0	4.5

Independent lines were generated for each TMT gene construct by targeting the respective TMT enzyme to the chloroplast (with tp) or the cytoplasm (without tp) of the squalene accumulating parental line (tpSQS+tpFPS #5), or botryococcene accumulating parental line (tpSSL1-3+tpFPS #10). Their triterpene content was determined by GC-MS in plants 2 to 4 months old, and the number of the transgenic plants for each engineered group that accumulate methylated squalene were scored (no methylated botryococcene was found). The average (ave) percentage of each molecular form of triterpene to the total triterpene and the number of plants with different conversion percentages were scored. The highest amount and average (ave) amount of conversion are shown. The number of plants with different conversions for each engineering effort were scored and highest and average of conversion were noted. The average total triterpene content for each group of lines is also noted.

Table 3.4 Chemical assessment of T0 transgenic lines targeting TMT enzyme activity to the chloroplast or the cytoplasm of wild type plants for their methylated squalene contents.

Constructs	Parental line	# of lines evaluated	# of lines having methylsqualene	triterpene/total			Conversion					Total triterpene (ave)
				C30	C31	C32	# <10%	# 10%-50%	# >50%	highest	ave	
tpTMT-1	Wildtpe	9	2	65%	0	35%	0	1	1	61%	41%	8.9
tpTMT-2	Wildtpe	7	0	100%	0	0	0	0	0	0	0	6.3
TMT-1	Wildtpe	14	3	24%	10%	66%	0	0	3	82%	72%	36
TMT-2	Wildtpe	5	2	38%	11%	50%	0	1	1	85%	67%	16.8
	Wildtpe	3	0	100%	0	0	0	0	0	0	0	8.2

Independent lines were generated for each construct consisting of one TMT gene targeted to the chloroplast (with tp) or the cytoplasm (without tp) of wild type plants, and their triterpene and methylated triterpene content was evaluated when the plants were 4-5 months old by GC-MS. The number of transgenic plants accumulating methylated squalene were scored and the percentage of plants accumulating non-methylated (C30), mono (C31) or dimethyl (C32) methylated squalene relative to total squalene (C30+C31+C32) reported. The average total triterpene content for each group of lines is also noted. The squalene (C30) content in three wild type plants was determined at the same growth stage as the transgenic lines.

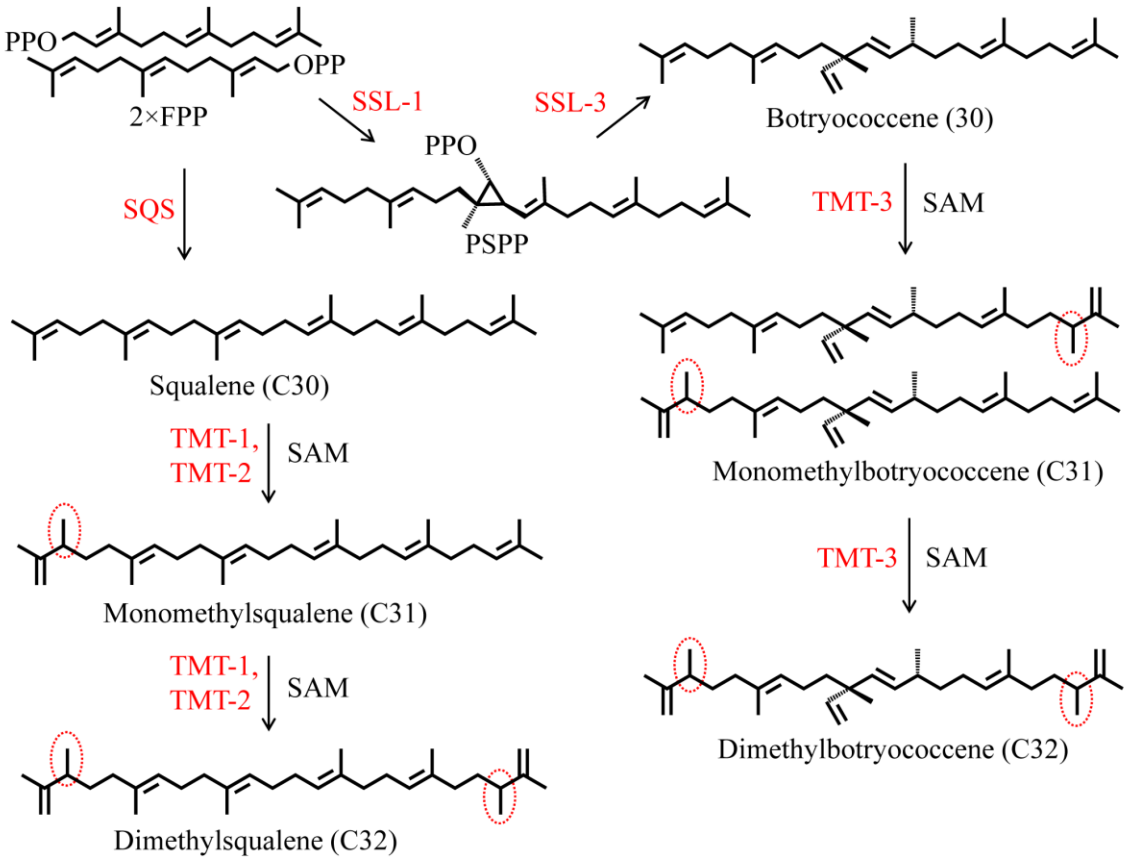


Figure 3.1 A depiction of the catalytic roles of the novel squalene synthase-like enzymes (SSL) and triterpene methyltransferases (TMT) in *Botryococcus braunii* race B and their putative contributions to the triterpene constituents (Niehaus et al., 2011; Niehaus et al., 2012). SSL-1 catalyzes the condensation of two farnesyl diphosphate molecules (FPP) to pre-squalene diphosphate (PSPP), which is converted to either squalene or botryococcene by SSL-2 or SSL-3, respectively. Squalene can also be synthesized from condensation of two FPP molecules catalyzed by squalene synthase (SQS) directly. TMT-1 and TMT-2 transfer the methyl donor group from AdoMet (SAM) to squalene to form mono- or di-methylated squalene, whereas TMT-3 acts on botryococcene to form mono- or di-methylated botryococcene (Niehaus et al., 2012).

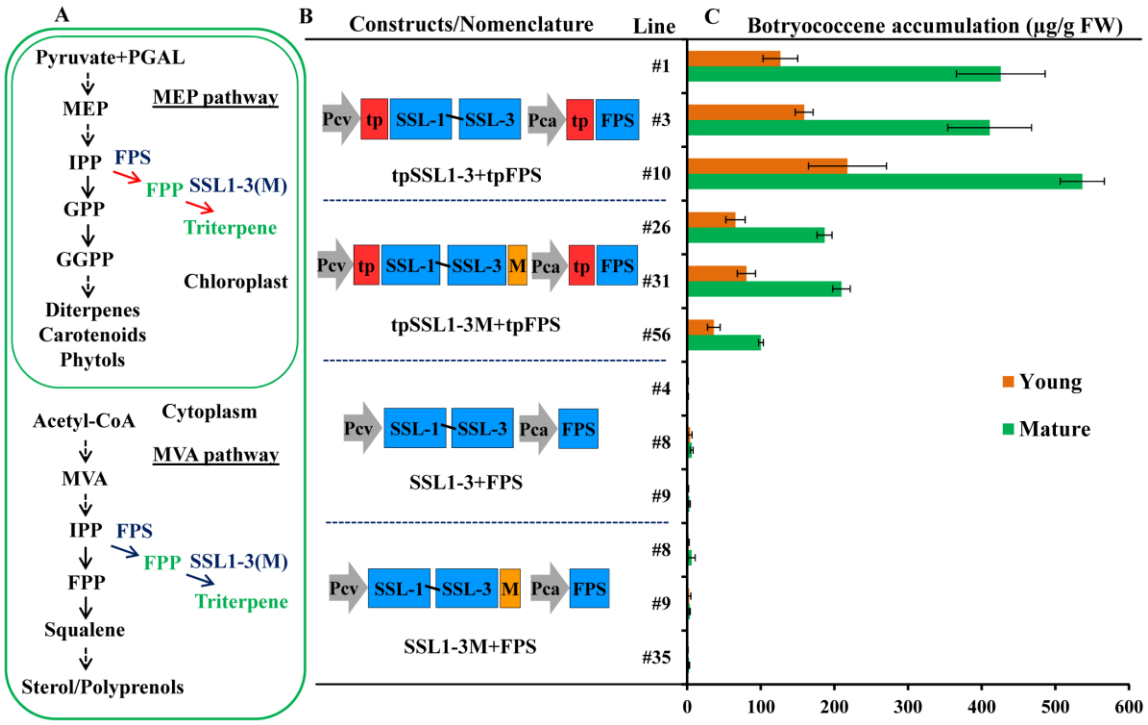


Figure 3.2 Triterpene content of independent T0 transgenic lines transformed for novel botryococcene synthase. Schematic outline of the mevalonate (MVA) and methylerythritol phosphate (MEP) pathways operating in plant cytoplasm and chloroplast compartment, respectively, and the conceptual strategies to divert carbon flux from these two pathways to biosynthesis of novel triterpenes by exogenous terpene synthase. Introduced enzymes are FPS (avian chicken FPP synthase) and SSL1-3(M) (a chimeric version botryococcene synthase with or without a membrane associating domain) (A). Wild type tobacco (*Nicotiana tabacum* accession KY 1068) was transformed with the indicated gene constructs (B), consisting of putative chimeric botryococcene synthase: SSL-1 fused to SSL-3 via a liner peptide with M (membrane associate domain) (SSL1-3M) or without M (SSL1-3) and the FPS gene. Both engineered terpene synthases are under the direction of two strong, constitutive viral promoters (cassava vein mosaic viral promoter, Pcv; cauliflower mosaic viral promoter, Pca, respectively). A plastid targeting signal sequence (tp) was fused to the 5' end of the respective genes. Thus, the constructs with tp will target the enzymes to the chloroplasts, and that without tp will target the enzyme to the cytoplasm. Antibiotic selected T0 lines propagated in the greenhouse were assessed from botryococcene accumulation at a relatively young age (1 month old, red) and mature age (6 month old, green) by GC-MS and GC-FID. Three

independent elite transgenic lines from each engineered construct are chosen to represent their capacity for botryococcene production (C).

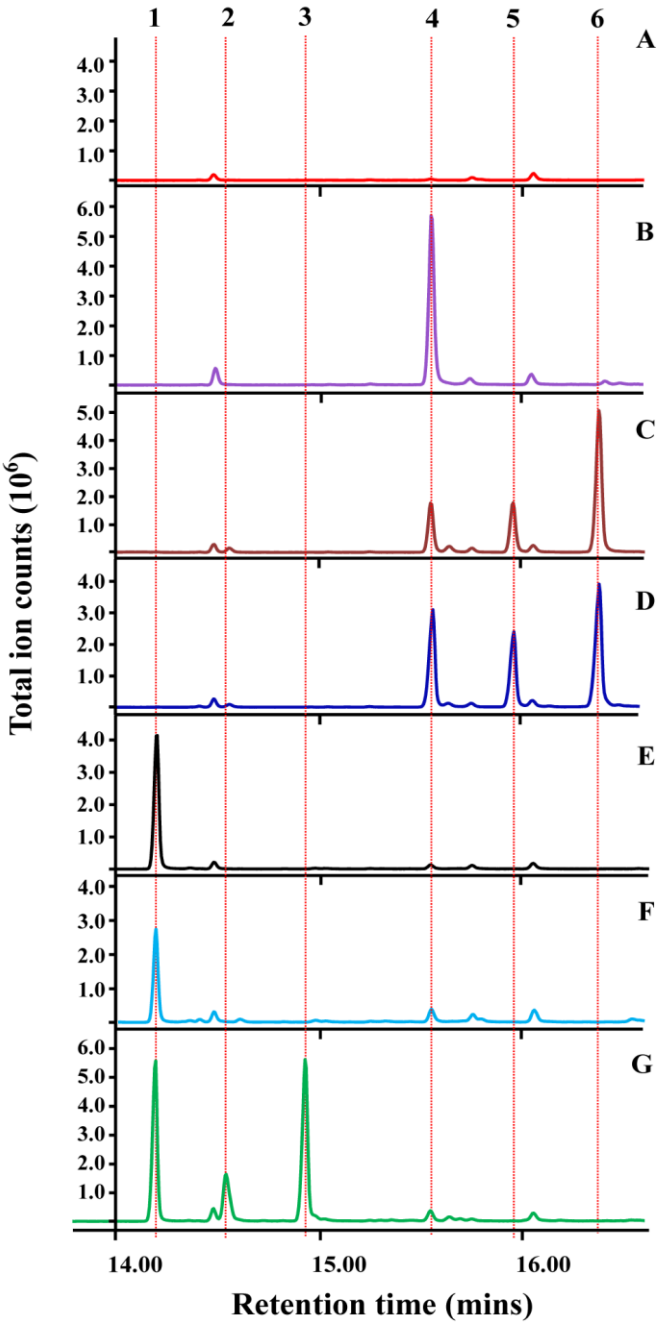


Figure 3.3 Triterpene and methylated triterpene contents were determined in leaf extracts from elite transgenic plants by GC-MS. Independent T0 transgenic lines generated by introducing each construct of tpSQS+tpFPS (B), tpSSL1-3+tpFPS (E) or tpSSL1-3M+tpFPS (F), into wild type plants (A) were propagated in the greenhouse and

their leaf extracts were assessed for triterpene and methylated triterpene accumulation by GC-MS. Select transgenic lines with a high level of squalene (B) were further transformed with tpTMT-1 (C) or tpTMT-2 (D), and lines with a high level of botryococcene (E) were engineered with tpTMT-3 (G). The non-methylated and methylated triterpene contents were determined in the secondly transformed transgenic lines (C, D and G) by GC-MS. The chromatograms are annotated for the elution behavior of (1), C30-botryococcene (C30); (2), C31-botryococcene; (3), C32-botryococcene; (4) C30-squalene; (5) C31-squalene; (6) C32-squalene.

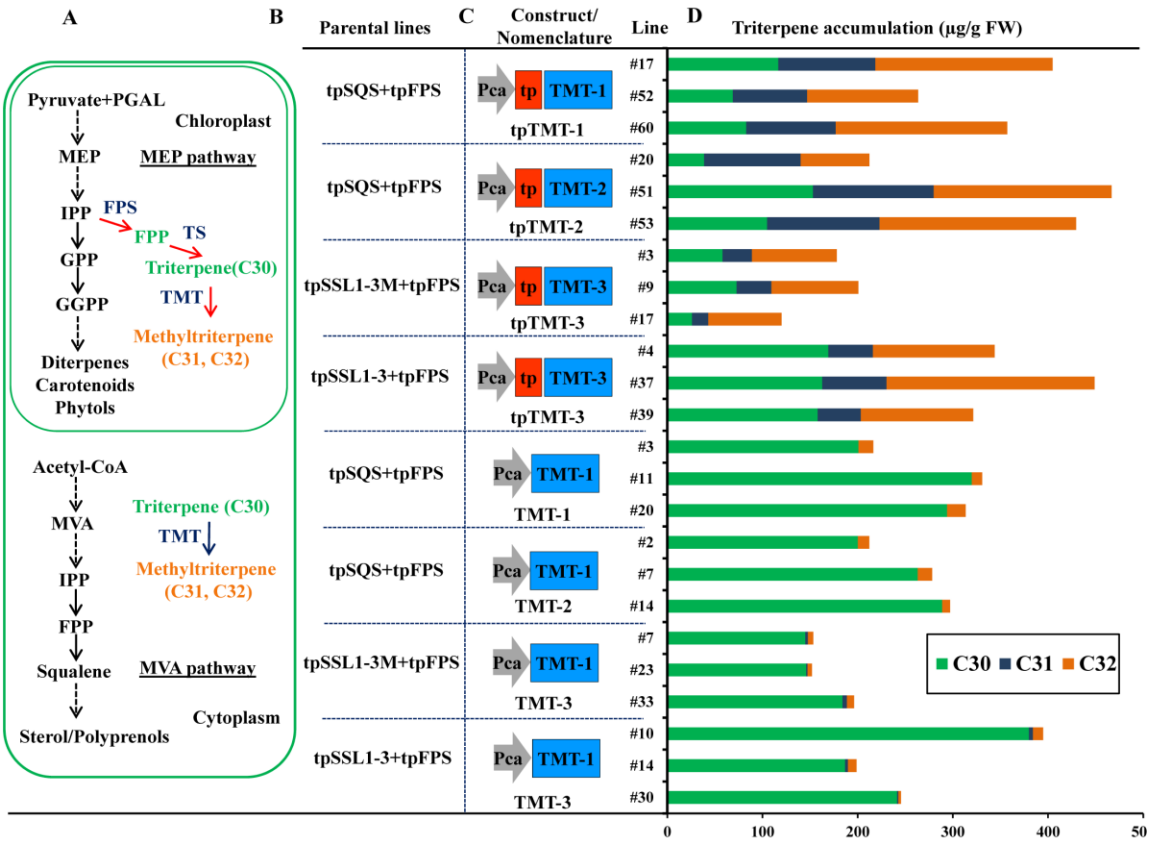


Figure 3.4 Methylated triterpene content in independent transgenic lines expressing TMT genes in particular subcellular compartments of high triterpene accumulating lines. Conceptual strategies to convert triterpenes accumulated in a transgenic plant cell where the carbon flux from MEP pathway was diverted to novel triterpenes biosyntheses by exogenous terpene synthase FPS and TS [SQS or SSL1-3(M)], into methylated triterpenes by cytosolic- or plastidic-targeted TMT are depicted in (A). The gene constructs harboring indicated TMT genes targeting to the chloroplast (with tp) or the cytoplasm (without tp) (C) were transformed into indicated T1 or T2 parental lines accumulating high level of squalene or botryococcene (B). The antibiotic selected T0 lines propagated in the greenhouse for 4 months were assessed from their triterpene and methylated triterpene content by GC-MS and GC-FID. The level of triterpene and methylated triterpene accumulation of three elite independent lines was shown as non-methylated (C30, green) mono- (C31, blue) and di-methylated (C32, orange) triterpene (D).

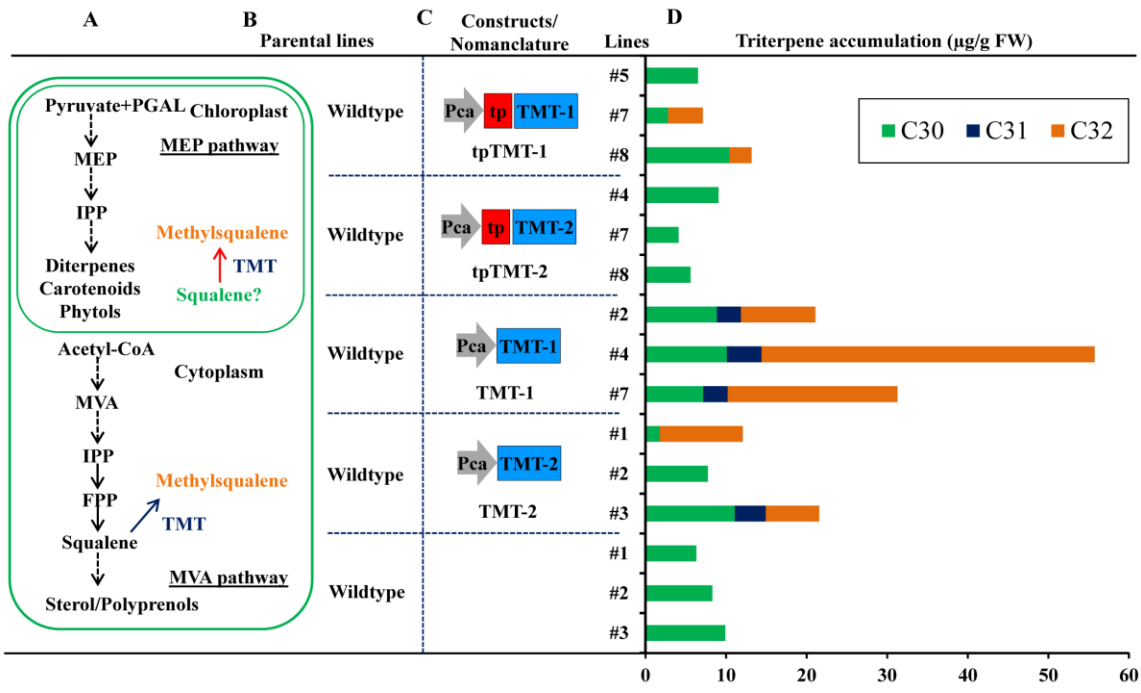


Figure 3.5 Methylated triterpene contents in transgenic lines expressing TMTs genes in wild type plants. The gene constructs harboring indicated TMT genes targeting to the chloroplast (with tp) or the cytoplasm (without tp) (C) were transformed into wild type plants (B). The antibiotic selected T0 lines propagated in the greenhouse were assessed for their triterpene content by GC-MS and GC-FID. The level of squalene and methylated squalene accumulation of three elite independent lines was shown as non-methylated (C30, green) mono- (C31, blue) and dimethyl- (C32, orange) methylated squalene (D).

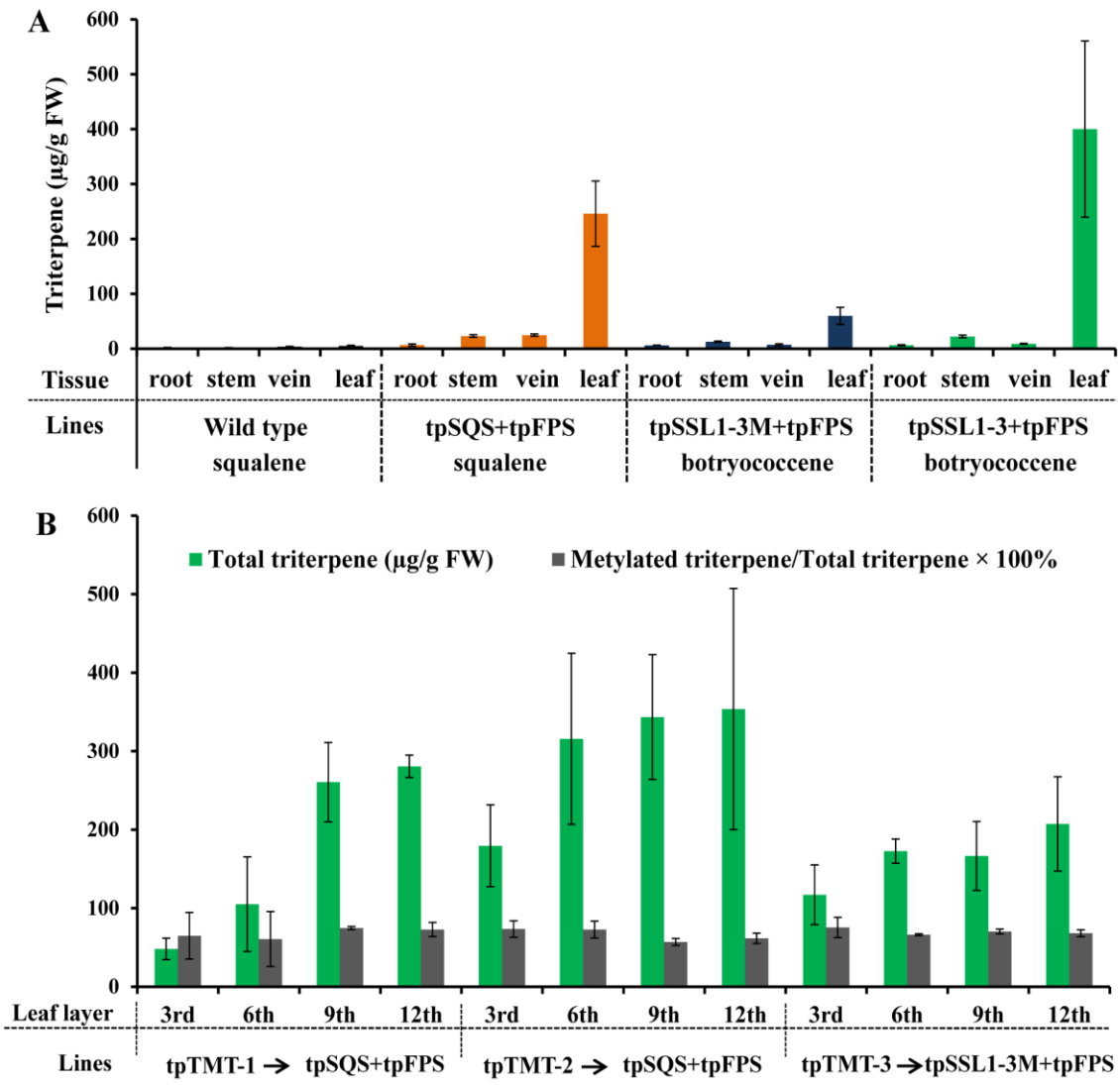


Figure 3.6 Triterpene content in different tissues and leaf positions. Triterpene accumulation in different tissues (A). Select transgenic lines and wild type plants were grown in the tissue culture for 3 months. The plant materials from different tissues (root, stem, vein and leaf) for each indicated line, were sampled for their triterpene content by GC-FID. The type of triterpene (squalene or botryococcene) they accumulated is indicated at the bottom of each column chart. The transgenic lines accumulating methylated triterpenes were grown in greenhouse, and the leaf materials at various leaf positions (counted from top to bottom) from three independent lines expressing the indicated TMT gene targeted to the chloroplast of high triterpene (squalene or botryococcene) accumulating plants, were sampled for their triterpene content after 4 months (B). The level of total tritepene is equivalent to the sum of non-methylated

triterpene (C30) and methylated (C31+C32). The percentage (%) of methylated triterpene (versus total triterpene) is shown.

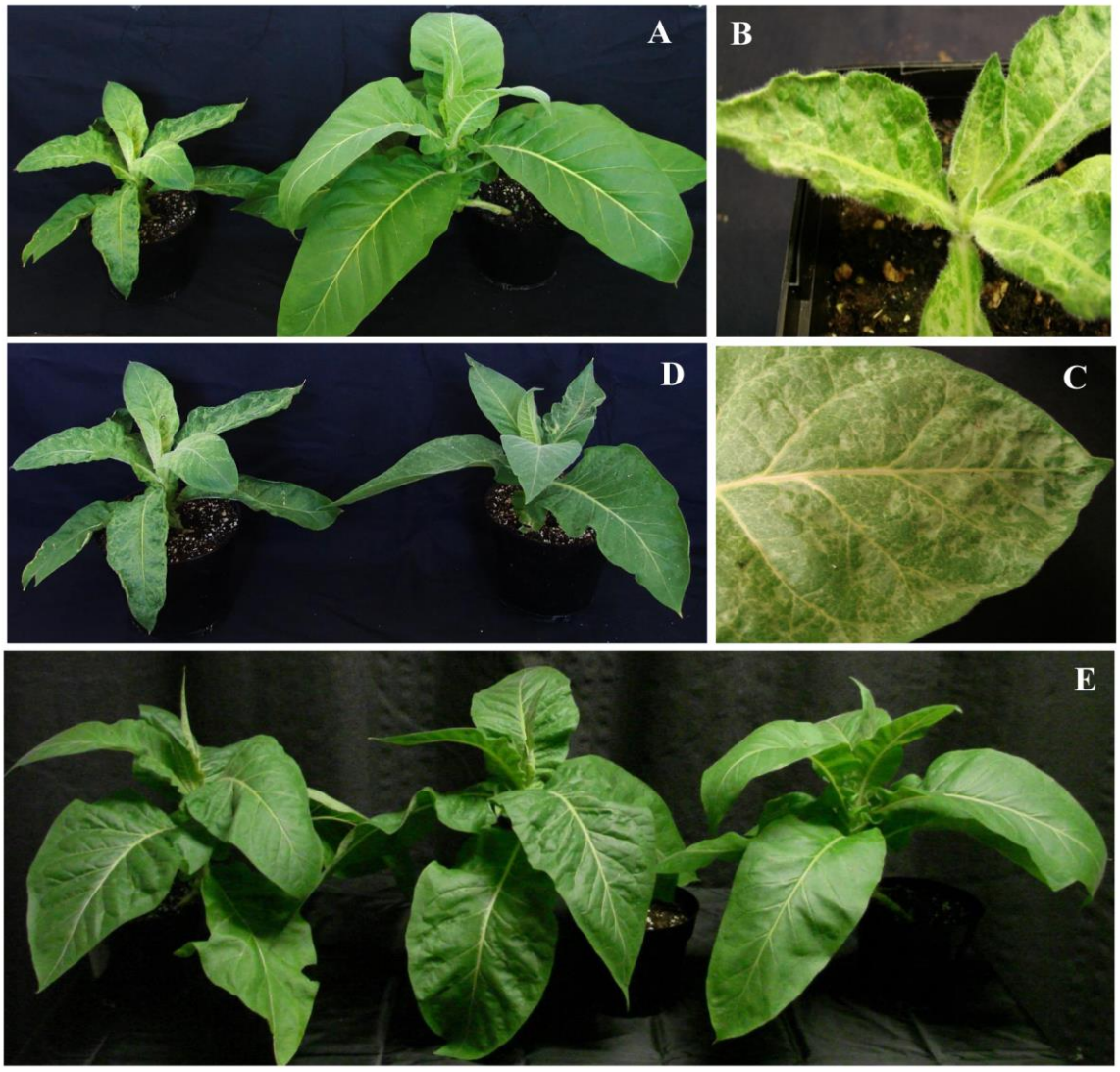


Figure 3.7 Example phenotypes of transgenic lines generated by introducing constructs of tpSSL1-3+tpFPS (A, left [side view]; B, left [top view]) or tpSQS+tpFPS (E, left), into wild type plants (A, right; B, right, E right). The line of tpSSL1-3+tpFPS exhibited unique phenotype of dwarfed, chlorosis, mottling and crinkly leave (C) relative to non-transgenic wild type plants, whereas tpSQS+tpFPS line did not. The transgenic lines expressing the TMT-3 targeted to the chloroplast of line tpSSL1-3+tpFPS (D, right) and lines engineered with TMT-1 or TMT-2 directed to the chloroplast of line tpSQS+tpFPS (E, middle) look comparable to their parental lines (D, right and E, left, respectively).

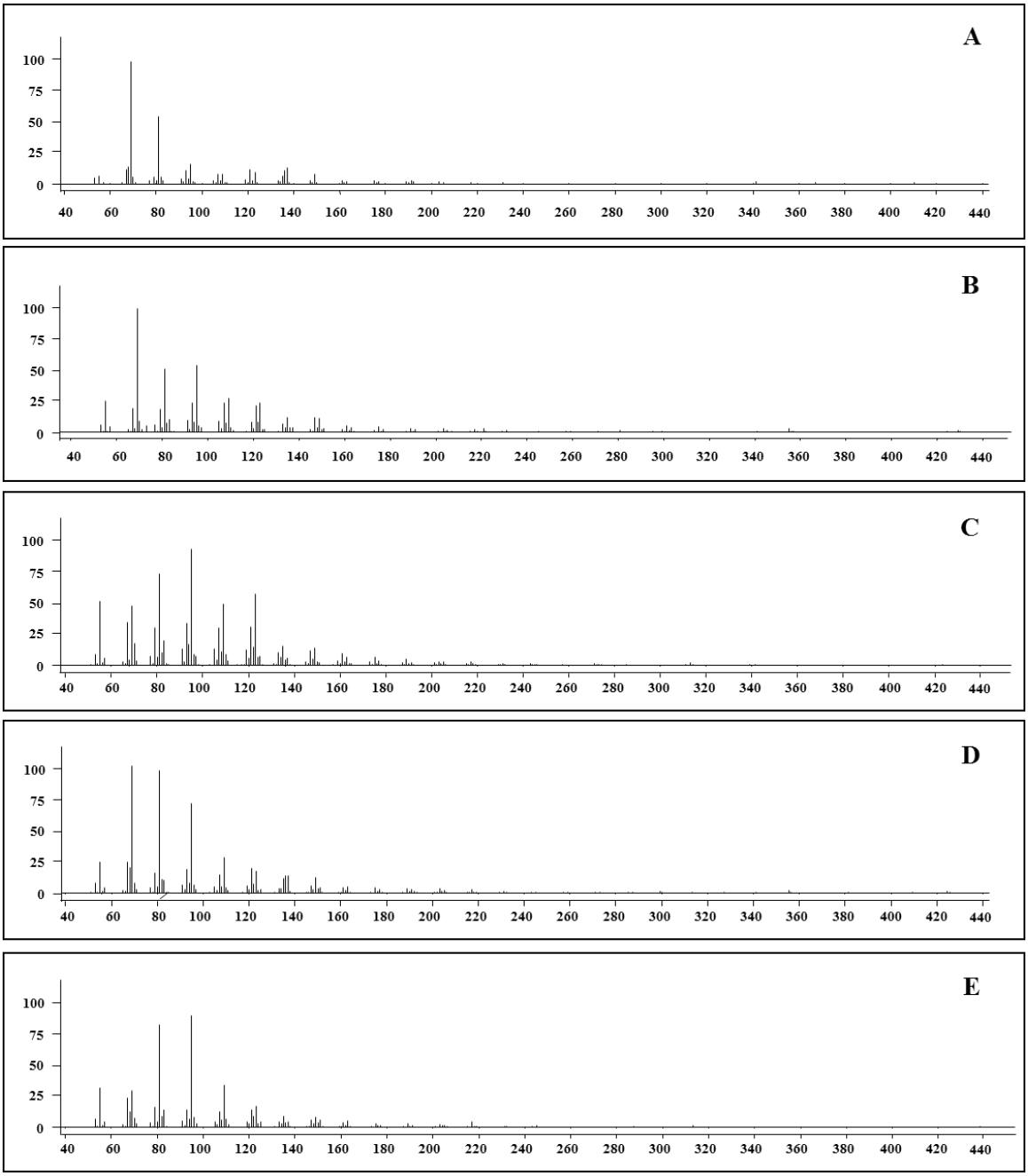
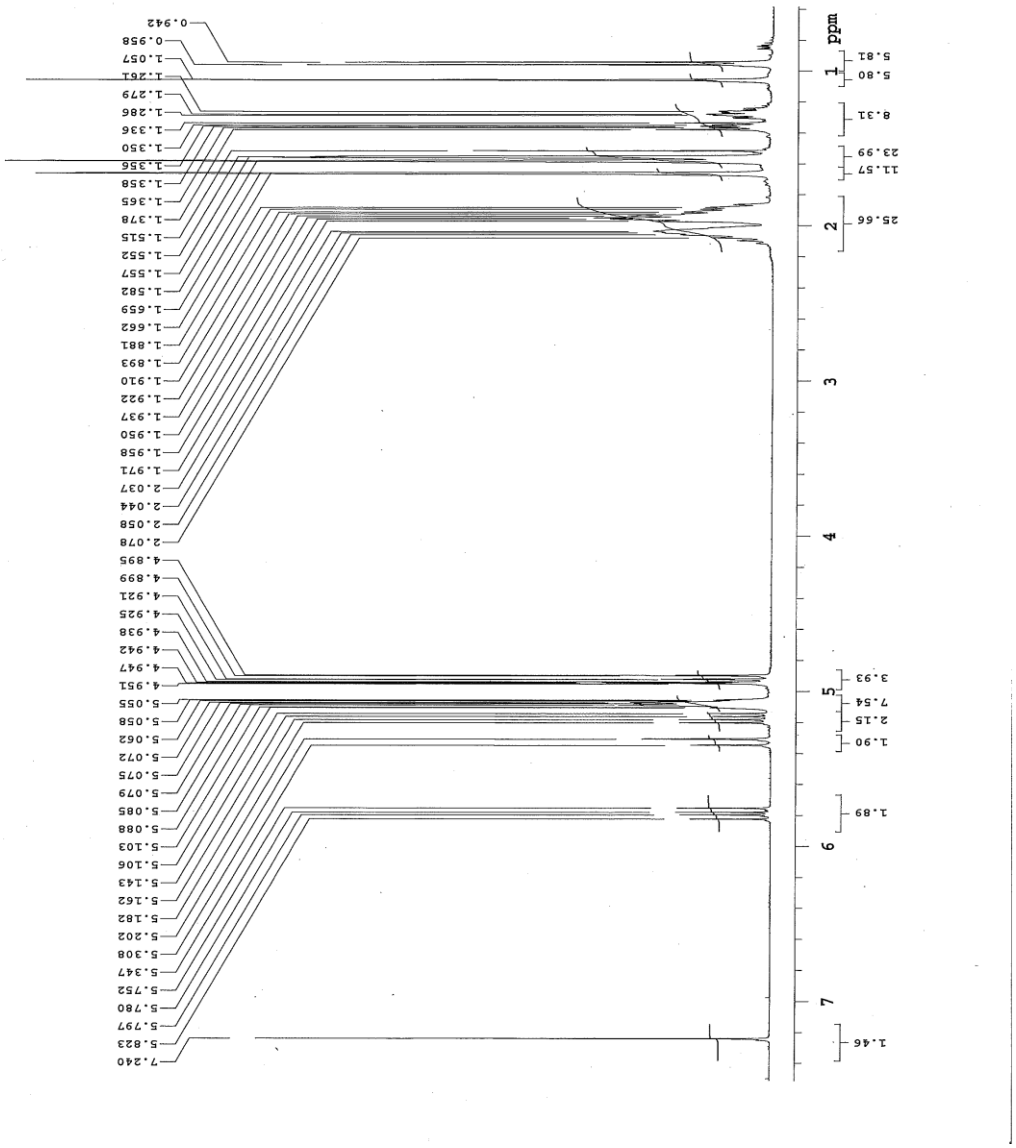
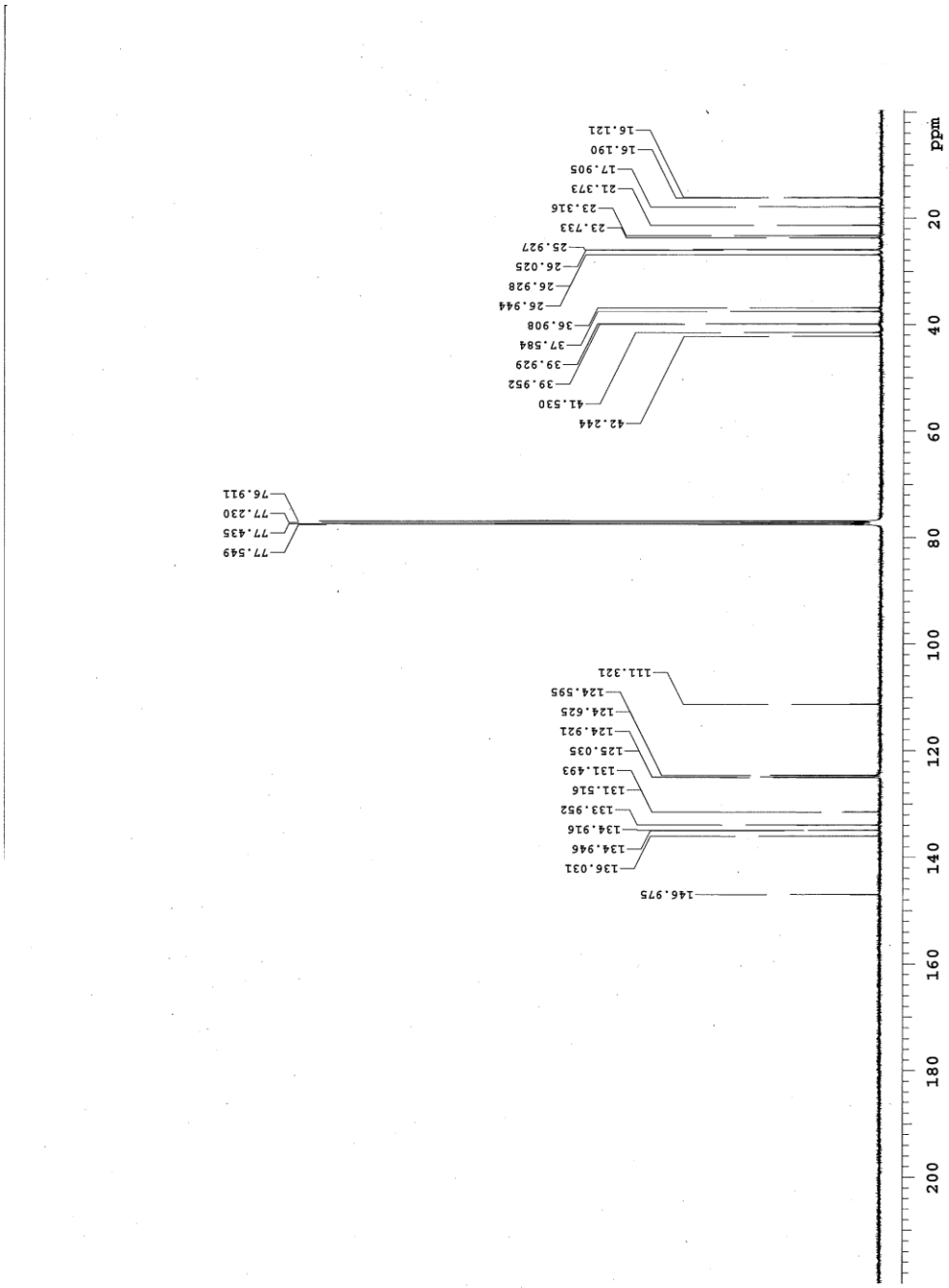


Figure 3.8 Mass spectra of C30 botryococcene (A) produced in transgenic lines expressing SSL1-3 or SSL1-3M in wild type plants, C31 botryococcene (B) and C32 botryococcene (C) produced in lines expressing TMT-3 in botryococcene accumulating line, C31 squalene (D) and C32 squalene (E) produced in lines expressing TMT-1, TMT-2 and TMT-3 in squalene accumulating lines.

Figure 3.9 (6E,10R,11E,13R,16E)-botryococcene (C30 botryococcene). Colorless oil. GC-MS mass: 410.5 amu (M+). C30H50. 1H-NMR (400 MHz) δH 5.82 (1H, dd), δH 5.33 (1H, dd), δH 5.20 (1H, dd), δH 5.05-5.11 (4H, m), δH 4.95 (1H, dd), δH 4.94 (1H, dd), δH 1.89-2.08 (13H, m), δH 1.66 (6H, s), δH 1.52-1.58 (12H, m), δH 1.26-1.38 (4H, m), δH 1.06 (3H, s) δH 0.94 (3H, d, J=7 Hz). 13C-NMR (100 MHz) δC 146.98, δC 136.03, δC 134.95, δC 134.92, δC 133.95, δC 131.52, δC 131.49, δC 125.04, δC 124.92, δC 124.63, δC 124.60, δC 111.32, δC 42.24, δC 41.53, δC 39.95, δC 39.93, δC 37.58, δC 39.91, δC 26.94, δC 26.93, δC 26.03, δC 25.93, δC 25.93, δC 23.73, δC 23.32, δC 21.37, δC 17.91, δC 17.91, δC 16.19, δC 16.12.



[1]H-NMR spectrum of isolated squalene produced in *planta*. (400 MHz, CDCl3).



^{13}C-NMR spectrum of isolated squalene produced in *planta*. (400 MHz, CDCl3).

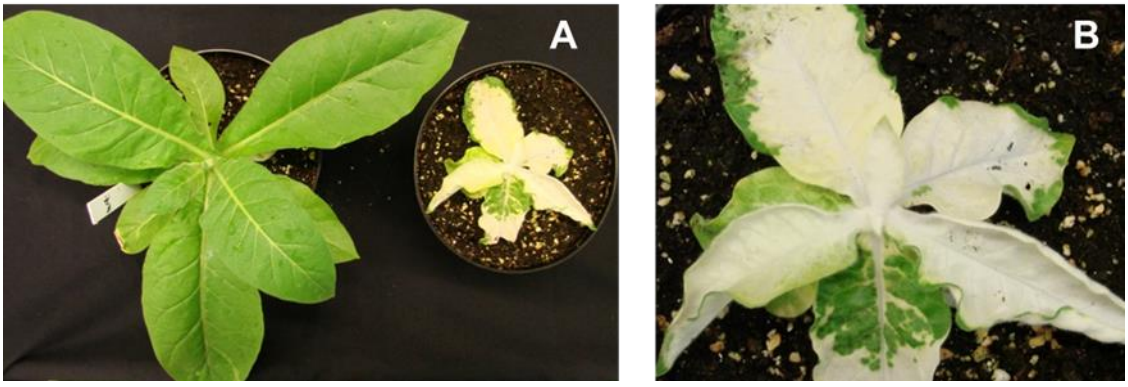


Figure 3.10 Transgenic lines generated by trichome specific expression of SSL1-3 and FPS directed to the chloroplast exhibit a distinct phenotype. Wild type plants (A, left) in comparison to transgenic lines (right) exhibit a strong chlorosis, mottling, dwarfed phenotype (B).

Chapter 4: Agronomic and chemical analyses of field grown transgenic tobacco engineered for triterpene and methylated triterpene metabolism

4.1 Summary

Squalene is a key, linear intermediate in isoprenoid metabolism to all classes of triterpenes and sterol biosynthesis and is itself highly valued for its use in cosmetics and industry. Another unique linear triterpene is botryococcene and its methylated derivatives generated by algae *Botryococcus braunii* race B. These are considered important as an alternative biofuel oils. In previous efforts, these high value triterpenes were successfully engineered into transgenic tobacco by introducing the key steps of triterpene metabolism into the particular subcellular compartments. In this study, the agronomic characteristics (height, biomass accumulation, leaf area), the photosynthetic capacity (photosynthesis rate, conductance, internal CO2 levels) and triterpene content of select triterpene accumulating lines grown in the field were measured and evaluated for three consecutive growing seasons. We observed that transgenic lines targeting the enzymes to the chloroplasts accumulates squalene about 50-150 fold greater than the lines expressing the enzymes in the cytoplasm and non-transgenic control lines, but the growth of some high accumulators was only slightly compromised and their photosynthesis rates were not affected. We also found that the transgenic lines directing the botryococcene metabolism to the chloroplast accumulates botryococccene 10-26 fold greater than the lines where the same enzymes were targeted to in the cytoplasm. But growth of these high accumulators was highly compromised while their photosynthesis rates remained unaffected. In addition, in the transgenic lines targeting the TMT to the chloroplasts of high squalene accumulators, 55%-65% of total squalene was methylated, whereas in the lines expressing TMT in the cytoplasm only 6% of squalene was methylated. The growth of these methylated triterpene accumulating lines was even more compromised than that of non-methylated squalene lines.

4.2 Background and introduction

Tobacco is an annually–grown herbaceous plant that produces over 2500 compounds including terpenes, alkaloids, flavonoids and anthocyanins (Nugroho and Verpoorte, 2002). Tobacco has been considered primarily as cash-crop grown for it tobacco use products that has abundant amount of nicotine, nornicotine, anabasine, and anatabine,

thought important for its consumer preference. In most commercial tobacco varieties, nicotine represents 90–95% of the total alkaloid content of the leaf (Siminszky et al., 2005). Tobacco has received much more attention recently because genetically enhanced tobacco has been suggested as an alternative platform for pharmaceuticals and biofuel production. For example, plant derived vaccines are expected to solve the vaccines shortage and be marketed in the near future (Yusibov et al., 2011). As an industrial biomass crop, it can generate up to 170 tons／ha of green tissues when grown for biomass production (Andrianov et al., 2010).

Engineering high level production of terpenes in tobacco has been achieved using a novel strategy that has been successfully applied in generating large amounts of sesquiterpene, monoterpene and triterpene products (Table 4.1). This strategy relies on the diversion of carbon flux from MEP pathway by over-expression and targeting an avian farnesyl diphosphate (FPP) synthase (FPS) gene with a heterologous terpene synthase gene in the chloroplasts (Figure 4.1). Engineering FPS to create a plastidic FPP pool was key to the success of this strategy because FPP biosynthesis is lacking in the plastidic organelle and hence is not subject to any known innate regulation in the chloroplasts (Kappers et al., 2005). In contrast, cytosolic biosynthesis of FPP is highly regulated by transcriptional or posttranscriptional mechanisms operating in the cytoplasm (Janowski et al., 1996; Gardner and Hampton, 1999). Thus, the strategy derives from the putative unlimited supply of IPP/DMAPP that can be diverted from the MEP pathway by the action of FPS to yield novel pools of plastidic FPP. These plastidic pools could then be utilized by FPP dependent terpene synthase targeted to the chloroplasts to result in high level production of the novel terpene(s).

This strategy makes it possible for large-scale production of highly valued terpene products, such as the triterpene and their methylated derivatives generated by the green algae *Botryococcus braunii*, race B., which is considered as an alternative biofuel oil (Niehaus et al., 2011). Use of tobacco as a platform for this biofuel production is promising because it can utilize photosynthesis to directly convert solar energy and CO_2 into energy–abundant liquid fuels. Engineering transgenic tobacco for production of high level of squalene (C30), botryococcene (C30), and their methylated derivatives (C31–C32) were described in the previous chapter (Wu et al., 2012; Jiang et al., unpubliched).

However, how much impact such biochemical/metabolic engineering might have on the overall physiology and how much yield of oil could be achieved in the field is yet unknown. Therefore, to gain a better appreciation for the robustness of the triterpene accumulation trait and its impact on overall growth performance, multiple transgenic lines were grown in field trials from late May to early September from 2012 to 2014 and evaluated for their agronomic characteristcs and triterpene content.

4.3 Results

4.3.1 Overall growth characteristics of transgenic tobacco

Agronomic characteristics (height, biomass accumulation, leaf area) of selecttransgenic lines expressing plastidic or cytosolic enzymes under the direction of the trichome-specific or constitutive promoters (Table 4.3, 4.4, and 4.5) were monitored throughout a typical field season of approximately 90 days, but the final measurements at the termination of the field are used here to simplify the comparisons. Likewise, measurements of CO_2 fixation (photosynthesis rate), transpiration (conductance) and internal levels (Ci) were also taken at various times during the field trials, but the final determination at ~60 days after planting are used here to compare physiological measurements of fitness.

All these growth characteristics were significantly greater for the 2013 field study versus 2012 and 2014 (Figure 4.2). For instance, the wild type plants grown in 2013 exhibited height, biomass (leaf weight and total weight) and leaf areas measurements 1.5 to 2-times greater than that in 2012 and 2014. Rainfall from late May to early September (especially the June) in 2013 and 2014 was significant greater than that during the same period of 2012 (Table 4.2), which that might account for some of these overall difference.

4.3.2 Agronomic performance of transgenic tobacco engineered for constitutive squalene biosynthesis and accumulation

Agronomic performance of select T2 homozygous transgenic lines constitutively targeting FPS and squalene synthase to the chloroplast compartment (G1, G8, I8 and H5) were compared to that for a plant line constitutively expressing the transgenes targeted to the cytoplasm (C5), as well as the control, non-transgenic parent line of tobacco, cultivar 1068 (Figure 4.2 and table 4.6). The H5 line, which has a significant

level of squalene accumulation (Figure 4.8), only grew to about 82 and 63% the height, 73 and 46 % the total weight, 71 and 38 % the leaf weight, 79 and 51% the leaf area of that for wild type (WT) plants in 2012 and 2013, respectively. In contrast, the G1 line, accumulating the highest level of squalene among all the transgenic lines, exhibited only a slight decrease in growth, which was about 96 and 90% the height, 99 and 98% the total weight, 97 and 89% the leaf weight, 110 and 87% the leaf area of that for WT plants in 2012 and 2013 respectively (Figure 4.2). This indicates that squalene accumulation does not necessarily correlate with a decrease in agronomic performance for these high squalene accumulating lines. The growth reduction of the H5 line is perhaps due to some additional genetic changes resulting from the transformation process, which might include so-called non–specific position effects (ectopic effect) of where the transgenes integrated into the genomic DNA of the plant. Line C5 which accumulates only a low level of squalene (Figure 4.8) similar to WT plants, also exhibited a pronounced decrease of overall performance of 62% and 56% total weight, 65% and 62% leaf weight, 68% and 60% leaf area of that for WT plants in 2012 and 2013, respectively (Table 4.6 A and B). The other two high squalene accumulating lines G8 and I8 showed a modest growth reduction, less than that observed for H5 but more than that for G1 (Table 4.6 A and B). Therefore although all four independent high accumulators were derived from the same constructs, their growth characteristics are far from identical, indicating that insertional effects have a bigger impact on plant growth than squalene accumulation. This notion was supported by line G1, which accumulates the highest levels of squalene but whose growth performance is directly comparable to the WT plants, indicating that agronomic performance of the transgenic lines accumulating squalene were not necessarily compromised.

4.3.3 Agronomic performance of transgenic tobacco engineered for squalene biosynthesis and accumulation by trichomes

A different trend in growth reduction was observed in the transgenic lines engineered for trichome specific squalene biosynthesis and accumulation. The line R13, which was grown as a heterozygous, segregating population of plants, was derived by transformation with a construct targeting squalene biosynthesis to the chloroplasts of trichomes. Those plants determined to accumulate high level of squalene (Figure 4.8) were considered to have inherited the transgenes and denoted as R13+, while those plants having squalene levels comparable to WT were considered to be siblings having

lost the transgene during segregation and denoted as R13-. The R13+ plants showed the most dramatic decrease of overall growth compared with WT plants (about 70% and 81% the height, 53% and 56% the total weight, 51% and 51% the leaf weight, 63% and 65% the leaf area of that for WT plants in 2012 and 2013 respectively) (Figure 4.2 and Table 4.6A and B). The R13+ plants exhibited stunted growth and chlorotic symptoms, which were previously reported when these plants were grown under greenhouse conditions (Wu et al., 2012). The homozygous sibling line R6 exhibited a similar growth reduction. But the R13– plants that presumably lost transgene during segregation performed equally to the WT, which is taken as evidence that the phenotypic consequences are directly attributable to the transgene construct (Figure 4.2 and Table 4.6 A and B).

In contrast, the lines targeting only SQS (A7) or SQS with FPS to the cytoplasm (N10), or only SQS to the chloroplast (D4) had squalene levels as low as WT plants and showed only a slight decrease, if any, in growth performance relative to the WT control. For instance, A7 had about 110 and 100% the height, 89 and 75% the total weight, 94 % and 78% the leaf weight, 93% and 77% the leaf area of that for WT plants in 2012 and 2013 respectively; D4 had about 112% and 98% the height, 113% and 85% the total weight,117% and 88% the leaf weight, 115% and 86% the leaf area of that for WT plants in 2012 and 2013 respectively; and N10 had about 102% and 99% the height, 85% and 73% the total weight, 87% and 75% the leaf weight, 83% and 70% the leaf area of that for WT plants in 2012 and 2013 respectively. These results indicate that targeting trichome specific expression of enzymes to the cytoplasm, or engineering only a partial pathway for squalene biosynthesis to the chloroplast, results in only low level accumulation of squalene (Figure 4.8) without an effect on growth performance.

4.3.4 The photosynthetic capacity of transgenic lines engineered by squalene metabolism.

In order to test if engineering squalene metabolism to the chloroplasts or cytoplasm has any impact on photosynthetic capacity in these transgenic lines, CO_2 fixation rates (photosynthesis rate), transpiration (conductance) and internal CO_2 levels (Ci) were measured at about 60 days after planting. Overall, most of the transgenic lines accumulating variable levels of squalene, via plastidic or cytosolic targeting under the direction of constitutive or trichome-specific promoters, showed higher conductance (a

measure of water movement and transpiration) than WT plants, but did not exhibit a significant difference in photosynthesis or their ability to concentrate CO_2 (Ci) compared with WT. The exceptions were lines H5, R13 and R6 that exhibited a modest decrease in their photosynthesis rate, a more pronounced increase in conductance and a slightly increase in Ci compared to WT plants over two growing seasons (Figure 4.3). Altered photosynthetic and gas exchange capacity in these lines could possibly be correlated with their growth reduction.

The inhibition of photosynthesis in these three lines is unlikely to be the result of squalene accumulation, because the likewise developed line G1 had the highest level of squalene accumulation without an obvious impact on its photosynthetic capacity (Figure 4.3). Hence, the reason for their reduction in photosynthesis could vary on a case-by-case basis. For instance, the H5 line targeting the squalene synthase (SQS) and farnesyl diphosphate synthase (FPS) enzymes to the chloroplasts directed by constitutive promoters may have had genes related to photosynthesis disrupted by the inserted transgenes by some ectopic effect or genetic changes resulting from the transformation/regeneration protocol. For the lines R6 and R13, their reduction in photosynthesis could be caused by some unique mechanism arising from the chimeric trichome specific promoters expressing the transgenes during a crucial time period during development of photosynthetically active cells. However, squalene accumulation in the various transgenic lines does not appear to correlate with altered photosynthetic capacity. Regardless, the measured higher conductance rates in most of transgenic lines indicate that they might be more drought sensitive than wild type plants.

4.3.5 Agronomic performance of transgenic tobacco engineered for botryococcene accumulation

The agronomic characteristics and photosynthesis of T1 transgenic lines expressing botryococcene synthase SSL1-3M and FPS targeted to the chloroplast (line td26) or the cytoplasm (line 13M), and those lines targeting SSL1-3 and FPS to the chloroplast (line tc10 and tc3) or the cytoplasm (line 136) were measured in 2013 and 2014. Because these lines were planted as segregating populations, plants accumulating botryococcene were considered to harbor the transgene (denoted as “+”), while those that did not were considered to have lost the transgene during segregation were denoted as “–”. Of the non-botryococcene accumulating plants, lines tc10–, tc3–, td26–, 13M–, 136–, most of

these plants performed directly comparable to the wild type plants with regard to the agronomic characteristics (height, biomass accumulation, leaf area) and photosynthetic measurements (CO_2 fixation rates, transpiration and internal CO_2 levels). In contrast, for the lines accumulating high level of botrococcene (tc10+, tc3+ and td26+), their overall agronomic performance was significantly reduced compared with WT (Figure 4.4 and table 4.7). For example, tc10+ with the highest accumulation of botryococcene was only 74 and 74% the height, 65 and 68% the total weight, 58 and 66% the leaf weight, and 68 and 58% the leaf area of that for WT plants in 2013 and 2014, respectively. These plants also exhibited a higher conductance (1.9 and 1.2 fold increase relative to WT plants in 2013 and 2014, respectively), and a slightly higher Ci (1.2 and 1.1 fold increase relative to WT plants in 2013 and 2014, respectively) (Table 4.7). Similar levels of reduction in growth and increases in conductance and Ci were also observed in other high botryococcene accumulating lines like td26 and tc3. However, the lines 136+ and 13M+ expressing the enzymes in the cytoplasm, accumulating a low level of botryococcene performed directly comparable to the WT plants with regard to overall growth and photosynthesis. Interestingly, all the high accumulating lines showed a crinkle, mottling and chlorosis leaf phenotype (Figure 4.12), whereas, the non-accumulators or low accumulators did not (Figure 4.12). Altogether, plant growth of transgenic lines with high levels of botryococcene was significantly compromised, but photosynthesis rates were not affected.

4.3.6 Agronomic performance of transgenic tobacco engineered for methylated triterpene accumulation

To test if engineering methylation of the accumulating triterpene in the transgenic lines would have any impact on plant growth and photosynthesis, the T1 transgenic lines expressing triterpene methyl transferase (TMT) genes targeted to the cytoplasm or chloroplast of respective high squalene or botryococcene accumulators were also evaluated in field performance assays. These studies were complemented by using different TMTs. TMT-1 and TMT-2 were previously documented to selectively methylate squalene, while TMT-3 was described as having much greater specificity for botryococcene. Those plants for each line that were determined to have methylated triterpenes were considered to have inherited the TMT transgene and denoted as +, while those that did not accumulate methylated triterpenes were considered to have lost the TMT gene during segregation and are denoted as –. The lines accumulating high

level of methylated squalene by targeting the enzymes to the chloroplast (tpT1+ and tpT2+), showed a significant growth reduction relative to that for WT plants, about 73 and 70% the height, 59 and 65% the leaf weight, 57 and 61% the total weight, 65 and 70% the leaf area of that for WT plants, respectively (Figure 4.6 and table 4.8A). These two high methylated triterpene accumulating lines also showed a growth reduction relative to their sibling lines (tpT1–, tpT2–) only accumulating non–methylated squalene, exhibiting about 94 and 87% the height , 69 and 73% the leaf weight, 72 and 71% the total weight, 86 and 80% the leaf area of that for tpT1– and tpT2–, respectively (Table 4.8B). However, the lines expressing the TMT–1 in the cytoplasm (T1G+) with low level of methylated squalene performed almost equally to its sibling lines (T1G–), accumulating only non–methylated squalene (Table 4.8B), although both lines showed a modest decrease of growth compared to WT plants (70 and 68% height, 81 and 90% the total weight, 75 and 83% the leaf weight, 90 and 95% the leaf area of that for WT plants, respectively) (Table 4.8A).

Similar trends in growth reduction were also observed in the transgenic lines engineered for methylated botryococcocenes. The high level methylated botryococcene accumulators tpT3tc+ and tpT3td+ expressed TMT-3 targeted to the chloroplast showed 68 and 64% the height, 61 and 84% the leaf weight, 61 and 77% the total weight, 81 and 99% the leaf area of that for WT plants, respectively. Besides, these two high level methylated botryococcene accumulators also showed a significant growth reduction relative to their sibling lines where the TMT-3 transgene had segregated out and no methylated botryococcenes was evident. tpT3tc+ and tpT3td+ exhibited about 98 and 89% the height, 68 and 89% the total weight, 70 and 86% the leaf weight, 95 and 93% the leaf area of that for tpT3tc– and tpT3td–, respectively (Table 4.8B). The lines targeting TMT-3 to the cytoplasm (T3tc+ and T3td+), which accumulated only very low levels of methylated botryococcenes, were also reduced in their growth relative to the wildtype plants, but only slightly decreased relative to their sibling lines (T3tc– and T3td–) accumulating only non-methylated botryococcenes (Table 4.8A and B).

Photosynthesis in most of the transgenic lines engineered for methylated botryococcene production was seemingly affected, expect for lines tpT3td+ and T3td+ in comparison to wildtype plants (Figure 4.7). Similar to the squalene and methylated squalene accumulating plants, most of methylated botryococcene accumulating lines (tpT1+, tpT2+, tpT3tc+, tpT3td+, T1G+, T3tc+,T3td+) showed a slightly higher conductance and

Ci than their respective sibling line (tpT1–, tpT2–, tpT3tc–, tpT3td–, T1G–, T3tc–,T3td–) that accumulated only non-methylated triterpene, or to WT plants (Figure 4.7).

These results altogether indicated that that the lines engineered for methylated triterpene production were modestly impacted in growth performance relative to wildtype plants, but did show a slightly greater decrease in growth relative to the non-methylated triterpene accumulating lines.

4.3.7 Development dependent accumulation of squalene over three growing seasons

The triterpene content was determined for the uppermost, fully expanded leaf of 6 to 9 plants for each transgenic line grown in three replicate rows at different growing stages: early (25 days), middle (45-50 days) and late stage (80-100 days, after topping) for three growing seasons. Squalene content for each transgenic line targeting squalene biosynthesis to the chloroplasts or the cytoplasm under the direction of constitutive or trichome-specific promoters was determined by GC-FID and GC-MS.

By comparing the squalene level in each line at different growing stages, we found that squalene accumulated in a developmental-dependent manner, especially for most of the high squalene accumulators such as G1. Squalene levels in middle to late stage of leaf development were significantly higher than measured in early stages of development (Figure 4.8). However, unlike some tobacco specific metabolites such as nitrosamines and nicotine that accumulate predominantly during the “ripening” stage (after apical meristem topping), squalene seemed to accumulate to the highest levels during the middle stages of leaf development. Transgenic line G1, for instance, accumulated the highest level of squalene, upwards of 700 µg/g fresh weight, in the middle stages of leaf development when averaged over the 3 years of field testing (Figure 4.8). A similar pattern of squalene accumulation was also evident in other lines, albeit the absolute amounts accumulating were not nearly as high, such as for H5 (or its silbing H9) grown for all three growing seasons, and for G8 and I8 grown in 2012 and 2014 (Figure 4.8).

Among all the transgenic lines under the direction of constitutive promoters, G1 accumulated squalene 95-, 48-, and 61- fold higher than that found in wild type plants growing at the same time in 2012, 2013 and 2014 respectively, and 87-, 150-, and 62-

fold for squalene accumulation by C5 (or its sibling line C12) that targeted squalene biosynthesis to the cytoplasm. Other independent T2 lines engineered with the same constructs, such as I8 and G8, also yielded a higher level of squalene accumulation, which is significantly higher than WT and cytosolic engineered lines (Figure 4.8). Except for the H5 line, other high accumulators did not show any significant growth or biomass reduction relative to wildtype plants, indicating that their growth was not significantly compromised by squalene accumulation, which makes them potential candidates for large-scale application.

Among the lines under direction of the trichome specific promoters, R13 and its homozygous sibling R6 expressing SQS with FPS in the chloroplasts accumulates a significant higher level of squalene than other lines. For example, the squalene level for R13+ in the middle stage is 74 and 128 fold greater than that for N10 that expressed SQS with FPS in the cytoplasm, 71 and 37 fold greater than that for A7 line expressing only SQS in the cytoplasm, but 22 and 21 fold greater than that for D4 line that expresses only SQS in the chloroplasts, in 2012 and 2013 respectively (Figure 4.8). However, when considering intense growth reduction for these high accumulators under direction of trichome specific promoters, they may not be suitable for large-scale production. Overall, these results indicate transgenic lines engineered for squalene metabolism in the chloroplast accumulate significantly higher levels of squalene than the lines engineered for cytosolic squalene biosynthesis, and the high squalene accumulating lines exhibited a developmental-dependent accumulation pattern when grown under field conditions.

4.3.8 Development dependent accumulation of botryococcene over two growing seasons

T1 transgenic lines expressing SSL1-3M with FPS targeted to the chloroplast (td26) or the cytoplasm (13M), and the lines directing SSL1-3 to the chloroplast (tc10 and tc3) or the cytoplasm (136) were measured for their botryococcene content. Those plants for each line that were determined to accumulate botryococcene were considered to have inherited the transgene expression cassette and denoted as “+”, while those not accumulating botryococcene were considered to have lost transgene cassette during segregation and denoted as “–”. The tc10+ line expressing SSL1-3 targeted to the chloroplast accumulated the highest level of botryococcene in the middle and late stages

of leaf development, which were about 2.2 and 2.5 times more than that accumulated in line td26+ expressing SSL1-3M targeted to the chloroplast grown in 2013 and 2014 respectively (Figure 4.9). This result coincides well with their greenhouse performance wherein plants harboring botryococcene synthase SSL1-3 targeted to the chloroplasts exhibited two times higher productivity than SSL1-3M. We also found directing the botryococcene biosynthesis to the chloroplasts resulted in significantly higher botryococcene accumulation than when the enzymes were targeted to the cytoplasm (Figure 4.9). Transgenic lines tc10+ and td26+ targeting SSL1-3 and SSL1-3M respectively in the chloroplast in the middle stage of leave development in 2014 accumulated botryococcene 26- and 10-fold greater than that generated by lines 136+ and 13M+ expressing these two enzymes respectively in the cytoplasm. The low production of botryococcene by cytosolic engineering indicated the limited substrates that could be under regulation in the cytoplasm for triterpene biosynthesis. In addition, similar to what we found for squalene accumulating lines, botryococcene also accumulated in a developmental–dependent manner, as the highest level of botryococcene was determined in their middle stage of leave development rather than the young and late stages for most of the high accumulating lines (Figure 4.9).

4.3.9 Development dependent accumulation of methylated triterpene accumulation in 2014

The squalene and methylated squalene content in transgenic lines tpT1 and tpT2 targeting TMT-1 and TMT-2 respectively to the chloroplasts of high squalene accumulating line G1, and line T1G expressing TMT-1 in the cytoplasm of line G1 were evaluated in 2014. As described earlier, the plants for each heterzygous line were segregated by their methylated triterpene accumulation: those determined to accumulate methylated triterpenes were considered to have inherit TMT expression cassette and denoted as “+”, while those that did not were considered to have lost the TMT gene during segregation and denoted as “–”. We found that tpT1+ and tpT2+ lines accumulated a large proportion of methylated squalene, accounting for 56% and 57% of total triterpene accumulating in line tpT1+ in its early and middle stages of leaf development respectively, and for 66% and 70% of total triterpene in line tpT2+ in its early and middle stages of leaf development, respectively. In contrast, only 6% of total triterpene was converted to methylated squalene in line T1G+ in both its early and middle stages of leaf development. This result is consistent with what we observed with

these lines grown under the greenhouse conditions and demonstrated again that TMT enzymes can access the major pool of squalene C30 that is present in the chloroplast in contrast to a small amount of C30 squalene present in the cytoplasm. We also found the level of total triterpenes for each line in middle stage of leave development is about 1.3-1.7 fold greater than their accumulation in early stage, but the ratio of methylated squalene to total squalene in different stages remains the same (Figure 4.10). The other interesting observation is that total triterpene in the high methylated squalene accumulating lines tpT1+ and tpT2+ was 1.2 to 2.2 fold greater than that in their sibling lines tpT1– and tpT2– accumulating only non-methylated squalene at the same stage, and indicating that reduction of the non-methylated squalene pool triggers the biosynthesis of additional squalene, which necessarily means an enhanced flux down the MEP pathway for triterpene production (Figure 4.10).

Transgenic lines tpT3tc and tpT3td targeting TMT-3 to the chloroplasts of high botryococcene accumulating line tc10 and td26, respectively, and lines T3tc and T3td expressing TMT-3 in the cytoplasm of line tc10 and td26, respectively, were evaluated for their botryococcene and methylated botryococene content in 2014. Similar to what we found in transgenic lines engineered for methylated squalene metabolism, the tpT3tc+ and tpT3td+ lines accumulated a high proportion of methylated botryococcene, accounting for 66% and 73% of total triterpene in line tpT3tc+ in its early and middle stages of leaf development, respectively, and for 71% and 59% of total triterpene in line tpT3td+ in its early and middle stages of leaf development, respectively (Figure 4.11). In contrast, 51% and 26% of total botryococcene was converted to methylated botryococcene in line T3tc+ in its early and middle stages, respectively; 12% and 14% of total botryococene was methylated in line T3tc+ and T3tc- in its early and middle stages of development, respectively. An unexpected high level of methylation (51%) was found in line T3tc+ in the early stage for some unknown reasons. The total triterpene accumulated in each line in middle stage of leave development is about 1.3-3.6 fold greater than their accumulation in early stage (Figure 4.11). We also observed a slight higher level of total triterpene in high methylated botryococcene lines (tpT3tc+, tpT3td+) than that in their respective silbling lines (tpT3tc-, tpT3td-), accumulated only non-methylated botryococcene (tpT3tc-, tpT3tc-) and that in their respective parental lines (tc10+, td26+) (Figure 4.11).

4.4 Discussion

In this study, we investigated what, if any, the impact of engineered triterpene metabolism might have on various parameters of plant growth and photosynthesis. Any impact could be caused directly by the accumulation of transgene encoded proteins themselves, distortion of normal physiology by triterpene accumulation in particular cells and intracellular compartments, by the depletion of essential substrates or co-factors metabolites like IPP and SAM, or by indirect consequences of the genetic engineering/regeneration protocol itself. While the current studies could not hope to critically address all of these possibilities, the information should be helpful for uncovering insights into the physiological consequences of such metabolic engineering efforts.

We firstly observed that most of the transgenic lines exhibited different levels of growth reduction relative to the wild type plant. This was not surprising because the transgenic plants may have some growth deficiencies compared to the non-transgenic control plant, due to the non–specific position effects of transgene insertion into the plant genome. Therefore, in order to determine if the growth reduction was due to some insertional event, change due to tissue culture regeneration of the various transgenic lines, or could be a consequence of the introduced terpene metabolism, we chose to grow and evaluate multiple independent lines generated with the same gene construct. It was already promising to find that the best squalene accumulators, G1 and its sibling line G8, only exhibited a slight decrease of growth. This supports a contention that no deleterious and or direct effects of the transgenes on growth occurred.

In contrast, all the high botryococcene accumulating lines exhibited a specific phenotype outcome. The plants were stunted, emerging leaves showed a transient mottling phenotype with essential little chlorophyll/carotenoid accumulation around the petiole/main vein intersections of the leaf that appears to be corrected over time, and gross morphology of the leaves is distinctly different from control plants that also seemed to be moderated as the leaves grew to their more mature forms. These phenotypes were different from anything seen with the squalene accumulating lines, even those high squalene accumulating lines that were somewhat stunted in appearance. The phenotypes were also reliable observed in successive growth seasons, and thus not simply induced by environmental conditions. Considering the biosynthetic

similarities between botryococcene metabolism and squalene metabolism, the greater impact of botryococcene on phenotypic outcome might be associated with its distinct structure differences to squalene and the recognition of squalene as a natural, native constituent but botryococcene not. Plant lines accumulating methylated botryococcene exhibited even more dramatic phenotypic outcomes consistent with these molecules becoming more physically un-natural in tobacco and possibly more toxic.

Surprisingly, photosynthetic rates in the transgenic lines were not or only marginally affected. But many of the transgenic lines, especially the high triterpene accumulators, exhibited a higher conductance than wildtype plants. Water conductance was also found to be higher in the plant lines accumulating methylated triterpenes, but this was not more significant than the lines accumulating non-methylated triterpene, which suggests that this measure, water conductance, is not sufficient to account for the adverse phenotypes observed. However, we would predict on the basis of the water conductance measure that the triterpene accumulating lines would be more drought sensitive.

Another overarching goal in a study like this was to determine if the triterpene yield of the transgenic plants grown under field condition was stable and comparable to that of greenhouse grown plants. The determination of triterpene content in the field grown plants demonstrated that the accumulation levels was developmental stage dependent, similar to those findings with greenhouse grown plants (Wu et al., 2012). However, more striking was that the levels of triterpene accumulation in greenhouse grown plants were capitulated in the field trial plants as well. For instance, the squalene levels tested for greenhouse grown G1 was 200-600 µg/g fw (data not shown) compared to field grown G1 accumulating, on average over 3 growth seasons, 438 µg/g fw. Hence, we conclude that the metabolic engineered trait for triterpene production is indeed stably inherited and expressed under a wide range of growth conditions, field versus greenhouse in particular.

Triterpene production for each transgenic line did vary between different growing seasons, and the weather conditions between different growing seasons could be a factor influencing triterpene production and accumulation. For example, the rainfall from May to Aug in 2013 was 1.6 and 1.3 times more than that in 2012 and 2014, respectively. This difference between dry and wet seasons directly leads to about 2 times more biomass production in wet season than the dry season. However, squalene level per leave area decreased with the increased biomass accumulation in 2013. Squalene

production for most of high accumulators grown at the middle stage in 2012 was about 1.5 times more than that of same lines grown in 2013. Hence, it was counter-initiative to what we would have predicted. We would have predicted a direct correlation between biomass accumulation and triterpene accumulation. This was not observed and suggests that there must be additional regulatory mechanisms controlling carbon allocation under these different growth conditions. If we were to understand these mechanisms, then perhaps we could use this information to further augment triterpene accumulation beyond what was observed with the existing plant materials.

4.5. Material and method

4.5.1 Seedling preparation and float setup

Seeds for all the transgenic lines and WT were germinated without any selection in the soil in the greenhouse. After approximately 2 weeks, the seedlings were randomly picked and transferred to the sterilized float beds filled with sterilized soil. The float beds were kept on water beds of 3 and 5 inches. Greenhouse temperatures were maintained around 72^{o}F (70 to 75^{o}F) during the day and 60 to 70^{o}F during the night. The initial water beds were prepared with 4.2 lbs of 20-10-20 fertilizer per 1000 gallons float water with 1 fl oz Terramaster 4EC per 100 gallons float water as a preventative treatment for fungal diseases. Fertilizer and Terramaster treatments were reapplied every two weeks. It generally took about 8 weeks from seeding to when plantlets were ready for transplanting to the field.

Seed stocks were generated at the University of Kentucky propagated in greenhouse facilities certified by authorization by the UK Biosafety Office. Appropriate APHIS field permits for performing field trials with the transgenic materials were obtained for each of the annual field trials. All equipment used for propagating transgenic materials were clearly labeled and segregated for use with transgenic plant use only. The floatbed flats were transported to the designated University of Kentucky Spindletop field site according to APHIS guidelines and the field site treated pre-, during and post-planting as dictated by APHIS rules for transgenic trials. The field sites were randomly inspected by APHIS inspectors during all 3 years of field trials.

4.5.2 Field plot setup

All the field work complied with the performance standards as required by USDA–APHIS. Thirty to thirty-six transplants for each line were randomly selected from the greenhouse propagation trays for the field planting. Plants were moved from the greenhouse to the field locations in an enclosed trailer to minimize environmental exposure. Three replicate rows of each line with 12 plants in a row were grown with standard plant and row spacing. They were planted in a designated field area with a minimum 50 foot perimeter area around the transgenic test area to maintain the field site was free of sexually compatible species to tobacco. Out-crossing of the transgenic lines was prevented by toping plants showing flowering buds, and maintaining an isolation distance of at least 1,320 feet between the transgenic plots and any non–transgenic tobacco. A distance of at least 5,280 feet was maintained between the transgenic plots and any open pollinated seed tobacco plots. Equipment used to transplant the plants was thoroughly cleaned before leaving the plot area.Cleaning included visual inspection to remove obvious plants or plant parts, water rinse followed by application of 10% bleach solution, second rinse and final inspection to insure no viable plant tissue remained. Any remaining plants were discarded in an area adjacent to the plots and incorporated into the soil. This area was monitored and tilled as needed to destroy volunteer plants over 4 year period.

Standard tobacco agricultural practices were implemented to control insects, weeds, and pathogens. The test plot was visited at least once a week to document observations regarding plant growth and morphology. Test plots were monitored weekly for weed, disease and insect infestation and noted conditions documented. The test plots were clearly marked to ensure that their identity was maintained throughout the field trial. The field plot was visited at least once per week after planting for the duration of the growing season to ensure and document the maintenance of the non–flowering stage. Any plants showing signs of flowering were topped. No seed pods were observed during the 3 field trial growth seasons.

Plants were harvested approximately 10–12 weeks after planting. Harvested plants were measured and weighed at a measuring station neighboring the plot, leaf samples collected and the remaining plant material returned to the plot for incorporation into the soil. Test plots will be monitored for volunteers at monthly intervals, and any volunteer plants destroyed prior to flowering, for at least one year post-harvest or until no volunteer plants are observed. Leaf material was transported back to the University of Kentucky for leaf area determination, after which the material was steam killed before disposal.

4.5.3 Photosynthesis and squalene determination

Leaf discs of 2 cm diameter were collected from the upper most, fully expanded leaves at several times during each field trial and at different leaf positions at least once per each field trial. The leaf materials was then extracted and triterpene levels determined by GC-MS as described by (Wu et al., 2012). The photosynthetic gas exchange measurements of first fully expanded leaves were determined at atmospheric concentrations of CO_2 and a saturating irradiance of 1,500 micromoles photons m^{-2} s^{-1} using a LI–COR 6400 portable photosynthesis system according to (Salvucci and Crafts-Brandner, 2004).

Tables and figures of chapter 4

Table 4.1 Yield of various terpenes in tobacco plants engineered for novel plastidic biosynthetic capacities.

Terpene type	Terpenes	Target gene	Yield µg/g FW (max)	Host	Reference
Monoterpene	R–(+) Limonene	Limonene synthase (Citrus rind)	400 ng/g(Greenhouse)	Nicotiana tobacum	Wu et al., 2006
Sesquiterpene	Patchoulol	Patchoulol synthase (Patchouli)	30 µg/g (Greenhouse)	Nicotiana tobacum	Wu et al., 2006
	Amorpha–4, 11–diene	Amorpha–4,11–diene synthase (Artemisia annua L)	25 µg/g (Greenhouse)	Nicotiana tobacum	Wu et al., 2006
triterpene	Squalene	Squalene synthase (yeast)	1000 µg/g (greenhouse)	Nicotiana tobacum	Wu et al., 2012
	Botryococene	Botryococene synthase (Botryococcus Braunii)	600 µg/g (field)	Nicotiana tobacum	Jiang et al., unpublished
	Methylated triterpene	Triterpene methyltransferase (Botryococcus Braunii)	550 µg/g (Greenhouse)	Nicotiana tobacum	Jiang et al., unpublished

Table 4.2 Kentucky Monthly Precipitation (2012–2014) (Inches)

Year\Mon	May	Jun	Jul	Aug	Total (growing season)
2012	4.25	1.5	5.54	3.17	14.46
2013	4.44	6.23	6.42	4.14	21.23
2014	3.65	3.98	3.67	5.43	16.73

data collected from University of Kentucky Ag weather center

Table 4.3 Transgenic lines engineered with various constructs that directed triterpene metabolism to different cellular compartments under the direction of constitutive or trichome specific promoters used for the 2012 field study.

Construct	Nomenclature	Generation	Genotype	Target compartment	Expression type	Triterpene type
WT	1068	N	N	N	N	squalene
e2SQS–32	A7	T2	Homozygous	cytoplasm	trichome	squalene
e2tpSQS–50	D4	T2	Homozygous	chloroplast	trichome	squalene
e2SQS+e2FPS–26	N10	T2	Homozygous	cytoplasm	trichome	squalene
e2tpSQS+e2tpFPS–23	R13	T2	Heterzygous	chloroplast	trichome	squalene
SQS+FPS–31	C5	T2	Homozygous	cytoplasm	constitutive	squalene
tpSQS+tpFPS–5	G8	T2	Homozygous	chloroplast	constitutive	squalene
tpSQS+tpFPS–5	G1	T2	Homozygous	chloroplast	constitutive	squalene
tpSQS+tpFPS–15.1	H5	T2	Homozygous	chloroplast	constitutive	squalene
tpSQS+tpFPS–42	I8	T2	Homozygous	chloroplast	constitutive	squalene

Wild type tobacco (*Nicotiana tabacum* accession 1068) was transformed with the respective gene constructs consisting of yeast squalene synthase (SQS) and the avian farnesyl diphosphate (FPS) gene inserted downstream of strong, viral promoters (35S cauliflower mosaic viral promoter and cassava vein mosaic viral promoter) for constitutive expression. Double 35S enhancers (e2) fused to trichome specific promoters, cembreinene synthase and hydroxlase promoters respectively (Wang et al., 2002a; Ennajdaoui et al., 2010) were used for trichome specific expression of two respective terpene synthase genes. A plastid targeting signal sequence (tp) was fused to the 5' end of the respective genes for targeting the enzymes to the chloroplast, and for the constructs without tp, the enzymes were expressed in the cytoplasm. T2 homozygous or heterozygous transgenic lines were selected for the squalene expression cassette based on separate antibiotic selection screens. The seeds for each line was germinated without selection and plants randomly selected for planning in the field test assessments.

Table 4.4 Transgenic lines engineered with various constructs that directed triterpene metabolism to different cellular compartments under the direction of constitutive or trichome specific promoters used for the 2013 field study.

Construct	Nomenclature	Generation	Genotype	Target compartment	Expression type	Triterpene type
WT	1068	N	N	N	N	squalene
e2SQS-32	A7	T2	Homozygous	cytoplasm	trichome	squalene
e2tpSQS-50	D4	T2	Homozygous	chloroplast	trichome	squalene
e2SQS+e2FPS-26	N10	T2	Homozygous	cytoplasm	trichome	squalene
e2tpSQS+e2tpFPS-23	R13	T2	Heterzygous	chloroplast	trichome	squalene
SQS+FPS-31	C5	T2	Homozygous	cytoplasm	constitutive	squalene
tpSQS+tpFPS-5	G8	T2	Homozygous	chloroplast	constitutive	squalene
tpSQS+tpFPS-5	G1	T2	Homozygous	chloroplast	constitutive	squalene
tpSQS+tpFPS-15.1	H5	T2	Homozygous	chloroplast	constitutive	squalene
tpSQS+tpFPS-42	I8	T2	Homozygous	chloroplast	constitutive	squalene
tpSSL1-3+tpFPS-10	tc10	T1	Heterzygous	chloroplast	constitutive	botryococcene
tpSSL1-3M+tpFPS-26	td26	T1	Heterzygous	chloroplast	constitutive	botryococcene

Transgenic lines engineered for botryococcene biosynthesis were generated by transforming the respective gene constructs consisting of botryococcene synthase gene (SSL1-3) with (SSL1-3M) or without M (ER membrane target sequence) and the avian farnesyl diphosphate (FPS) gene inserted downstream of strong, viral promoters (35S cauliflower mosaic viral promoter and cassava vein mosaic viral promoter), into wild type tobacco (*Nicotiana tabacum* accession 1068). A plastid targeting signal sequence (tp) was fused to the 5' end of the respective genes for targeting the enzymes to the chloroplast, and for the constructs without tp, the enzymes were expressed in the cytoplasm. The seeds for the putative heterzygous T1 lines engineered for botryococcene metabolism were germinated without selection and plants randomly selected for planning in the field test assessments. These two T1 lines (tc10 and td26) were both shown to be segregated in the field based on whether or not accumulating botryococcene. Those plants accumulating botryococcene were considered to harbor the transgene (denoted as “+”), while those that did not were considered to have lost the transgene during segregation were denoted as “–”. T2 transgenic lines homozygous or heterozygous for squalene biosynthesis were described in table 4.3.

Table 4.5 Transgenic lines engineered with various constructs that directed triterpene metabolism to different cellular compartments under the direction of constitutive or trichome specific promoters used for the 2014 field study.

Construct	Nomenclature	Generation	Genotype	Target compartment	Expression type	Triterpene type
WT	1068	N	N	N	N	squalene
tpSQS+tpFPS–5	G8	T2	Homozygous	chloroplast	constitutive	squalene
tpSQS+tpFPS–5	G1	T3	Homozygous	chloroplast	constitutive	squalene
SQS+FPS–31	C12	T2	Homozygous	cytoplasm	constitutive	squalene
tpSQS+tpFPS–15.1	H5	T2	Homozygous	chloroplast	constitutive	squalene
tpSSL1+3–tpFPS–3	tc3	T1	Heterzygous	chloroplast	constitutive	botryococcene
tpSSL1+3–tpFPS–10	tc10	T1	Heterzygous	chloroplast	constitutive	botryococcene
tpSSL1+3–tpFPS–10	td26	T1	Heterzygous	chloroplast	constitutive	botryococcene
SSL1–3+FPS–6	13–6	T1	Heterzygous	cytoplasm	constitutive	botryococcene
SSL1–3M+FPS–35	13M–35	T1	Heterzygous	cytoplasm	constitutive	botryococcene
tpTMT1–G1-53	tpT1	T1	Heterzygous	chloroplast	constitutive	Methylsqualene
tpTMT2–G1–51	tpT2	T1	Heterzygous	chloroplast	constitutive	Methylsqualene
TMT1–G1–20	T1G	T1	Heterzygous	cytoplasm	constitutive	Methylsqualene
tpTMT3–tc10–37	tpT3tc	T1	Heterzygous	chloroplast	constitutive	Methylbotryococcene
tpTMT3–td31–35	tpT3td	T1	Heterzygous	chloroplast	constitutive	Methylbotryococcene
TMT3–tc10–10	T3tc	T1	Heterzygous	cytoplasm	constitutive	Methylbotryococcene
TMT3–td31–2	T3td	T1	Heterzygous	cytoplasm	constitutive	Methylbotryococcene

Methylated triterpene accumulating lines were generated by each indicated construct with one of three TMT genes targeted to the chloroplast (with tp) or the cytoplasm (without tp) of squalene (tpSQS+tpFPS-5, G1), or botryococcene (tpSSL1-3+tpFPS-10, tc10 or tpSSL1-3M+tpFPS-31, td31) accumulating lines. The seeds for the putative heterzygous T1 lines engineered for methylated triterpene metabolism were germinated without selection and plants randomly selected for planning in the field test assessments. These T1 lines were all shown to be segregated in the field based on whether or not accumulating methylated triterpene. Those plants were determined to accumulate methylated triterpene were considered to harbor the transgene (denoted as “+”), while those that did not were considered to have lost the transgene during segregation were denoted as (–). Transgenic lines homozygous or heterozygous for squalene or botryococcene biosynthesis were described in table 4.3 and 4.4.

Table 4.6 Growth reduction of squalene accumulating lines compared to wildtype plants and genetic segregating siblings (–) that did only accumulate squalene grown in 2012 (A) and 2013 (B).

A

#/WT (%)	Height	Leaf Weight	Total Weight	Leaf area
WT	100	100	100	100
G1	96	99	97	110
G8	94	87	87	107
I8	93	76	75	86
H5	82	73	71	79
C5	98	62	65	68
A7	110	89	94	93
N10	102	85	87	83
R13+	70	53	51	63
R13–	101	99	99	95
P1	93	58	63	57
D4	112	113	117	115

B

#/WT (%)	Height	Leaf Weight	Total Weight	Leaf area
WT	100	100	100	100
G1	90	98	89	87
G8	88	80	73	81
I8	93	90	84	87
H5	63	46	38	51
C5	103	56	62	60
A7	100	75	78	77
N10	99	73	75	70
R13+	81	56	51	65
R13–	94	93	97	92
R6	74	61	53	68
D4	98	85	88	86

The percentages of four differerent growth parameters of each indicated transgenic line to that of wildtype plants were represented. Line R13 was a T1 generation seed lot and shown to be heterozygous for the triterpene expression cassette based on separate antibiotic selection screens. The R13 seed was germinated without selection and plants randomly selected for planning in the field test assessments. Individual R13 plants in replicate rows were subsequently screened for their ability to accumulate squalene (+) or not (-) and the data for performance of (+) and (-) plants within a row pooled for their comparison. Nomenclature for the various lines is the same as given in table 4.3.

Table 4.7 Growth reductions of botryococcene accumulating lines (+) compared to wildtype plants and genetic segregating siblings (-) that did only accumulate non-methylated triterpene grown in 2013 (A) and 2014 (B).

A

#/WT (%)	Height	Leaf Weight	Total Weight	Leaf area
WT	100	100	100	100
G1	90	98	89	87
td26+	75	60	55	59
td26–	96	81	84	92
tc10+	74	65	58	68
tc10–	103	88	88	79

B

#/WT (%)	Height	Leaf Weight	Total Weight	Leaf area
WT	100	100	100	100
G1	76	99	95	104
tc10+	74	68	66	58
tc10–	92	111	114	113
tc3+	72	63	60	66
tc3–	97	123	122	124
td26+	76	80	75	91
td26–	95	127	129	115
13M+	91	81	91	84
13M–	93	109	124	99
136+	88	99	114	109
136–	89	95	98	117

The percentages of four differerent growth parameters of each indicated transgenic line to that of wildtype plants were shown. The seeds for each line were germinated without selection and plants randomly selected for planning in the field test assessments. Individual plants in replicate rows for each line were subsequently screened for their ability to accumulate botryococcene (+) or not (-) and the data for performance of (+) and (-) plants within a row pooled for their comparison. Nomenclatures for the various lines were described in table 4.3, 4.4 and 4.5.

Table 4.8 Growth reduction of methylated triterpene accumulating lines compared to wild type plants (A) and genetic segregating siblings (-) (B) that did only accumulate non-methylated triterpene during the 2014 growth season.

A

#/WT (%)	Height	Leaf Weight	Total Weight	Leaf Area
WT	100	100	100	100
G1	76	99	95	104
tc10+	74	68	66	58
Td26+	76	80	75	91
tpT1+	73	59	57	65
tpT1–	77	85	78	75
tpT2+	70	65	61	70
tpT2–	80	89	78	87
tpT3tc+	68	61	61	81
tpT3tc–	69	89	86	85
tpT3td+	64	84	77	99
tpT3td–	72	95	90	106
T1G+	70	81	75	90
T1G–	68	90	83	95
T3tc+	80	69	71	69
T3tc–	82	85	88	89
T3tdS+	56	64	56	68
T3td–	62	65	58	77

B

+/– (%)	Height	Leaf Weight	Total Weight	Leaf Area
tpT1+/tpT1–	94	69	72	86
tpT2+/tpT2+	87	73	71	80
tpT3tc+/tpT3tc–	98	68	70	95
tpT3td+/tpT3td–	89	89	86	93
T1G+/T1G–	104	90	90	95
T3tc+/T3tc–	98	81	81	78
T3td+/T3td–	90	98	97	88

The percentages of four differerent growth parameters of lines accumulating methylated triterpene to that of lines only accumulating non-methylated triterpene or non-transgenic wild type control lines were shown. The seeds for each transgenic line were germinated without selection and plants randomly selected for planning in the field test assessments. Individual plants in replicate rows for each line were subsequently screened for their ability to accumulate methylated triterpene (+) or not (-) and the data for performance of (+) and (-) plants within a row pooled for their comparison. Nomenclatures for the various lines are described in table 4.3 4.4 and 4.5.

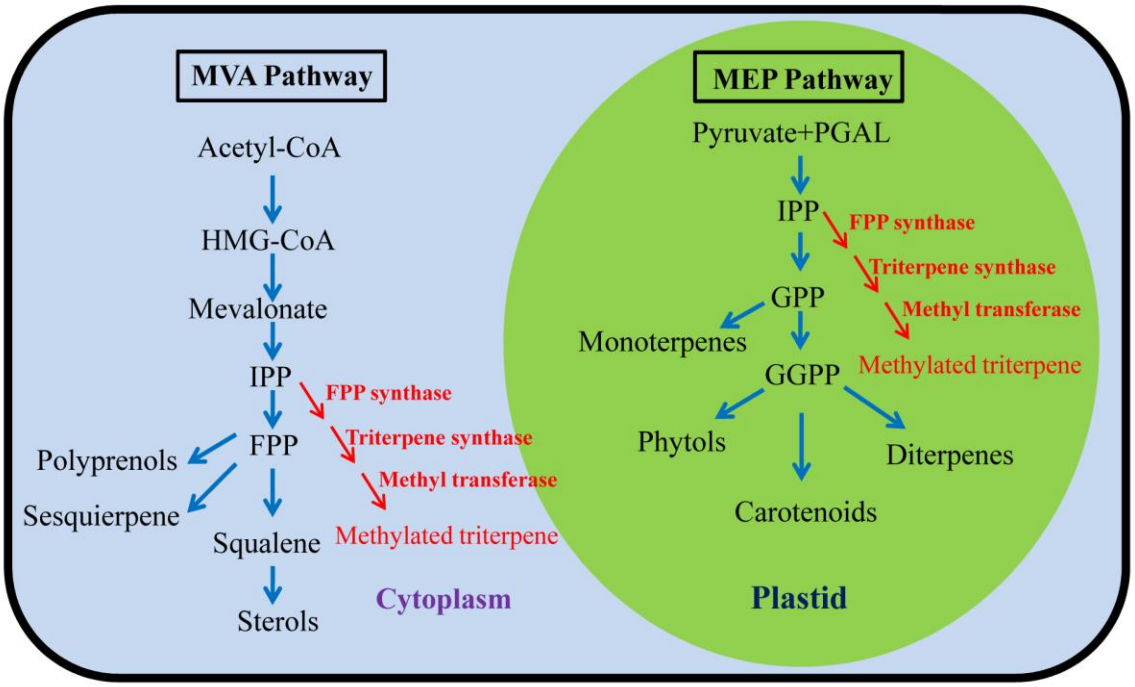


Figure 4.1 Outline of the two terpene biosynthetic pathways that operated in plants and strategies for engineering novel triterpene metabolism from the MVA (cytosolic) or MEP (plastidic) pathways.

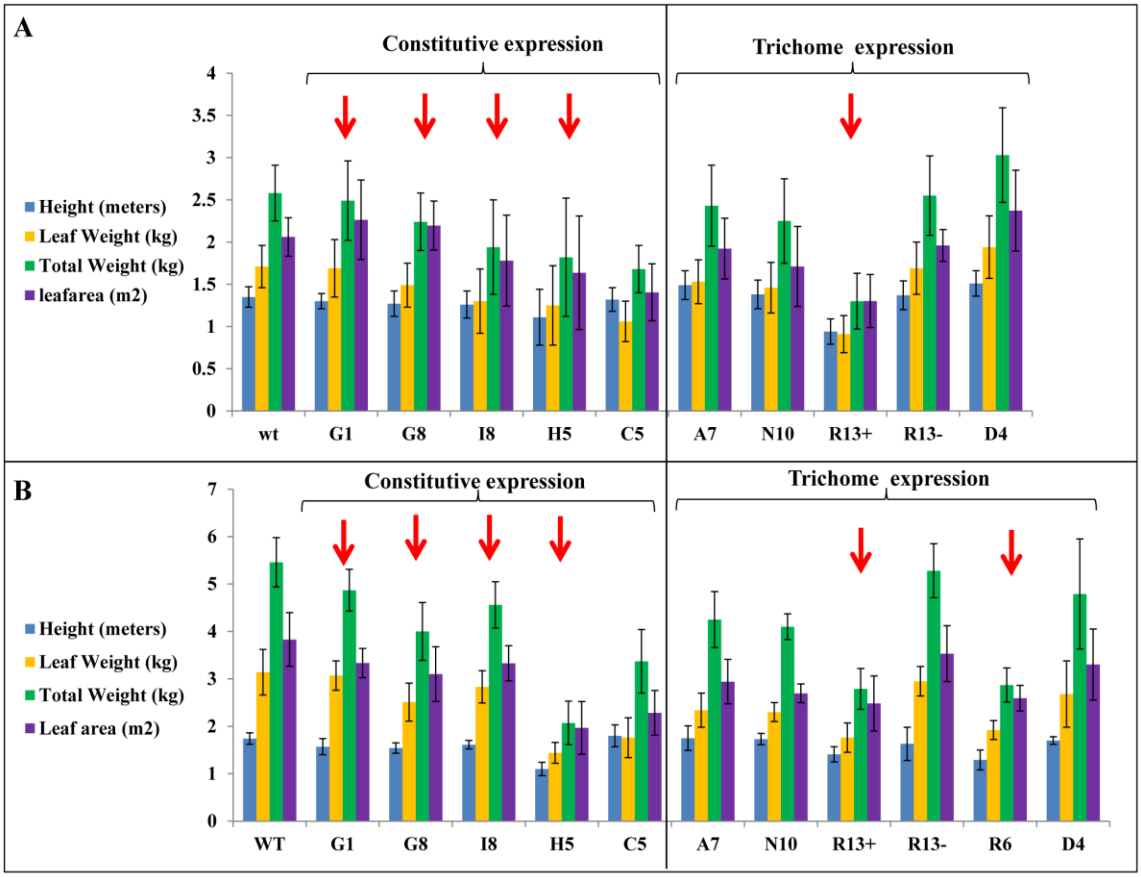


Figure 4.2 The growth performance of transgenic tobacco engineered for squalene metabolism. Transgenic lines targeting SQS with FPS to the cytoplasm (C5) or chloroplasts (G1, G8, I8, H5) under the direction of constitutive promoters or targeting SQS with FPS (R13) or only SQS (D4) to the chloroplasts, or targeting expressing of SQS with FPS (N10) or only SQS (A7) to the cytoplasm of trichomes were grown in field studies of 2012 (A) and 2013 (B). The height, leaf weight, total weight and leaf area of 9 individual of plants from 3 independent rows for each respective line were measured at the end of field season. An heterologous, segregating population of line R13 was planted in both years and are distinguished as those plants that accumulated high levels of squalene (>50 µg/g fw),denoted as "+", and the others having wild type levels of squalene are denoted as "-". The high squalene accumulating lines (>50 µg/g fw) are highlighted by red arrows. The values shown are the averages of 6-9 determinations from 9 individual plants from 3 independent rows.

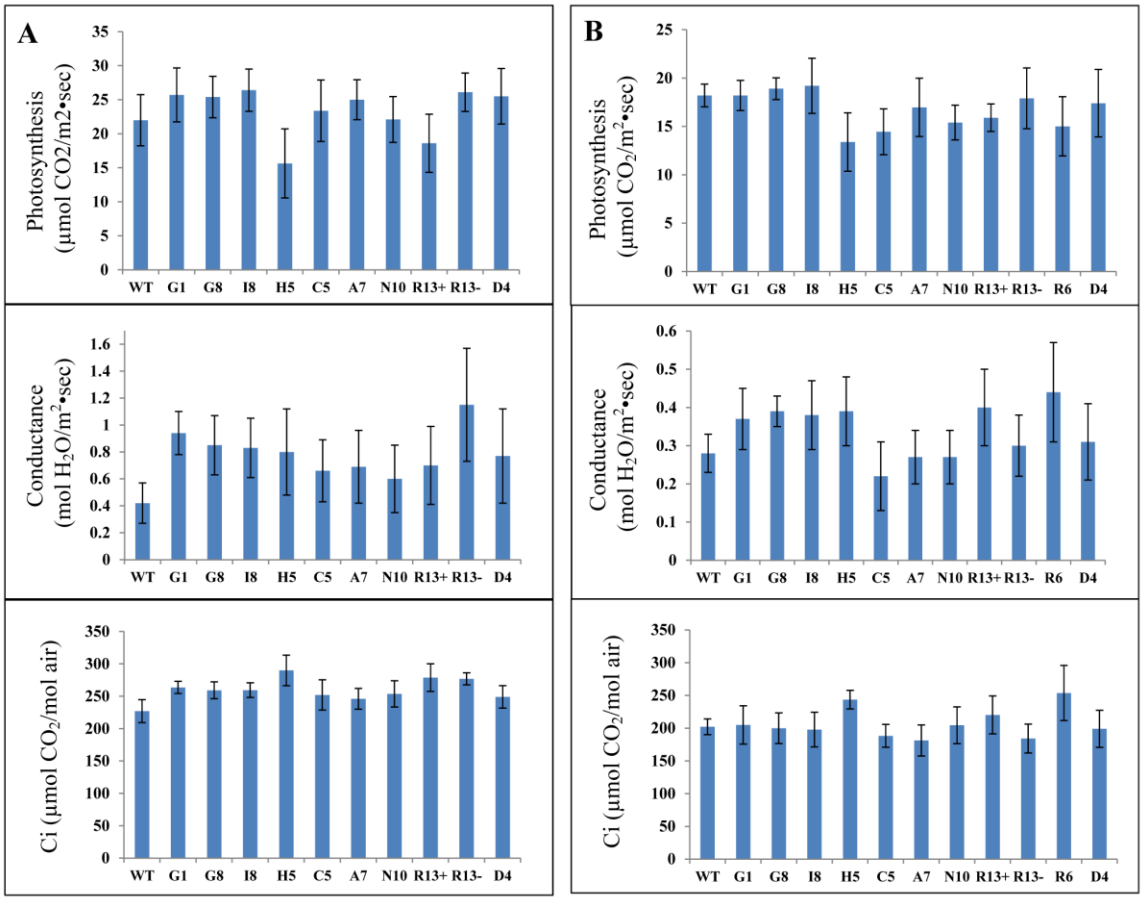


Figure 4.3 Photosynthetic parameters of transgenic tobacco engineered with squalene metabolism grown in the field of 2012 (A) and 2013 (B). The photosynthesis rates, water conductance and Ci of 9 individual of plants from 3 independent rows for each respective line were measured in the middle of each growth season. The values shown are the averages from 6-9 determinations. Nomenclature for the various lines is the same as given in figure 4.2.

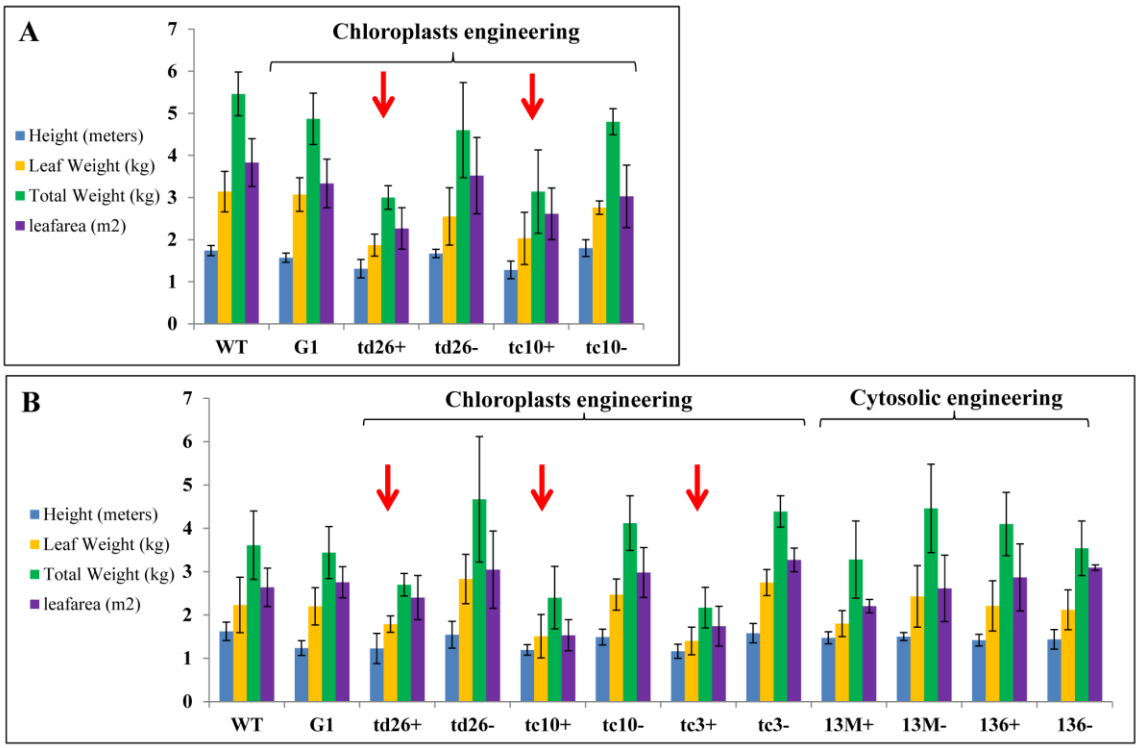


Figure 4.4 Growth performance of transgenic tobacco engineered for botryococcene metabolism to particular compartments grown in the field of 2013 (A) and 2014 (B). Transgenic lines expressing SSL1-3 or SSL1-3M with FPS targeted to the chloroplast (line tc10 or tc3 and line td26 respectively) or the cytoplasm (line 136 and line 13M respectively) were grown and evaluated in the field in comparison to control wild type plants as well as those engineered for high squalene accumulation (G1). The botryococcene accumulating lines represent genetic segregating T1 population where those plants accumulating botryococcene were denoted as "+" and are homozygous or heterozygous for the botryococcene expression cassette, and those not accumulating botryococcene were denoted as "-" and represent homozygous plants without any transgene cassettes. Plant height, leaf weight, total weight and leaf area of 9 individual plants from 3 independent rows for each respective line were measured at the end of growth cycle for each growing season. The values shown are the averages of 6-9 determinations. The high botryococcene accumulators (>50 µg/g fw) are highlighted by red arrows.

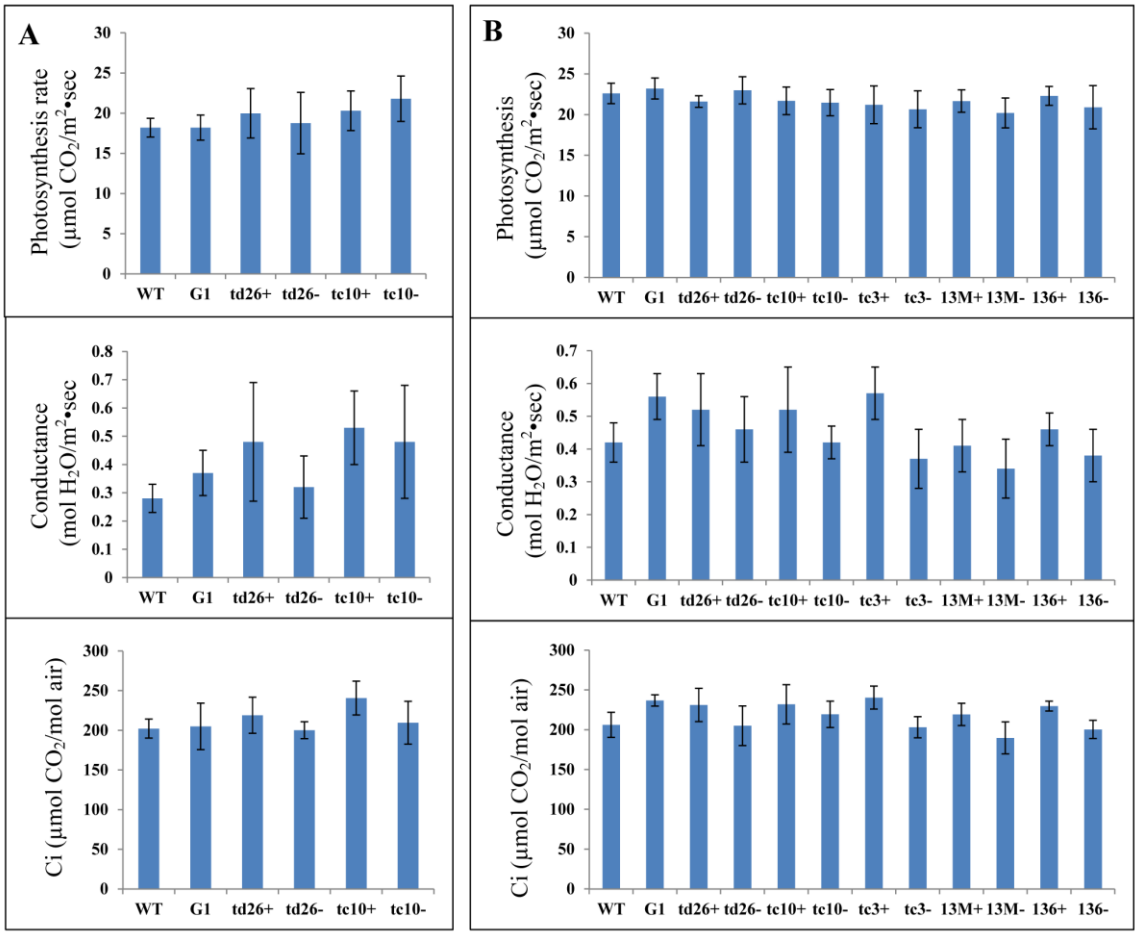


Figure 4.5 Photosynthetic parameters of transgenic tobacco lines engineered for botryococcene accumulation grown in 2013 (A) and 2014 (B) field trials. The photosynthetic rates, water conductance and Ci of 9 individual of plants from 3 independent rows for each respective line for were measured in the middle of each growing season. The values shown are the averages from 6-9 determinations. The nomenclature for the various lines is given in Figure 4.4.

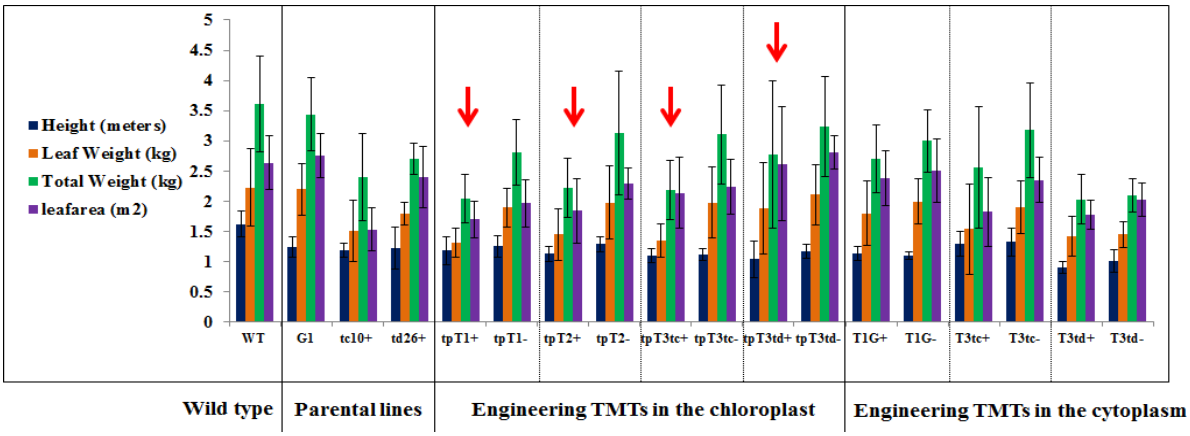


Figure 4.6 Growth performance of transgenic tobacco lines engineered for methylated triterpene accumulation. Transgenic lines expressing TMT-1 and TMT-2 targeted to the chloroplast (line tpT1 and tpT2 respectively) or directing TMT-1 to the cytoplasm (line T1G) of the high squalene accumulating line (G1), and lines that expressing TMT-3 in the chloroplasts (tpT3tc) or the cytoplasm (T3tc) of high botryococcene accumulating line (tc10) or targeting TMT-3 to the chloroplasts (tpT3td) or the cytoplasm (T3td) of high botryococcene accumulating line (td26) were grown in field studies of 2013 (A) and 2014 (B). Plant height, leaf weight, total weight and leaf area of 9 plants for each line was measured at the end of the growing season by measuring 6-9 individual of plants from 3 independent rows. The highest botryococcene accumulating lines are highlighted by red arrows. The plant lines engineered for methylation of the triterpenes and planted in the 2013 and 2014 field seasons were segregating populations representing homozygous/heterozygous for the methylation expression cassette as well as homozygous without the expression cassette. Those plants for each line that accumulated methylated triterpene are denoted as “+”, and those that did not accumulate methylated triterpenes are denoted as “-”.The values shown are the averages of 6-9 determinations.

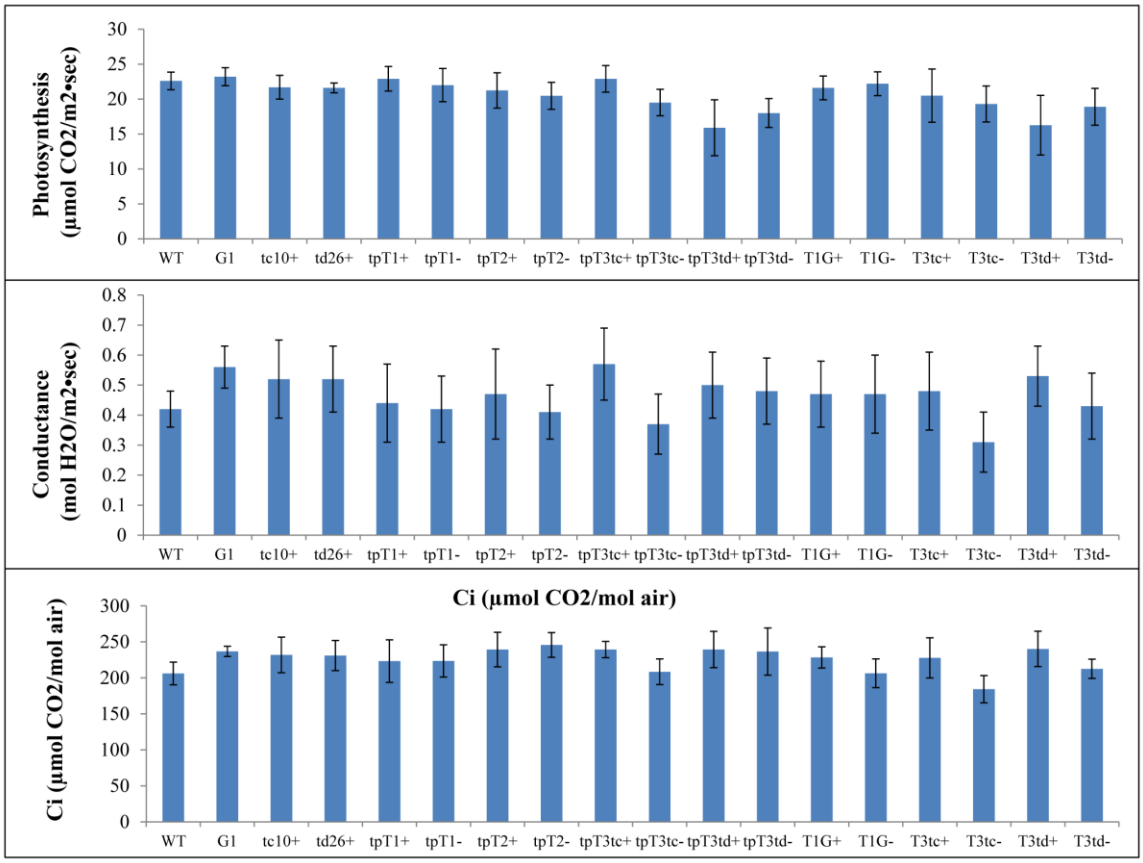


Figure 4.7 Photosynthetic parameters of transgenic tobacco engineered for methylated triterpene accumulation grown in 2014. The photosynthetic rates, water conductance and Ci of respective lines were measured in the middle of each growth season by evaluating 9 individual of plants from 3 independent rows. The values shown are the averages from 6-9 determinations. Nomenclature for the various lines is given if fig. 4.6.

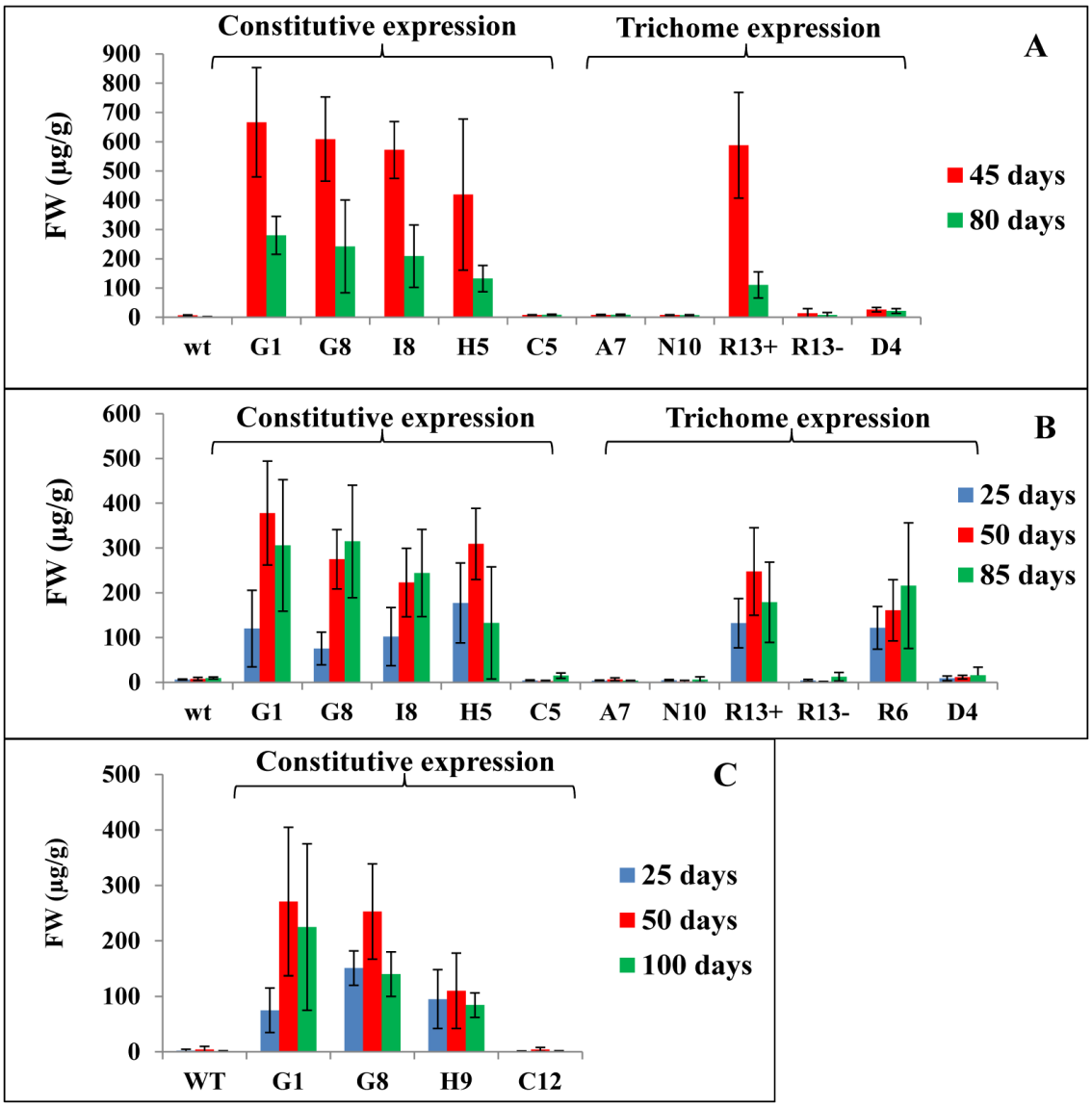


Figure 4.8 Squalene accumulation in transgenic lines grown in 2012 (A), 2013 (B) and 2014 (C). The levels of squalene were determined for the uppermost, fully expanded leaf of 6 to 9 plants for each transgenic line grown in three replicate rows at different growing stages: 25 days (blue), 50 days (red), and 80-100 days (green) after they grown in the field. The values shown are the averages of 6-9 determinations. Nomenclature for the various lines is given in Tables 4.3-4.5.

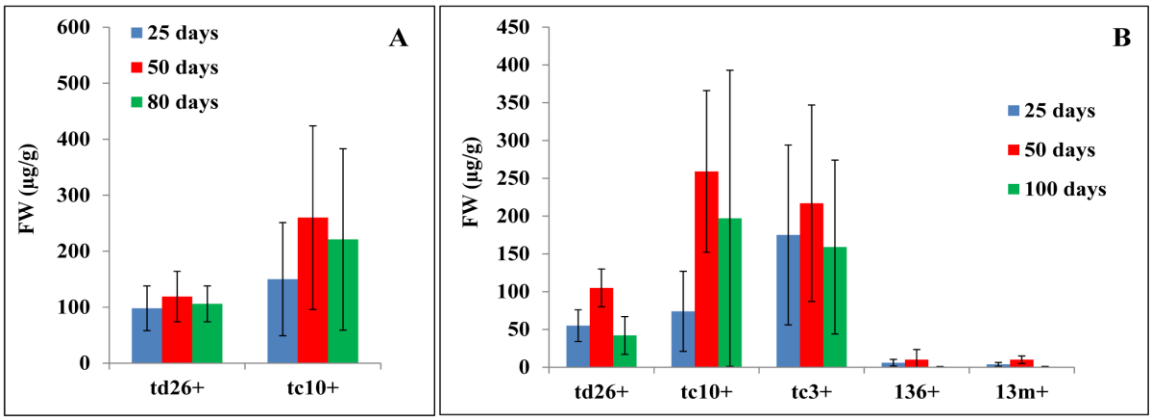


Figure 4.9 Botryococcene accumulation in transgenic lines grown in 2013 (A) and 2014 (B). The levels of botryococcene were determined for the uppermost, fully expanded leaf of 6 to 9 plants for each transgenic line grown in three replicate rows at different growing stages: 25 days (blue), 50 days (red), and 80-100 days (green) after they planted in the field. The values shown are the averages of 6-9 determinations. Nomenclature for the lines are defined in Tables 4.3-4.5.

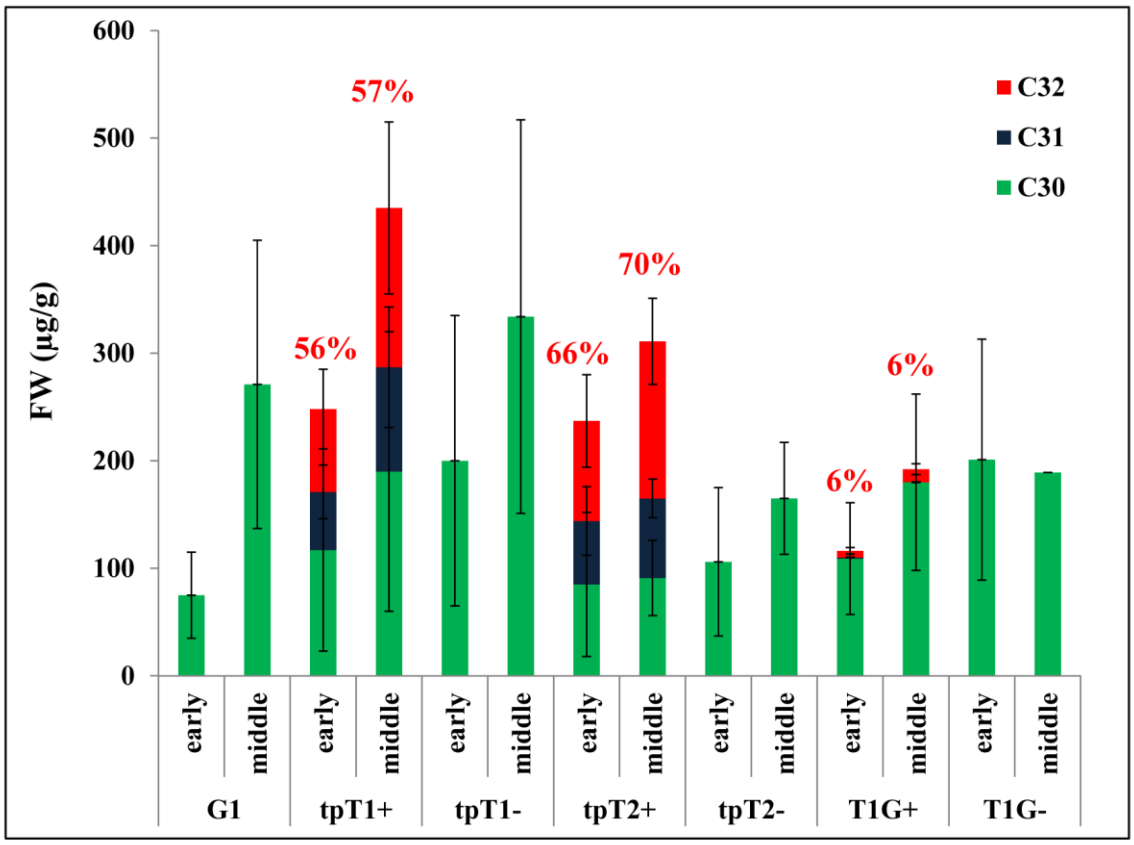


Figure 4.10 Methylated squalene accumulation in transgenic lines field grown in 2014. The levels of botryococcene were determined for the uppermost, fully expanded leaf of 6 to 9 plants for each transgenic line grown in three replicate rows at different growth stages: early (25 days) and middle (50 days) stages. Transgenic lines expressing TMT-1 and TMT-2 targeted to the chloroplast (line tpT1 and tpT2 respectively) or directing TMT-1 to the cytoplasm (line T1G) of the high squalene accumulating line (G1) were evaluated. Those plants for each line that accumulated methylated triterpene were denoted as "+", and those accumulated only non-methylated triterpene were denoted as "-". The percentage of methylated squalene (C31+C32) to total squalene (C30+C31+C32) was highlighted above the line column. The values shown are the averages of 6-9 determinations. Nomenclature for the lines are defined in Tables 4.3-4.5.

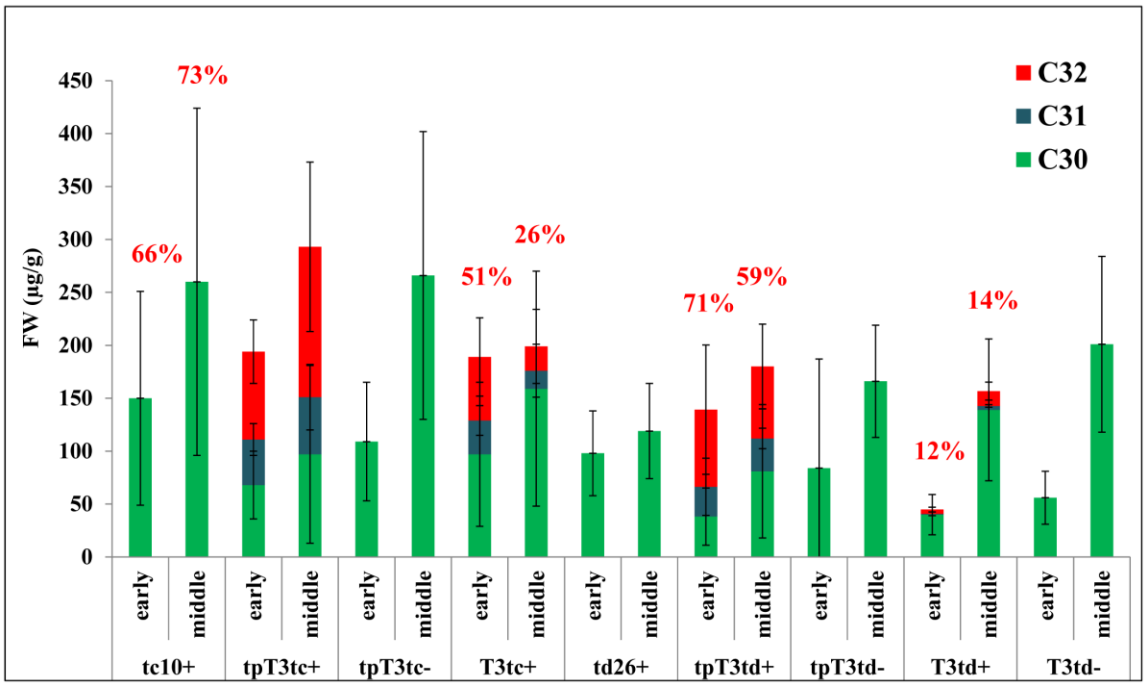


Figure 4.11 Methylated botryococcene accumulations in transgenic lines grown in 2014. The levels of botryococcene were determined for the uppermost, fully expanded leaf of 6 to 9 plants for each transgenic line grown in three replicate rows at different growing stages: early (25 days) and middle stages (50 days). Transgenic lines tpT3tc and tpT3td expressing TMT-3 targeted to the chloroplast of line tc10 and td26 respectvely or directing TMT-1 to the cytoplasm of line tc10 and td26 respectively were evaluated. Those plants for each line that accumulated methylated triterpene were denoted as "+", and those accumulated only non-methylated triterpene were denoted as "-". The percentage of methylated squalene (C31+C32) to total squalene (C30+C31+C32) was highlighted above the line column. The values shown are the averages of 6-9 determinations. Nomenclature for the lines are defined in Tables 4.4-4.5.

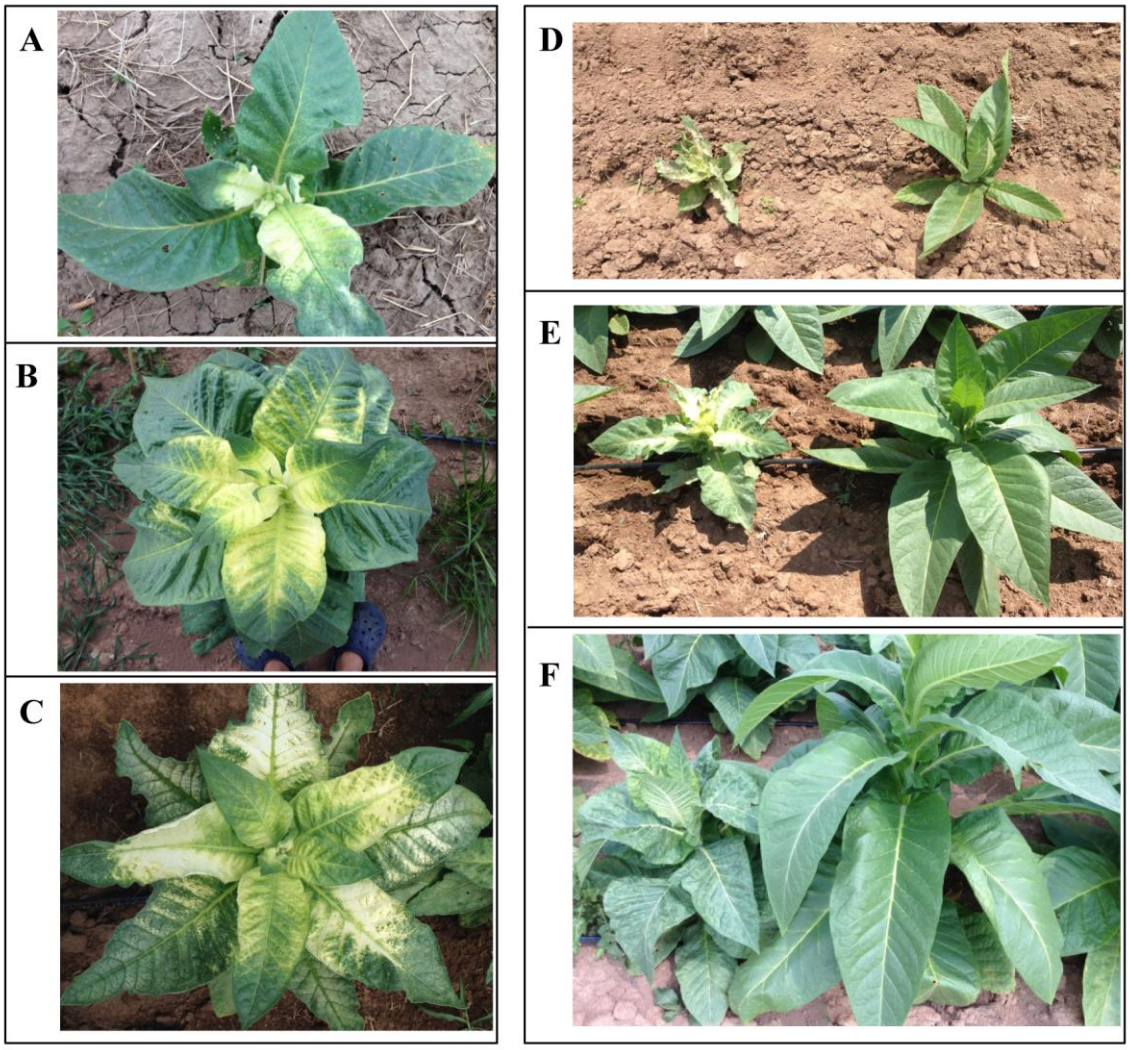


Figure 4.12 Phenotypes of botryococcene accumulating lines grown in the field trials. High botryococcene accumulators segregated and exhibited a unique phenotype (chrinkle, mottling and chlorosis) at their early (A, D left), middle (B, E left) and late (C, F left) stages of growth relative to their non-botryococcene accumulating siblings (D right, E right, and F right).

Chapter 5: Conclusion remarks

The current work proved that redirection of C5 (IPP/DMAPP) precursors from the MEP pathway in the tobacco chloroplast can successfully build novel triterpene products in plants. In addition to earlier studies, this strategy has been demonstrated to be widely applicable to metabolic engineering of various types of terpene compounds including monoterpenes, sesquiterpenes, and now triterpenes in tobacco plants. However, one should not overlook that the accumulation level of each type of terpene differed significantly, in that triterpene production is usually within range from 200 to 1000 µg/g fw, whereas sesquiterpene production did not exceed 30 µg/g fw and monoterpene limonene accumulates to approximately 1 µg/g fw. One reason why the lower molecular weight terpenes don't accumulate is because they are volative and lost to the gas phase. In fact, Wu et al. (2007) reported that sesquiterpene emission of genetically engineered plants probably exceeded that which accumulated by at least 10-fold. A second reason for this could be that these terpene synthases for each type have pronounced different catalytic efficiencies For instance, truncated yeast squalene synthase hasa catalytic efficiency for FPP reflected in Kms equal to 40 µM and Kcats equal to 3.3 S^{-1} (Zhang et al., 1993), relative to values of 4.45±0.56 µM and 0.43×10^{-3} s^{-1} for sesquiterpene synthases PTS (Deguerry et al., 2006). A third reason could be if the introduced metabolism occurs naturally in the compartments being engineered, it may be subject to endogenous regulation. For example, it could be very difficult to elevate GPP pool in the chloroplast for monoterpene production even if a exogenous GPS were used, because GPP biosynthesis that is normally occurring in the chloroplast and could be regulated by plastidic mechanisms, resulting in a relatively low level of monoterpene production. In contrast, engineering FPP, sesquiterpene, and triterpene biosynthesis into the chloroplast inserts a non-native pathway into the chloroplast compartment, which escapes from any native regulation, leads to a higher production of sesquiterpenes and triterpene production. It should also be noted that success of this strategy also replies on the divertion of a carbon flux pathway at an earlier point in the pathway, where the main flux might be monitored by the accumulation of downstream end-products that can feedback to positively enhance/increase overall flux down the pathway.

Metabolic engineering of higher plant chloroplasts offers a great potential for improving isoprenoid yields for many reasons: 1) there are many plastids within each higher plant cell (up to 50) in which their genomes may be directly engineered themselves and

resulting in a large gene dosage effect (up to 1000 copies per plastid); 2) since they are the site of photosynthesis, carbon flux in chloroplasts is robust and theoretically could provide for an unlimited supply of precursors; 3) chloroplasts may offer a good environment for exogenous protein folding, expression and activity; 4) chloroplasts appear especially suitable for heterologous isoprenoid production due to lax endogenous regulation of the MEP pathway as compared to the MVA pathway operating in the cytosol; and 5) there is now the possibility to improve photosynthetic efficiency, which in theory should increase production of engineered compounds (Kebeish et al., 2007).

A recent review by Heinig et al. (2013) discussed the current challenges in conducting subcellular targeting in plant metabolic engineering. They suggested important considerations to ensure substrate availability, whether this can be overcome by co-introducing transporters, upstream catalytic enzymes to increase pathway flux, or suppressing endogenous pathways which bleed away precursors or cofactors. While the most successful subcellular targeting efforts so far have used nuclear-encoded pro-peptides, that include transit peptides that direct the final protein to its intracellular destination, further development and optimization of plastid-encoded enzymes could allow for even higher titers of products. While engineering constructs into the chloroplast has been possible since the late 20th century (McBride et al., 1995; Daniell et al., 1998; Kota et al., 1999), and high protein titers have been obtained, engineering chloroplast-encoded catalytically active enzymes has been less successful. Hence, there appears to be an inherent limitation in the chloroplast for the biosynthesis of high levels of catalytically competent enzymes, a problem that currently doesn't seem to be receiving considerable attention.

Figure of chapter 5

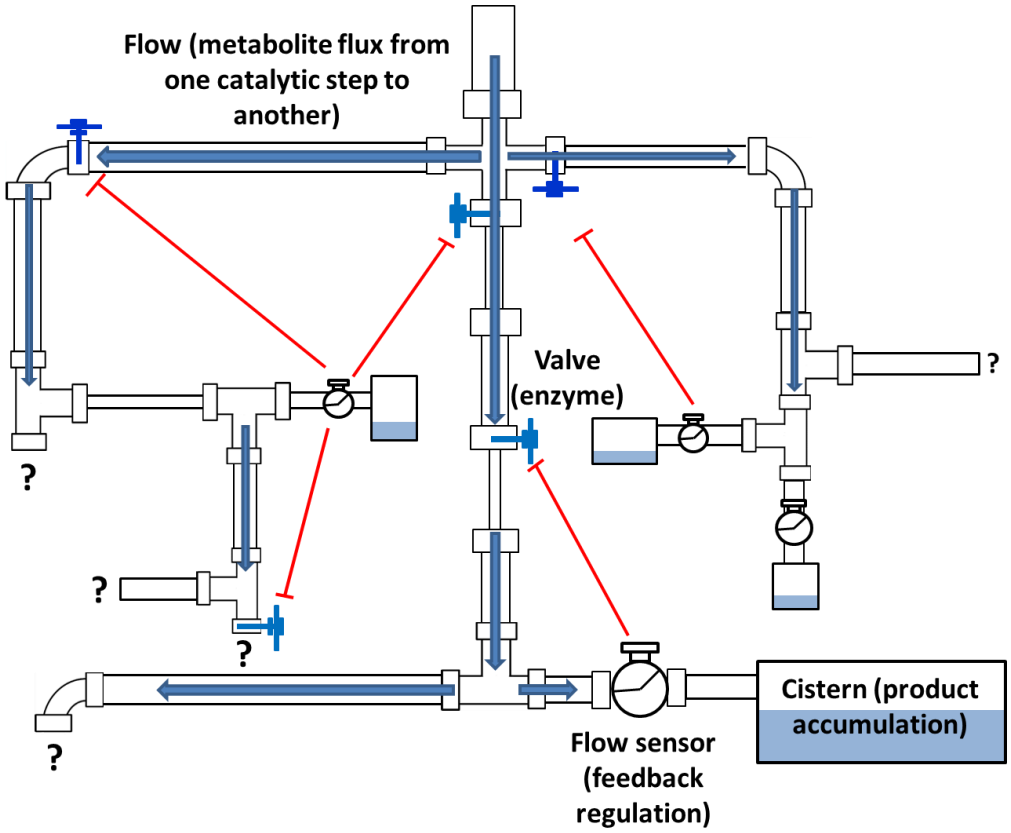


Figure 5.1 Plumbing model and analogy for metabolic flux. One can imagine that carbon flux throughout the cell can be modeled as a system of connected pipes. The pipe junctions represent enzymes and the connecting pipes represent the reactions catalyzed by those enzymes—with thinner pipes representing rate-limiting steps. In this model the water collects in various cisterns which represent the final metabolic product. These cisterns have a certain maximum capacity (maximum amount of end product which can accumulate) and this is monitored by various sensors (feedback mechanisms), which can alter valves (blue tee-shaped objects) that represent regulatory proteins (*e.g.* transcription factors) controlling flux through the various connecting pipes or reactions. However, there are many steps designated by question marks for which we do not know how they might interact with our metabolic network.

Bibliography:

Aharoni A, Giri AP, Deuerlein S, Griepink F, de Kogel W-J, Verstappen FW a, Verhoeven H a, Jongsma M a, Schwab W, Bouwmeester HJ (2003) Terpenoid metabolism in wild-type and transgenic Arabidopsis plants. Plant Cell 15: 2866–84

Ahlert D, Stegemann S, Kahlau S, Ruf S, Bock R (2009a) Insensitivity of chloroplast gene expression to DNA methylation. Mol Genet Genomics 282: 17–24

Ahlert D, Stegemann S, Kahlau S, Ruf S, Bock R (2009b) Insensitivity of chloroplast gene expression to DNA methylation. Mol Genet Genomics 282: 17–24

Ajikumar PK, Xiao W-H, Tyo KEJ, Wang Y, Simeon F, Leonard E, Mucha O, Phon TH, Pfeifer B, Stephanopoulos G (2010) Isoprenoid pathway optimization for Taxol precursor overproduction in Escherichia coli. Science 330: 70–74

Alban C, Tardif M, Mininno M, Brugière S, Gilgen A, Ma S, Mazzoleni M, Gigarel O, Martin-Laffon J, Ferro M, et al (2014) Uncovering the protein lysine and arginine methylation network in Arabidopsis chloroplasts. PLoS One 9: e95512

Andrianov V, Borisjuk N, Pogrebnyak N, Brinker A, Dixon J, Spitsin S, Flynn J, Matyszczuk P, Andryszak K, Laurelli M, et al (2010) Tobacco as a production platform for biofuel: overexpression of Arabidopsis DGAT and LEC2 genes increases accumulation and shifts the composition of lipids in green biomass. Plant Biotechnol J 8: 277–87

Anterola A, Shanle E, Perroud PF, Quatrano R (2009) Production of taxa-4(5),11(12)-diene by transgenic Physcomitrella patens. Transgenic Res 18: 655–660

Bai C, Twyman RM, Farré G, Sanahuja G, Christou P, Capell T, Zhu C (2011) A golden era-pro-vitamin A enhancement in diverse crops. Vitr Cell Dev Biol - Plant 47: 205–221

Banerjee A, Sharma R, Chisti Y, Banerjee UC (2002) Botryococcus braunii: a renewable source of hydrocarbons and other chemicals. Crit Rev Biotechnol 22: 245–279

Bao X, Katz S, Pollard M, Ohlrogge J (2002) Carbocyclic fatty acids in plants: biochemical and molecular genetic characterization of cyclopropane fatty acid synthesis of Sterculiafoetida. Proc Natl Acad Sci U S A 99: 7172–7177

Bao X, Thelen JJ, Bonaventure G, Ohlrogge JB (2003) Characterization of cyclopropane fatty-acid synthase from Sterculia foetida. J Biol Chem 278: 12846–12853

Benfey PN, Ren L, Chua NH (1990) Combinatorial and synergistic properties of CaMV 35S enhancer subdomains. EMBO J 9: 1685–1696

Besumbes Ó, Sauret-Güeto S, Phillips MA, Imperial S, Rodríguez-Concepción M, Boronat A (2004) Metabolic engineering of isoprenoid biosynthesis in Arabidopsis

for the production of taxadiene, the first committed precursor of taxol. Biotechnol Bioeng 88: 168–175

Beyer P (2010) Golden Rice and “Golden” crops for human nutrition. N Biotechnol 27: 478–481

Bouvier F, Linka N, Isner J-C, Mutterer J, Weber APM, Camara B (2006) Arabidopsis SAMT1 defines a plastid transporter regulating plastid biogenesis and plant development. Plant Cell 18: 3088–105

Bouvier F, Rahier A, Camara B (2005) Biogenesis, molecular regulation and function of plant isoprenoids. Prog Lipid Res 44: 357–429

Buckingham, J. Dictionary of Natural Products (Chapman & Hall, London (CD ROM),

(2003).

Chappell J, Nable R (1987) Induction of sesquiterpenoid biosynthesis in tobacco cell suspension cultures by fungal elicitor. Plant Physiol 85: 469–73

Chen J-L, Fang H-M, Ji Y-P, Pu G-B, Guo Y-W, Huang L-L, Du Z-G, Liu B-Y, Ye H-C, Li G-F, et al (2011) Artemisinin biosynthesis enhancement in transgenic Artemisia annua plants by downregulation of the β-caryophyllene synthase gene. Planta Med 77: 1759–65

Choi YE, Lim S, Kim H-J, Han JY, Lee M-H, Yang Y, Kim J-A, Kim Y-S (2012) Tobacco NtLTP1, a glandular-specific lipid transfer protein, is required for lipid secretion from glandular trichomes. Plant J 70: 480–91

Closa M, Vranová E, Bortolotti C, Bigler L, Arró M, Ferrer A, Gruissem W (2010) The Arabidopsis thaliana FPP synthase isozymes have overlapping and specific functions in isoprenoid biosynthesis, and complete loss of FPP synthase activity causes early developmental arrest. Plant J 63: 512–525

Clouse SD (2011) Brassinosteroid signal transduction: from receptor kinase activation to transcriptional networks regulating plant development. Plant Cell 23: 1219–1230

Cragg GM (1998) Paclitaxel (Taxol??): A success story with valuable lessons for natural product drug discovery and development. Med Res Rev 18: 315–331

Croteau RB, Davis EM, Ringer KL, Wildung MR (2005) (-)-Menthol biosynthesis and molecular genetics. Naturwissenschaften 92: 562–77

Curien G, Job D, Douce R, Dumas R (1998) Allosteric activation of Arabidopsis threonine synthase by S- adenosylmethionine. Biochemistry 37: 13212–13221

Daniell H (2006) Production of biopharmaceuticals and vaccines in plants via the chloroplast genome. Biotechnol J 1: 1071–1079

Daniell H, Datta R, Varma S, Gray S, Lee SB (1998) Containment of herbicide resistance through genetic engineering of the chloroplast genome. Nat Biotechnol 16: 345–348

Daniell H, Kumar S, Dufourmantel N (2005) Breakthrough in chloroplast genetic engineering of agronomically important crops. Trends Biotechnol 23: 238–245

Davidovich-Rikanati R, Lewinsohn E, Bar E, Iijima Y, Pichersky E, Sitrit Y (2008) Overexpression of the lemon basil ??-zingiberene synthase gene increases both mono- and sesquiterpene contents in tomato fruit. Plant J 56: 228–238

Deguerry F, Pastore L, Wu S, Clark A, Chappell J, Schalk M (2006) The diverse sesquiterpene profile of patchouli, Pogostemon cablin, is correlated with a limited number of sesquiterpene synthases. Arch Biochem Biophys 454: 123–136

DellaPenna D (2005) A decade of progress in understanding vitamin E synthesis in plants. J Plant Physiol 162: 729–37

Dewick, P.M.. Medicinal Natural Products (Wiley & Sons, New York, 2002).

Dewick PM (2009) Medicinal Natural Products: A Biosynthetic Approach: Third Edition. Med Nat Prod A Biosynthetic Approach Third Ed 1–539

Dhingra V, Vishweshwar Rao K, Lakshmi Narasu M (2000) Current status of artemisinin and its derivatives as antimalarial drugs. Life Sci 66: 279–300

Diretto G, Al-Babili S, Tavazza R, Papacchioli V, Beyer P, Giuliano G (2007) Metabolic engineering of potato carotenoid content through tuber-specific overexpression of a bacterial mini-pathway. PLoS One 2: e350

Diretto G, Tavazza R, Welsch R, Pizzichini D, Mourgues F, Papacchioli V, Beyer P, Giuliano G (2006) Metabolic engineering of potato tuber carotenoids through tuber-specific silencing of lycopene epsilon cyclase. BMC Plant Biol 6: 13

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

Farhi M, Marhevka E, Ben-Ari J, Algamas-Dimantov A, Liang Z, Zeevi V, Edelbaum O, Spitzer-Rimon B, Abeliovich H, Schwartz B, et al (2011) Generation of the potent anti-malarial drug artemisinin in tobacco. Nat Biotechnol 29: 1072–1074

Farré G, Bai C, Twyman RM, Capell T, Christou P, Zhu C (2011) Nutritious crops producing multiple carotenoids - a metabolic balancing act. Trends Plant Sci 16: 532–540

Feng L-L, Yang R-Y, Yang X-Q, Zeng X-M, Lu W-J, Zeng Q-P (2009) Synergistic re-channeling of mevalonate pathway for enhanced artemisinin production in transgenic Artemisia annua. Plant Sci 177: 57–67

Fontecave M, Atta M, Mulliez E (2004) S-adenosylmethionine: nothing goes to waste. Trends Biochem Sci 29: 243–9

Fraser PD, Enfissi EMA, Bramley PM (2009) Genetic engineering of carotenoid formation in tomato fruit and the potential application of systems and synthetic biology approaches. Arch Biochem Biophys 483: 196–204

Gardner RG, Hampton RY (1999) A highly conserved signal controls degradation of 3-hydroxy-3- methylglutaryl-coenzyme a (HMG-CoA) reductase in eukaryotes. J Biol Chem 274: 31671–31678

Giovanelli J, Mudd SH, Datko AH (1985) Quantitative analysis of pathways of methionine metabolism and their regulation in lemna. Plant Physiol 78: 555–560

Glikson M, Lindsay K, Saxby J (1989) Botryococcus—A planktonic green alga, the source of petroleum through the ages: Transmission electron microscopical studies of oil shales and petroleum source rocks. Org Geochem 14: 595–608

Graham IA, Besser K, Blumer S, Branigan CA, Czechowski T, Elias L, Guterman I, Harvey D, Isaac PG, Khan AM, et al (2010) The genetic map of Artemisia annua L. identifies loci affecting yield of the antimalarial drug artemisinin. Science 327: 328–331

Harjes CE, Rocheford TR, Bai L, Brutnell TP, Kandianis CB, Sowinski SG, Stapleton AE, Vallabhaneni R, Williams M, Wurtzel ET, et al (2008) Natural genetic variation in lycopene epsilon cyclase tapped for maize biofortification. Science 319: 330–333

Harvey BG, Wright ME, Quintana RL (2010) High-density renewable fuels based on the selective dimerization of pinenes. Energy and Fuels. pp 267–273

Heinig U, Gutensohn M, Dudareva N, Aharoni A (2013a) The challenges of cellular compartmentalization in plant metabolic engineering. Curr Opin Biotechnol 24: 239–246

Heinig U, Gutensohn M, Dudareva N, Aharoni A (2013b) The challenges of cellular compartmentalization in plant metabolic engineering. Curr Opin Biotechnol 24: 239–246

Hillen LW, Pollard G, Wake L V, White N (1982a) Hydrocracking of the oils of Botryococcus braunii to transport fuels. Biotechnol Bioeng 24: 193–205

Hillen LW, Pollard G, Wake L V, White N (1982b) Hydrocracking of the oils of Botryococcus braunii to transport fuels. Biotechnol Bioeng 24: 193–205

Hong G-J, Hu W-L, Li J-X, Chen X-Y, Wang L-J (2009) Increased Accumulation of Artemisinin and Anthocyanins in Artemisia annua Expressing the Arabidopsis Blue Light Receptor CRY1. Plant Mol Biol Report 27: 334–341

Hong G-J, Xue X-Y, Mao Y-B, Wang L-J, Chen X-Y (2012) Arabidopsis MYC2 Interacts with DELLA Proteins in Regulating Sesquiterpene Synthase Gene Expression. Plant Cell 24: 2635–2648

Horsch RB (1993) COMMERCIALIZATION OF GENETICALLY-ENGINEERED CROPS. Philos Trans R Soc London Ser B-Biological Sci 342: 287–291

Houtz RL, Stults JT, Mulligan RM, Tolbert NE (1989) Post-translational modifications in the large subunit of ribulose bisphosphate carboxylase/oxygenase. Proc Natl Acad Sci U S A 86: 1855–1859

Huang L-C, Hsiao L-J, Pu S-Y, Kuo C-I, Huang B-L, Tseng T-C, Huang H-J, Chen Y-T (2012) DNA methylation and genome rearrangement characteristics of phase change in cultured shoots of Sequoia sempervirens. Physiol Plant 145: 360–8

Huang Z, Dale Poulter C (1989) Tetramethylsqualene, a triterpene from Botryococcus braunii var. showa. Phytochemistry 28: 1467–1470

Huffaker A, Kaplan F, Vaughan MM, Dafoe NJ, Ni X, Rocca JR, Alborn HT, Teal PEA, Schmelz EA (2011) Novel acidic sesquiterpenoids constitute a dominant class of pathogen-induced phytoalexins in maize. Plant Physiol 156: 2082–2097

Inagaki YS, Etherington G, Geisler K, Field B, Dokarry M, Ikeda K, Mutsukado Y, Dicks J, Osbourn A (2011) Investigation of the potential for triterpene synthesis in rice through genome mining and metabolic engineering. New Phytol 191: 432–448

Jander G, Joshi V (2009) Aspartate-Derived Amino Acid Biosynthesis in Arabidopsis thaliana. Arabidopsis Book 7: e0121

Janowski BA, Willy PJ, Devi TR, Falck JR, Mangelsdorf DJ (1996) An oxysterol signalling pathway mediated by the nuclear receptor LXR alpha. Nature 383: 728–731

Kagan RM, Clarke S (1994) Widespread occurrence of three sequence motifs in diverse S-adenosylmethionine-dependent methyltransferases suggests a common structure for these enzymes. Arch Biochem Biophys 310: 417–427

Kappers IF, Aharoni A, van Herpen TWJM, Luckerhoff LLP, Dicke M, Bouwmeester HJ (2005) Genetic engineering of terpenoid metabolism attracts bodyguards to Arabidopsis. Science (80-) 309: 2070–2072

Kebeish R, Niessen M, Thiruveedhi K, Bari R, Hirsch H-J, Rosenkranz R, Stäbler N, Schönfeld B, Kreuzaler F, Peterhänsel C (2007) Chloroplastic photorespiratory bypass increases photosynthesis and biomass production in Arabidopsis thaliana. Nat Biotechnol 25: 593–599

Keeling CI, Bohlmann J (2006) Genes, enzymes and chemicals of terpenoid diversity in the constitutive and induced defence of conifers against insects and pathogens. New Phytol 170: 657–675

Kegge W, Pierik R (2010) Biogenic volatile organic compounds and plant competition. Trends Plant Sci 15: 126–132

Kim SW, Keasling JD (2001) Metabolic engineering of the nonmevalonate isopentenyl diphosphate synthesis pathway in Escherichia coli enhances lycopene production. Biotechnol Bioeng 72: 408–415

Kempinski C, Jiang Z, Bell S, Chappell J. Metabolic engineering of higher plants and algae for isoprenoid production. Adv Biochem Eng Biotechnol. 2015; DOI 10.1007/10_2014_290

Kirby J, Romanini DW, Paradise EM, Keasling JD (2008) Engineering triterpene production in Saccharomyces cerevisiae-beta-amyrin synthase from Artemisia annua. FEBS J 275: 1852–1859

Kohlen W, Ruyter-Spira C, Bouwmeester HJ (2011) Strigolactones: a new musician in the orchestra of plant hormones. Botany 89: 827–840

Kota M, Daniell H, Varma S, Garczynski SF, Gould F, Moar WJ (1999) Overexpression of the Bacillus thuringiensis (Bt) Cry2Aa2 protein in chloroplasts confers resistance to plants against susceptible and Bt-resistant insects. Proc Natl Acad Sci U S A 96: 1840–1845

Kozbial PZ, Mushegian AR (2005) Natural history of S-adenosylmethionine-binding proteins. BMC Struct Biol 5: 19

Krings U, Berger RG (2010) Terpene bioconversion - How does its future look? NatProdComm 5: 1507–1522

Kumar S, Hahn FM, Baidoo E, Kahlon TS, Wood DF, McMahan CM, Cornish K, Keasling JD, Daniell H, Whalen MC (2012) Remodeling the isoprenoid pathway in tobacco by expressing the cytoplasmic mevalonate pathway in chloroplasts. Metab Eng 14: 19–28

Kusano T, Berberich T, Tateda C, Takahashi Y (2008) Polyamines: Essential factors for growth and survival. Planta 228: 367–381

Lange BM, Mahmoud SS, Wildung MR, Turner GW, Davis EM, Lange I, Baker RC, Boydston R a, Croteau RB (2011a) Improving peppermint essential oil yield and composition by metabolic engineering. Proc Natl Acad Sci U S A 108: 16944–16949

Lange BM, Mahmoud SS, Wildung MR, Turner GW, Davis EM, Lange I, Baker RC, Boydston R a., Croteau RB (2011b) Improving peppermint essential oil yield and composition by metabolic engineering. Proc Natl Acad Sci 108: 16944–16949

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

Lee DW, Lee S, Lee G-J, Lee KH, Kim S, Cheong G-W, Hwang I (2006) Functional characterization of sequence motifs in the transit peptide of Arabidopsis small subunit of rubisco. Plant Physiol 140: 466–483

Lewinsohn E, Schalechet F, Wilkinson J, Matsui K, Tadmor Y, Nam KH, Amar O, Lastochkin E, Larkov O, Ravid U, et al (2001) Enhanced levels of the aroma and flavor compound S-linalool by metabolic engineering of the terpenoid pathway in tomato fruits. Plant Physiol 127: 1256–1265

Liscombe DK, Louie G V, Noel JP (2012) Architectures, mechanisms and molecular evolution of natural product methyltransferases. Nat Prod Rep 29: 1238–50

Liscombe DK, Usera AR, O'Connor SE (2010) Homolog of tocopherol C methyltransferases catalyzes N methylation in anticancer alkaloid biosynthesis. Proc Natl Acad Sci U S A 107: 18793–18798

Lu X, Zhang L, Zhang F, Jiang W, Shen Q, Zhang L, Lv Z, Wang G, Tang K (2013) AaORA, a trichome-specific AP2/ERF transcription factor of Artemisia annua, is a positive regulator in the artemisinin biosynthetic pathway and in disease resistance to Botrytis cinerea. New Phytol 198: 1191–1202

Lücker J, Bouwmeester HJ, Schwab W, Blaas J, Van Der Plas LHW, Verhoeven HA (2001) Expression of Clarkia S-linalool synthase in transgenic petunia plants results in the accumulation of S-linalyl-β-D-glucopyranoside. Plant J 27: 315–324

Lücker J, Schwab W, Franssen MCR, Van Der Plas LHW, Bouwmeester HJ, Verhoeven H a. (2004) Metabolic engineering of monoterpene biosynthesis: Two-step production of (+)-trans-isopiperitenol by tobacco. Plant J 39: 135–145

Ma D, Pu G, Lei C, Ma L, Wang H, Guo Y, Chen J, Du Z, Wang H, Li G, et al (2009) Isolation and characterization of AaWRKY1, an Artemisia annua transcription factor that regulates the amorpha-4,11-diene synthase gene, a key gene of artemisinin biosynthesis. Plant Cell Physiol 50: 2146–2161

Mahmoud SS, Croteau RB (2001) Metabolic engineering of essential oil yield and composition in mint by altering expression of deoxyxylulose phosphate reductoisomerase and menthofuran synthase. Proc Natl Acad Sci U S A 98: 8915–8920

Mandel MA, Feldmann KA, Herrera-Estrella L, Rocha-Sosa M, León P (1996) CLA1, a novel gene required for chloroplast development, is highly conserved in evolution. Plant J 9: 649–658

Manzano D, Fernández-Busquets X, Schaller H, González V, Boronat A, Arró M, Ferrer A (2004) The metabolic imbalance underlying lesion formation in Arabidopsis thaliana overexpressing farnesyl diphosphate synthase (isoform 1S) leads to

oxidative stress and is triggered by the developmental decline of endogenous HMGR activity. Planta 219: 982–92

Markus Lange B, Ahkami A (2013) Metabolic engineering of plant monoterpenes, sesquiterpenes and diterpenes-current status and future opportunities. Plant Biotechnol J 11: 169–196

Martin JL, McMillan FM (2002) SAM (dependent) I AM: The S-adenosylmethionine-dependent methyltransferase fold. Curr Opin Struct Biol 12: 783–793

Martin VJJ, Pitera DJ, Withers ST, Newman JD, Keasling JD (2003) Engineering a mevalonate pathway in Escherichia coli for production of terpenoids. Nat Biotechnol 21: 796–802

Masferrer A, Arró M, Manzano D, Schaller H, Fernández-Busquets X, Moncaleán P, Fernández B, Cunillera N, Boronat A, Ferrer A (2002) Overexpression of Arabidopsis thaliana farnesyl diphosphate synthase (FPS1S) in transgenic Arabidopsis induces a cell death/senescence-like response and reduced cytokinin levels. Plant J 30: 123–132

Mastalerz M, Hower JC (1996) Elemental composition and molecular structure of Botryococcus alginite in Westphalian cannel coals from Kentucky. Org Geochem 24: 301–308

McBride KE, Svab Z, Schaaf DJ, Hogan PS, Stalker DM, Maliga P (1995) Amplification of a chimeric Bacillus gene in chloroplasts leads to an extraordinary level of an insecticidal protein in tobacco. Biotechnology (N Y) 13: 362–365

Mehrshahi P, Stefano G, Andaloro JM, Brandizzi F, Froehlich JE, DellaPenna D (2013) Transorganellar complementation redefines the biochemical continuity of endoplasmic reticulum and chloroplasts. Proc Natl Acad Sci U S A 110: 12126–31

Metzger P (1985) Structures of some botryococcenes: branched hydrocarbons from the b-race of the green alga Botryococcus braunii. Phytochemistry 23: 2995–3002

Metzger P, Casadevall E, Coute A (1988) Botryococcene distribution in strains of the green alga Botryococcus braunii. Phytochemistry 27: 1383–1388

Misawa N (2009) Pathway engineering of plants toward astaxanthin production. Plant Biotechnol 26: 93–99

Moldowan JM, Seifert WK (1980) First discovery of botryococcane in petroleum. J Chem Soc Chem Commun 912

Montiel G, Breton C, Thiersault M, Burlat V, Jay-Allemand C, Gantet P (2007) Transcription factor Agamous-like 12 from Arabidopsis promotes tissue-like organization and alkaloid biosynthesis in Catharanthus roseus suspension cells. Metab Eng 9: 125–132

Morandini P (2013) Control limits for accumulation of plant metabolites: Brute force is no substitute for understanding. Plant Biotechnol J 11: 253–267

Munné-Bosch S, Alegre L (2001) Subcellular compartmentation of the diterpene carnosic acid and its derivatives in the leaves of rosemary. Plant Physiol 125: 1094–1102

Muñoz-Bertomeu J, Sales E, Ros R, Arrillaga I, Segura J (2007) Up-regulation of an N-terminal truncated 3-hydroxy-3-methylglutaryl CoA reductase enhances production of essential oils and sterols in transgenic Lavandula latifolia. Plant Biotechnol J 5: 746–758

Niehaus TD, Kinison S, Okada S, Yeo YS, Bell SA, Cui P, Devarenne TP, Chappell J (2012) Functional identification of triterpene methyltransferases from Botryococcus braunii race B. J Biol Chem 287: 8163–8173

Niehaus TD, Okada S, Devarenne TP, Watt DS, Sviripa V, Chappell J (2011) Identification of unique mechanisms for triterpene biosynthesis in Botryococcus braunii. Proc Natl Acad Sci U S A 108: 12260–12265

Nielsen J, Keasling JD (2011) Synergies between synthetic biology and metabolic engineering. Nat Biotechnol 29: 693–695

Nielsen MT, Jones GA, Collins GB (1982) Inheritance pattern for secreting and non-secreting glandular trichomes in tobacco. Crop Sci 22: 1051-1053

Niemi KJ, Adler J, Selman BR (1990) Protein methylation in pea chloroplasts. Plant Physiol 93: 1235–1240

Nishiyama R, Ito M, Yamaguchi Y, Koizumi N, Sano H (2002) A chloroplast-resident DNA methyltransferase is responsible for hypermethylation of chloroplast genes in Chlamydomonas maternal gametes. Proc Natl Acad Sci U S A 99: 5925–5930

Noel JP, Dixon RA, Pichersky E, Zubieta C, Ferrer JL (2003) Chapter two Structural, functional, and evolutionary basis for methylation of plant small molecules. Recent Adv Phytochem 37: 37–58

Nugroho LH, Verpoorte R (2002) Secondary metabolism in tobacco. Plant Cell Tissue Organ Cult 68: 105–125

Ohara K, Matsunaga E, Nanto K, Yamamoto K, Sasaki K, Ebinuma H, Yazaki K (2010) Monoterpene engineering in a woody plant Eucalyptus camaldulensis using a limonene synthase cDNA. Plant Biotechnol J 8: 28–37

Ohara K, Ujihara T, Endo T, Sato F, Yazaki K (2003) Limonene production in tobacco with Perilla limonene synthase cDNA. J Exp Bot 54: 2635–2642

Okada S, Devarenne TP, Chappell J (2000) Molecular characterization of squalene synthase from the green microalga Botryococcus braunii, race B. Arch Biochem Biophys 373: 307–317

Okada S, Murakami M, Yamaguchi K (1995) Hydrocarbon composition of newly isolated strains of the green microalga Botryococcus braunii. J Appl Phycol 7: 555–559

Patra B, Schluttenhofer C, Wu Y, Pattanaik S, Yuan L (2013) Transcriptional regulation of secondary metabolite biosynthesis in plants. Biochim Biophys Acta 1829: 1236–1247

Peralta-Yahya PP, Keasling JD (2010) Advanced biofuel production in microbes. Biotechnol J 5: 147–162

Peralta-Yahya PP, Ouellet M, Chan R, Mukhopadhyay A, Keasling JD, Lee TS (2011) Identification and microbial production of a terpene-based advanced biofuel. Nat Commun 2: 483

Pulis AP, Aggarwal VK (2012) Synthesis of enantioenriched tertiary boronic esters from secondary allylic carbamates. Application to the synthesis of C30 botryococcene. J Am Chem Soc 134: 7570–7574

Ravanel S, Block M a, Rippert P, Jabrin S, Curien G, Rébeillé F, Douce R (2004) Methionine metabolism in plants: chloroplasts are autonomous for de novo methionine synthesis and can import S-adenosylmethionine from the cytosol. J Biol Chem 279: 22548–57

Reeves PH, Ellis CM, Ploense SE, Wu M-F, Yadav V, Tholl D, Chételat A, Haupt I, Kennerley BJ, Hodgens C, et al (2012) A regulatory network for coordinated flower maturation. PLoS Genet 8: e1002506

Renninger, N. & McPhee, D. Fuel compositions comprising farnesane and farnesane derivatives and method of making and using same. US patent 7,399,323 (2008).

Ro D-K, Paradise EM, Ouellet M, Fisher KJ, Newman KL, Ndungu JM, Ho KA, Eachus RA, Ham TS, Kirby J, et al (2006) Production of the antimalarial drug precursor artemisinic acid in engineered yeast. Nature 440: 940–943

Rodríguez A, San Andrés V, Cervera M, Redondo A, Alquézar B, Shimada T, Gadea J, Rodrigo MJ, Zacarías L, Palou L, et al (2011) Terpene down-regulation in orange reveals the role of fruit aromas in mediating interactions with insect herbivores and pathogens. Plant Physiol 156: 793–802

Rodríguez A, Shimada T, Cervera M, Alquézar B, Gadea J, Gómez-Cadenas A, De Ollas CJ, Rodrigo MJ, Zacarías L, Peña L (2014) Terpene down-regulation triggers defense responses in transgenic orange leading to resistance against fungal pathogens. Plant Physiol 164: 321–339

Rude MA, Schirmer A (2009) New microbial fuels: a biotech perspective. Curr Opin Microbiol 12: 274–281

Salim V, Yu F, Altarejos J, De Luca V (2013) Virus-induced gene silencing identifies Catharanthus roseus 7-deoxyloganic acid-7-hydroxylase, a step in iridoid and monoterpene indole alkaloid biosynthesis. Plant J 76: 754–65

Sallaud C, Rontein D, Onillon S, Jabès F, Duffé P, Giacalone C, Thoraval S, Escoffier C, Herbette G, Leonhardt N, et al (2009) A novel pathway for sesquiterpene biosynthesis from Z,Z-farnesyl pyrophosphate in the wild tomato Solanum habrochaites. Plant Cell 21: 301–317

Salvucci ME, Crafts-Brandner SJ (2004) Inhibition of photosynthesis by heat stress: The activation state of Rubisco as a limiting factor in photosynthesis. Physiol Plant 120: 179–186

Sauter M, Moffatt B, Saechao MC, Hell R, Wirtz M (2013) Methionine salvage and S-adenosylmethionine: essential links between sulfur, ethylene and polyamine biosynthesis. Biochem J 451: 145–54

Sawai S, Saito K (2011) Triterpenoid Biosynthesis and Engineering in Plants. Front Plant Sci. doi: 10.3389/fpls.2011.00025

Scheer M, Grote A, Chang A, Schomburg I, Munaretto C, Rother M, Söhngen C, Stelzer M, Thiele J, Schomburg D (2011) BRENDA, the enzyme information system in 2011. Nucleic Acids Res. doi: 10.1093/nar/gkq1089

Schillberg S, Fischer R, Emans N (2003) Molecular farming of recombinant antibodies in plants. Cell Mol Life Sci 60: 433–445

Schmidt-Dannert C, Umeno D, Arnold FH (2000) Molecular breeding of carotenoid biosynthetic pathways. Nat Biotechnol 18: 750–753

Schubert HL, Blumenthal RM, Cheng X (2003) Many paths to methyltransfer: A chronicle of convergence. Trends Biochem Sci 28: 329–335

Schwab W, Davidovich-Rikanati R, Lewinsohn E (2008) Biosynthesis of plant-derived flavor compounds. Plant J 54: 712–732

Seo JW, Jeong JH, Shin CG, Lo SC, Han SS, Yu KW, Harada E, Han JY, Choi YE (2005) Overexpression of squalene synthase in Eleutherococcus senticosus increases phytosterol and triterpene accumulation. Phytochemistry 66: 869–877

Shelar DB, Shirote PJ (2011) Natural product in drug discovery: back to future. Biomed Pharm J 4: 141-146

Siminszky B, Gavilano L, Bowen SW, Dewey RE (2005) Conversion of nicotine to nornicotine in Nicotiana tabacum is mediated by CYP82E4, a cytochrome P450 monooxygenase. Proc Natl Acad Sci U S A 102: 14919–14924

Snell KD, Peoples OP (2009) PHA bioplastic: A value-added coproduct for biomass biorefineries. Biofuels, Bioprod Biorefining 3: 456–467

Takahashi M, Terada Y, Nakai I, Nakanishi H, Yoshimura E, Mori S (2003a) Role of Nicotianamine in the Intracellular Delivery of Metals and Plant Reproductive Development. 15: 1263–1280

Takahashi M, Terada Y, Nakai I, Nakanishi H, Yoshimura E, Mori S, Nishizawa NK (2003b) Role of nicotianamine in the intracellular delivery of metals and plant reproductive development. Plant Cell 15: 1263–1280

Takahashi S, Yeo Y, Greenhagen BT, McMullin T, Song L, Maurina-Brunker J, Rosson R, Noel JP, Chappell J (2007) Metabolic engineering of sesquiterpene metabolism in yeast. Biotechnol Bioeng 97: 170–181

Tarshis LC, Yan M, Poulter CD, Sacchettini JC (1994) Crystal structure of recombinant farnesyl diphosphate synthase at 2.6-A resolution. Biochemistry 33: 10871–10877

Tholl D, Chen F, Petri J, Gershenzon J, Pichersky E (2005) Two sesquiterpene synthases are responsible for the complex mixture of sesquiterpenes emitted from Arabidopsis flowers. Plant J 42: 757–771

Tissier A (2012) Trichome Specific Expression : Promoters and Their Applications. *In* YÖ Çiftçi, ed, Transgenic Plants-Advances and Limitations. INTECH, pp 353–378

Trievel RC, Flynn EM, Houtz RL, Hurley JH (2003) Mechanism of multiple lysine methylation by the SET domain enzyme Rubisco LSMT. Nat Struct Biol 10: 545–552

Ubersax, J. & Platt, D. Genetically modified microbes producing isoprenoids patent world 1,414,52 A1 (2010).

Verdaguer B, de Kochko A, Beachy RN, Fauquet C (1996) Isolation and expression in transgenic tobacco and rice plants, of the cassava vein mosaic virus (CVMV) promoter. Plant Mol Biol 31: 1129–1139

Verdaguer B, De Kochko A, Fux CI, Beachy RN, Fauquet C (1998) Functional organization of the cassava vein mosaic virus (CsVMV) promoter. Plant Mol Biol 37: 1055–1067

Wagner GJ, Wang E, Shepherd RW (2004) New approaches for studying and exploiting an old protuberance, the plant trichome. Ann Bot 93: 3–11

Wallaart TE, Bouwmeester HJ, Hille J, Poppinga L, Maijers NCA (2001) Amorpha-4,11-diene synthase: Cloning and functional expression of a key enzyme in the biosynthetic pathway of the novel antimalarial drug artemisinin. Planta 212: 460–465

Wang C-T, Liu H, Gao X-S, Zhang H-X (2010) Overexpression of G10H and ORCA3 in the hairy roots of Catharanthus roseus improves catharanthine production. Plant Cell Rep 29: 887–894

Wang E, Gan S, Wagner GJ (2002a) Isolation and characterization of the CYP71D16 trichome-specific promoter from Nicotiana tabacum L. J Exp Bot 53: 1891–1897

Wang KL, Li H, Ecker JR (2002b) Ethylene Biosynthesis and Signaling Networks. Plant Cell 14: 131–152

Westfall PJ, Pitera DJ, Lenihan JR, Eng D, Woolard FX, Regentin R, Horning T, Tsuruta H, Melis DJ, Owens a., et al (2012) From the Cover: PNAS Plus: Production of amorphadiene in yeast, and its conversion to dihydroartemisinic acid, precursor to the antimalarial agent artemisinin. Proc Natl Acad Sci 109: E111–E118

Withers ST, Gottlieb SS, Lieu B, Newman JD, Keasling JD (2007) Identification of isopentenol biosynthetic genes from Bacillus subtilis by a screening method based on isoprenoid precursor toxicity. Appl Environ Microbiol 73: 6277–6283

Wu S, Jiang Z, Kempinski C, Eric Nybo S, Husodo S, Williams R, Chappell J (2012) Engineering triterpene metabolism in tobacco. Planta 236: 867–877

Wu S, Schalk M, Clark A, Miles RB, Coates R, Chappell J (2006) Redirection of cytosolic or plastidic isoprenoid precursors elevates terpene production in plants. Nat Biotechnol 24: 1441–1447

Wurtzel ET, Cuttriss A, Vallabhaneni R (2012) Maize Provitamin A Carotenoids, Current Resources, and Future Metabolic Engineering Challenges. Front Plant Sci. doi: 10.3389/fpls.2012.00029

Ying Z, Mulligan RM, Janney N, Houtz RL (1999) Rubisco small and large subunit N-methyltransferases. Bi- and mono- functional methyltransferases that methylate the small and large subunits of Rubisco. J Biol Chem 274: 36750–36756

Yu B, Lydiate DJ, Young LW, Schäfer U a, Hannoufa A (2008) Enhancing the carotenoid content of Brassica napus seeds by downregulating lycopene epsilon cyclase. Transgenic Res 17: 573–585

Yu Z-X, Li J-X, Yang C-Q, Hu W-L, Wang L-J, Chen X-Y (2012) The jasmonate-responsive AP2/ERF transcription factors AaERF1 and AaERF2 positively regulate artemisinin biosynthesis in Artemisia annua L. Mol Plant 5: 353–365

Yusibov V, Streatfield SJ, Kushnir N (2011) Clinical development of plant-produced recombinant pharmaceuticals: vaccines, antibodies and beyond. Hum Vaccin 7: 313–321

Zhang D, Jennings SM, Robinson GW, Poulter CD (1993) Yeast squalene synthase: expression, purification, and characterization of soluble recombinant enzyme. Arch Biochem Biophys 304: 133–143

Zhang L, Jing F, Li F, Li M, Wang Y, Wang G, Sun X, Tang K (2009) Development of transgenic Artemisia annua (Chinese wormwood) plants with an enhanced content of artemisinin, an effective anti-malarial drug, by hairpin-RNA-mediated gene silencing. Biotechnol Appl Biochem 52: 199–207

Zook M, Hohn T, Bonnen A, Tsuji J, Hammerschmidt R (1996) Characterization of Novel Sesquiterpenoid Biosynthesis in Tobacco Expressing a Fungal Sesquiterpene Synthase. Plant Physiol 112: 311–318

Zubieta C, Ross JR, Koscheski P, Yang Y, Pichersky E, Noel JP (2003) Structural basis for substrate recognition in the salicylic acid carboxyl methyltransferase family. Plant Cell 15: 1704–1716

Vita

1) Master in Plant Pathology, China Agricultural University, Beijing, China Jul. 2009

Bachelor in Plant Protection at Shenyang Agricultural University, Shenyang, China, Jul. 2006

2) No professional positions held.

3) Graduate research assistant in University of Kentucky, 2009-2015

4) Publications:

1. Kempinski C, Jiang Z, Bell S, Chappell J. Metabolic engineering of higher plants and algae for isoprenoid production. Adv Biochem Eng Biotechnol. 2015; DOI 10.1007/10_2014_290

2. Wu S, Jiang Z, Kempinski C, Eric Nybo S, Husodo S, Williams R, Chappell J. Engineering triterpene metabolism in tobacco. Planta. 2012; 236(3):867-77

3. Pathak KB1, Jiang Z, Ochanine V, Sharma M, Pogany J, Nagy PD. Characterization of dominant-negative and temperature-sensitive mutants of tombusvirus replication proteins affecting replicase assembly. Virology. 2013; 437(1):48-61

4. Li Y, Jia M, Jiang Z, Zhou T, Fan Z. Molecular variation and recombination in RNA segment 10 of rice black-streaked dwarf virus isolated from China during 2007-2010. Arch Virol. 2012; 157(7):1351-6