METABOLIC ENGINEERING TO ENHANCE TERPENE YIELD IN NICOTIANA TABACUM

A Dissertation

by

CHENG ZHAO

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Joshua S. Yuan
Committee Members,	Martin B. Dickman
	Timothy Devarenne
	Clint W. Magill
Head of Department,	Leland S. Pierson III

August 2018

Major Subject: Plant Pathology

ABSTRACT

Terpenoids provide a diverse range of applications in fuels, chemicals, specialty materials, nutraceuticals, and pharmaceuticals. Terpenes also provide unconventional sinks for the carbon fixed by photosynthesis. However, it is highly challenging to increase terpene production in higher plants, partially due to downstream conversion, toxicity to the cell and feedback regulations. My thesis research has mainly focused on developing strategies to address the challenges of terpene production in *Nicotiana tabacum*. First, a co-compartmentation strategy was design and implemented, in which biosynthesis and storage were co-compartmentalized via a synthetic droplet as an effective new strategy to improve the bioproduct yield, with squalene as a model compound. Instability of squalene was demonstrated by mathematic modeling and dark treatment experiments. Thus, a hydrophobic protein was designed and introduced into the tobacco chloroplast to generate a synthetic droplet for terpene storage. Simultaneously, squalene biosynthesis enzymes were introduced to chloroplasts together with the droplet-forming protein to co-compartmentalize the biosynthesis and storage of squalene. The strategy has enabled a significant increase of squalene yield without compromising plant growth. Confocal fluorescent microscopy imaging, Stimulated Raman Scattering microscopy, and droplet composition analysis confirmed the formation of synthetic storage droplets in chloroplasts. Second, an inter-compartmental 'pull and block' strategy was established to re-balance squalene biosynthesis and degradation via silencing squalene epoxidases. An actively expressed squalene epoxidase SQE3 was identified in the leaf of *N. tabacum*. An artificial microRNA was designed to target the consensus coding sequence of SQE3 to reduce the

downstream conversion of squalene. Simultaneously, the squalene biosynthesis pathway was over-expressed in *N. tabacum* chloroplasts. Suppression of SQE3 was confirmed by RT-PCR. Squalene conversion decrease in SQE3-suppressed plants was confirmed by measuring sterol levels by GC-MS. The squalene yield in SQE3 suppressed plants has achieved 3.98 mg/g fresh weight without comprising plant growth.

Overall, these studies established efficient photosynthetic platforms for terpene production. A sufficient sink capacity and a minimized downstream conversion was emphasized in the studies for high yield of bioproducts. The technology advances also provided new strategies for stabilizing high value bioproducts.

ACKNOWLEDGEMENTS

CONTRIBUTORS AND FUNDING SOURCES

NOMENCLATURE

35S	35S Cauliflower mosaic viral promoter
3D	3-Dimensional
acetyl-CoA	Acetyl coenzyme A
amiRNA	Artificial microRNA
ATP	Adenosine Triphosphate
BLAST	Basic Local Alignment Search Tool
cDNA	Complementary DNA
C-terminus	Carboxyl-terminus
DMAPP	Dimethylallyl Diphosphate
DMSO	Dimethyl Sulfoxide
DNA	Deoxyribonucleic Acid
DW	Dry Weight
ER	Endoplasmic Reticulum
FAME	fatty acid methyl ester
FPP	farnesyl pyrophosphate
G3P	glyceraldehyde 3-phosphate
GC-MS	Gas chromatography–mass spectrometry
GFP	Green Fluorescent Protein
IPP	Isopentenyl Diphosphate
MEcPP	Methylerythritol Cyclodiphosphate

MEP	2-C-methyl-D-erythritol 4-phosphate
MS	Murashige and Skoog
MVA	Mevalonate
N.D.	Not Detectable
NADH	Nicotinamide Adenine Dinucleotide
NADPH	Nicotinamide Adenine Dinucleotide Phosphate
nLC-LTQ	Nanoflow liquid chromatography Ion Trap-Orbitrap
N-terminus	Amino-terminus
PBS	Phosphate-Buffered Saline
PCR	Polymerase Chain Reaction
Pcv	Cassava Vein Mosaic Viral Promoter
PTS	Patchoulol Synthase
Pyr	Pyruvate
RNA	Ribonucleic Acid
RT-PCR	Reverse Transcription Polymerase Chain Reaction
RuBisCO	Ribulose-1,5-bisphosphate carboxylase/oxygenase
SDS-PAGE	Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis
SQE	Squalene Epoxidase
SQS	Squalene Synthase
SRS	Stimulated Raman Scattering
Tnos	*nos* Terminator
TP	Transit Peptide

WT   Wildtype

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION AND LITERATURE REVIEW

1.1 Opportunities and challenges of metabolic engineering of sink capacity in plants

      Plants and algae have the sink capability to accumulate diverse natural products synthesized from atmospheric carbon dioxide using sunlight energy. Plant secondary metabolites have been exploited as medicines, fragrances and foods supplements throughout civilization[1]. For instance, artemisinin has been found in plants documented in Chinese medicine book for treatment of malaria[2]. There are many compounds that can be biomanufactured in autotrophic organisms, such as rubber (polyterpenoids), and the anticancer drugs Taxol (diterpene)[3] and vinblastine[4]. More opportunities are to be mined the field of plant metabolic engineering, especially in photosynthetic environments, where plants contain a large spectrum of specialized metabolism[5]. Moreover, plants and green algae have different compartments to control biosynthesis, which is not replicable in prokaryotic organisms[6].

      With mature agriculture practices and rapid developing algae production techniques, biomanufacturing of natural products from CO2 and sunlight represents a sustainable, environmental friendly and cost-effective alternative compared with organic chemical manufacturing from petroleum[7,8]. Nevertheless, large scale industrial production is hindered by the relatively low titer of the metabolites *in planta*, as plants and algae evolve to produce a large spectrum of metabolites for sessile life, instead of for production of one target compound. Thus, metabolic engineering is necessary to alter cell properties into an efficient production system for target compounds. However, there are inherent challenges for engineering terpenes. First, terpene sink capacity is limited by the presence

downstream/modification pathways. For example, squalene, a triterpene, the precursor for sterol biosynthesis, is not accumulated in most of organisms and not accumulated even with overexpression of squalene synthases in cytosol[9]. Second, terpene biosynthesis pathways are highly regulated[10]. Third, toxicity of intermediates and bioproducts may cause intolerance of the cells. For example, the over-accumulation of methylerythritol cyclodiphosphate (MEcPP), the 2-C-methyl-D-erythritol 4-phosphate (MEP) pathway intermediate, led to global stress response and growth defects[11].

Among different engineering strategies, engineering metabolic sink capacity to increase the potential maximal yield of a metabolite has become critical for bioproduct yield. Classic metabolic engineering of terpene production focuses on the biosynthesis part such as identification of new terpene synthases, carbon repartition toward terpene biosynthesis and heterologous overexpression, etc[12,13]. Insufficient sink capacity can lead to a ceiling of bioproduct titer, which may be the result of feedback based down-regulation of upstream biosynthesis and activation of downstream conversion[14-16]. One critical prospect for altering sink capacity is engineering the accessibility of bioproducts to downstream conversion pathways and feedback signaling molecules. Another prospect is to manipulate biosynthesis and conversion pathways to enhance the yield of product.

This review is focused on the development of strategies to expand terpene sink in photosynthetic systems. We describe recent examples how researchers addressed the current challenges. This review covers the following topics: First, recent advances in metabolic engineering strategies to improve terpene biosynthesis, including increasing precursor and cofactor supply, increasing enzyme specificity and activity and manipulating transcriptional factors and heterologous over-expression of terpene synthases. Second,

recent advances in metabolic engineering strategies to prevent loss of target products, including knockdown and knockout of downstream enzymes, improved storage techniques and terpene production in special tissues and organs. Finally, integration of "omics" technologies and modeling to engineer higher titers of terpenes.

1.2 Background of terpene biosynthesis in plants

Terpenes (hydrocarbons based on C5 units) and Terpenoids (compounds with functional groups added to terpenes, which in this review are also referred to as terpenes) are the most structurally diverse group of natural products. Terpenes have broad functions and applications such as fuels, chemicals, fragrance, nutraceuticals, and pharmaceuticals[7,17]. For example, a valuable therapeutic agent paclitaxel, isolated from *Taxus brevifolia*, is used as chemical medication for a number of cancers.[18] Artemisinin, originally found in *Artemisia annua,* is used as an antimalaria medicine.[19] Other medications include squalene, which has been used as an adjuvant for H5N1 vaccines.[20] In nature, terpenes also have essential functions, from serving as defense signaling molecules to light harvesting pigments. [21,22] Terpenes are biosynthesized from 5-carbon common precursors, IPP (isopentenyl diphosphate) and DMAPP (dimethylallyl diphosphate), which can be produced in two pathways: the MVA (mevalonate) pathway in eukaryotes and the MEP (2-*C*-methyl-D-erythritol 4-phosphate) pathway in prokaryotes[7]. In higher plants, the MVA pathway and MEP pathway co-exists in plant cells but function in different compartments. The MEP pathway is located in plastids and is mainly responsible for the generation of monoterpenes (10-carbon), diterpenes (20-carbon) and tetraterpenes(40-carbon), whereas the MVA pathway is located in the cytosol, and is predominately responsible for the biosynthesis of sesquiterpene (15-carbon) and triterpene (30-carbon).

The first step of the MVA pathway condenses two acetyl-CoA molecules which are made

from pyruvate, the end product of the glycolysis pathway. Notably, the MEP pathway in

plastids directly incorporates fixed carbon into the terpene skeleton using ATP and

NADPH from light dependent reaction[23]. The MEP pathway is initiated with a

condensation reaction between pyruvate and glyceraldehyde 3-phosphate(G3P), one of the

intermediates of the light-independent reactions which occur when light is available.

Though the MVA and MEP pathways both produce common precursors (IPP/DMAPP),

the carbon flux through those pathways may vary significantly, depending on availability

of the substrates (acetyl-CoA; pyruvate and G3P), energy (ATP) and reducing equivalents

(NADH/NADPH) [24].

1.3 Engineering terpene biosynthesis is basic to establish a product sink.

Sink capacity requires a sufficient terpene biosynthesis rate to maintain the titers.

The supply of IPP and DMAPP determines the maximum flux of terpene biosynthesis.

There are two strategies to increase IPP or DMAPP supply, which are optimization of

endogenous MEP/MVA pathway and installing MEP/MVA pathway to new hosts. Efforts

has been made to enhance IPP/DMAPP flux to increase terpene yield.

In the MEP pathway, 1-deoxy-d-xylulose-5-phosphate synthase (DXS) was

suggested to be the rate limiting step. Overexpressing DXS in *Arabidopsis thaliana*

showed significant accumulation of various isoprenoids[25,26]. In addition, DXS was found to

be negatively regulated by IPP and DMAPP[27]. IPP and DMAPP were revealed to compete

with thiamine pyrophosphate for binding with DXS[28], which suggested the

overaccumulation of IPP/DMAPP could lead to downregulation of overall terpene carbon

flux. Recently, limited terpene synthase activity was discovered to be a key deterrent to

4

turn on carbon flux toward terpene biosynthesis[29]. The concept was demonstrated by a study, in which computational kinetic modeling predicting that limonene yield would increase linearly if limonene synthase activity increased 100-fold, and the model was validated by showing limonene yield positively correlated to different expression level of limonene synthase driven by different promoters and ribosome binding sites [30]. . Overall, the findings suggested that an efficient terpene synthase is crucial for enhancing total IPP/DMAPP flux.

In additional to tackling the bottlenecks in the endogenous MEP pathway, an ambitious strategy was developed to introduce a heterologous MVA pathway to bypass normal regulation. In a pioneer study, the whole yeast MVA pathway and amorpha-4,11-diene synthase were installed in *E.coli*, resulting in a considerable increase of production[31]. A similar strategy was tested in cyanobacteria; the MVA pathway was expressed in cyanobacteria with isoprene synthase and isoprene production increased 2.5 fold compared with isoprene only strains[32]. Mevalonate pathway was also introduced into tobacco plants by chloroplast transformation, leading to accumulation of multiple terpene derivatives, showing the potential for future engineering terpene production with an alternate terpene synthase[33]. Again, two keys strategies to expand terpene flux involve increasing IPP/DMAPP pools and more efficient terpene synthesis.

1.4 Downstream modification and degradation of terpenes

In addition to optimization of biosynthesis, successful engineering must consider harvesting and preventing loss of synthesized products. Possibilities include altering physicochemical properties of terpenes and localization of terpene products to expend the terpene sink. For example, engineering highly volatile terpenes in plant leaves may result

in difficulties of harvesting and quantification of terpene yield.[34] The importance of manipulating downstream conversion was highlighted in a study of engineering limonene production. A basic strategy to avoid the degradation is to suppress degradative enzyme activity. In the study, limonene content in essential oils increased 40 fold to 80% with suppression of limonene hydrolases[35].

The classic approach to alleviate downstream degradation and modification is to combine downstream gene knockouts and RNA silencing. Successful attempts have been made in both prokaryotic and eukaryotic photosynthetic systems. For example, squalene hopene cylase in *Synechocystis* sp. PCC6803 was knocked out, resulting in accumulation of squalene to 0.67 mg OD750^{-1}L^{-1}, 70-times higher than wild type which occurred without growth deficiency[36]. In the green alga *Chlamydomonas reinhardtii*, in which squalene is synthesized and oxidized in the cytosol, squalene epoxidases were silenced, leading to around 2 ug/mg cell dry weight, while overexpressing squalene synthase did not lead to overaccumulation of squalene with or without squalene epoxidase knockdown[37]. However, inactivation of squalene epoxidase1 in Arabidopsis lead to a lethal phenotype due to root development deficiency[9]. In the Arabidopsis system, even partial mutation of squalene epoxidases may lead to drought sensitive phenotype, indicating the plasticity of the squalene biosynthesis pathway is not high, which requires accurate control of the gene expression level[38]. With recent development of Crispr/Cas9 system, single point mutations and in-situ gene replacement became possible[39,40]. Thus, approaches to address the issue include knockout and knockdown of downstream degradation/modification pathways, compartmentalization of pathways and products and production in special tissues and

organs. The recent success with the approaches indicate that loss of products must be considered.

1.5 Localization of terpene production and storage techniques

Compartmentalization of synthetic pathways is a new dimension to consider besides linear design of any terpene metabolic pathway. In eukaryotic cells, biosynthetic reactions are divided by membranes enclosed organelles. Especially in higher plants, metabolisms of terpenes are highly compartmentalized in subcellular or intercellular levels. Besides inactivation of downstream enzymes, compartmentalization of pathways and products has emerged as an attractive strategy which can keep products out of the reach of downstream enzymes[41]. As mentioned above, the MVA pathway exits in the cytosol is predominantly responsible for the production of sesquiterpenes and triterpenes while the MEP pathway exits in plastids where it is mainly responsible for production of monoterpenes, diterpenes and tetraterpenes. One of advantages to compartmentalizing the terpene synthesis pathway is avoiding endogenous downstream of products. A typical approach to compartmentalize a pathway is to install a heterologous pathway into a compartment that contains common precursors. For instance, C15 farnesyl pyrophosphate synthase and a sesquiterpene PTS were overexpressed and targeted to plastid expression, resulting in high accumulation of patchnol and monoterpenes being produced in the cytosol for overall higher accumulation[42]. A similar study was performed on the triterpene squalene. In a recent study, artemisinin biosynthesis pathways were introduced to tobacco plastids and reaching 0.8 mg/g dry weight artemisinin which was sufficient for clinical applications[43].

Given that terpene synthesis pathway is compartmentalized, it is still questioned whether synthesized terpenes can be well protected within a cell compartment[44]. Terpenes are widely used as membrane penetration enhancers and vaccine adjuvants due to their capability of diffusing through a phospholipid bilayer membrane. Moreover, accumulation of terpene can result in feedback inhibition and detrimental effects on photosynthesis[15]. Therefore, it is pressing to develop storage techniques to further enhance terpene yield. In other studies, engineering the triacylglycerol synthesis pathway together with oleosin, a lipid droplet structural protein showed significant increase of lipid content in leaf[45]. In a recent study, a synthetic lipid droplet was designed based on the lipophilicity of squalene and the droplets were engineered within chloroplast of *N. tabacum*, leading to more than 2 fold increase in squalene yield to 2.6 mg/g fresh weight compared with squalene pathway compartmentalization-only plants[46]. Though the subcellular storage technique is a new area for bioproducts production, exploring this area as a means of reducing loss of product along with rebalancing biosynthesis and degradation will eventually expand the sink capacity of natural products.

2. CO-COMPARTMENTATION OF TERPENE BIOSYNTHESIS AND STORAGE VIA SYNTHETIC DROPLET*

2.1 Introduction

The classical approach of natural product engineering has mainly focused on pathway and enzyme optimization to enhance the metabolic flux toward target compounds, such as terpenoids[47]. Terpenoids are the largest group of natural products, with broad applications as fuels, chemicals, specialty materials, nutraceuticals, and pharmaceuticals[7,17]. Despite progress in engineering terpenoid biosynthesis, several inherent challenges have limited the further increase of bioproduct yield. First, many terpene compounds cannot accumulate to high levels due to the existence of downstream conversion pathways. For example, squalene production in plants, bacteria, and yeast is often hampered due to downstream modification by enzymes like hopene cyclase and squalene epoxidases[9,36]. Second, terpene biosynthesis is subject to extensive regulation, where the accumulation of end-product and intermediates often lead to feedback inhibition that inactivates key enzymes, down-regulates pathway gene expression, and even impacts the cell growth and physiology[24,48]. Finally, accumulation of certain terpene compounds can be toxic to cells[49].

In fact, nature has evolved mechanisms that address these challenges by storing terpenes in special plant structures, such as glandular trichomes and vascular tissues[50-52]. However, enhancing terpene production by over-expressing *FPS* (farnesyl pyrophosphate synthase) and *SQS* (squalene synthase) in the plastids in trichome led to mosaic and dwarf phenotypes[53]. Moreover, even though the compartmentation of the squalene biosynthetic

pathway in plastids could decrease the downstream conversion to a certain degree, squalene could still 'leak' out of the permeable chloroplast membrane according to Fick's law and Overton's Rule, and be oxidized by the downstream squalene epoxidases[54]. Squalene accumulation will thus heavily depend on available intracellular storage technologies.

Inspired by the natural lipid storage mechanisms, we designed and established a synthetic storage droplet for the tobacco chloroplasts that accumulates squalene by reducing product conversion, removal, and toxicity. It is well established that oleosin plays an essential role in creating the topology for phospholipid monolayers and forms the lipid droplets that maintain the hydrophobic environment for neutral lipid storage[55]. Previous studies have demonstrated that over-expressing oleosin protein can enhance lipid droplet formation and increase lipid accumulation in plant seeds and leaves[45,56]. However, no study has shown that high value terpene accumulation can be enhanced by introducing lipid droplets. In fact, the strategies for enhancing lipid yield cannot be readily translated into terpene accumulation because the target compound biosynthesis needs to be co-located in the same compartment with lipid droplet. In plants, oleosin is generally targeted to ER, where the triacylglycerol is produced and lipid droplet is formed[57]. A successful design would require the co-localization of terpene biosynthesis and droplet formation for squalene storage to avoid the downstream modification of squalene epoxidase in the cytosol.

In this section, I first demonstrate the potential leakage of squalene from chloroplasts with mathematical modeling and experimental verification. A chloroplast-targeted hydrophobic protein was designed to enable the formation of synthetic droplets

and co-compartmentation of biosynthesis and storage of squalene in plastids. The co-compartmentation strategy led to a significant increase of squalene accumulation at a record level without compromising plant growth and photosynthesis. The new strategy for co-compartmentation of biosynthesis and storage thus opened new avenues for enhancing the target bioproduct yield.

2.2 Material and methods

2.2.1 Plasmids construction

The codon optimized sequences of *FPS* (GI:3915686) and *SQS* (GI:729468) were synthesized with the chloroplast transit peptide from *A. thaliana* at the N-terminus of *FPS* and *SQS*. DNA fragment of *FPS* and *SQS* were ligated to downstream of 35S promoter and Cassava vein mosaic virus promoter respectively, and upstream of the *nos* terminator via overlap PCR. The PCR products were subsequently incorporated by Gibson assembly (NEW ENGLAND Biolab) into pCAMBIA2300 that was digested with *EcoRI* and *HindIII* to generate plant binary expression vector *FPS-SQS*.[58]

The DNA fragment containing the coding sequence of *oleosin* (GI:30686782) was cloned from genomic DNA of *A. thaliana* by PCR. Truncation of *oleosin* was performed by PCR with primers HP-F1 (5'-ATGAAAGCTGCAACTGCTGTCAC-3') and HP-R1 (5'-TTAGTAAATCCAAGAGAAAACG-3'). The chloroplast transit peptide of *RuBisCO* small subunit from *Solanum tuberosum* was fused to the N-terminus of *oleosin* and *HP* respectively, and the fragments were subsequently fused to the C-terminus of *GFP* to generate DNA fragments of *HPG* and *OG* by overlapping PCR. Four versions of oleosin were ligated to the 35S promoter and terminator of octopine synthase using overlap PCR. The PCR products were inserted into FPS-SQS to generate FPS-SQS-HP, FPS-SQS-HPG,

FPS-SQS-OP and FPS-SQS-OG by Gibson Assembly. All vector constructs were confirmed by DNA sequencing.

2.2.2   Plant material and transformation

Tobacco plants (*Nicotiana tabacum* line Ti1068) were transformed with *Agrobacterium tumefaciens* GV3101 containing FPS-SQS, FPS-SQS- HP, FPS-SQS-HPG, FPS-SQS-OP and FPS-SQS-OG using standard leaf disc tobacco transformation method, followed by selection of 25 mg/L antibiotic G418 for 2 weeks and 75 mg/L G418 for 2 weeks. Ten to thirty independent positive transformants were generated per construct. Positive transformants were screened by measuring squalene content, and verified by PCR and RT-PCR (not shown). T1 progeny were germinated on Murashige and Skoog (MS) medium with 50 mg/L G418. Plants were grown in standard greenhouse condition with 16 h light /8 h darkness cycle.

2.2.3   Confocal microscopy

Oil droplets were visualized by confocal laser scan microscopy (Olympus FV1000 confocal microscope). Nile red staining was carried out following the protocols as following: Plant tissue containing epidermis and mesophyll cells were peeled from the abaxial side of tobacco leaves, and vaccum infiltrated for 5 min in phosphate-buffered saline (PBS) solution containing 5 mg/L Nile red. The stained tissues were rinsed with PBS 3 times and observed under confocal microscopy (Olympus FV1000 confocal microscope).

GFP fluorescence was stimulated at 488 nm, and the emission signal was collected at 510 nm. Nile red signal was excited at 543 nm, with emission at 603 nm. For chloroplast imaging, chlorophyll autofluorescence was stimulated at 488 nm with emission at 650 nm.

12

2.2.4   Squalene analysis

0.5 g fresh leaf samples were ground to fine powder in liquid nitrogen. The ground samples were extracted using 3 mL hexane, with 15 µg of α-cedrene as the internal standard for 2 hours, followed by centrifugation at 3000g for 5 min. To purify squalene, 1 mL of supernatant was added to a silica gel column, and washed with hexane until the pigments reached the end of column. The purified eluate was concentrated to 6 mL by a stream of nitrogen gas prior to a gas-chromatography mass spectrometry (SHIMADZU & GCMS-QP2010 SE) analysis using standard protocols. Briefly, 1 µL of elute was analyzed by GC-MS with SH-Rxi-5Sil MS column in split mode. Helium was used as the carrier gas. Injection temperature was set to 220.0 °C. The GC-MS program was first held at 40 °C for 30 second, then increased at a rate of 20 °C min-1 to 120 °C, then at a rate of 6 °C min-1 to 200 °C, then at a rate of 20 °C min-1 to 260 °C, then at a rate of 5 °C min-1 to 310 °C, and finally hold for 3 min.

2.2.5   Droplet isolation

Tobacco chloroplasts were isolated from fresh leaves as previously described[59]. Isolated chloroplasts were subsequently diluted in an extraction buffer (100 mM sodium phosphate buffer, pH 7.2 300 mM NaCl, and 1.2 M sucrose; 10 mg of cetyl-trimethyl-ammonium bromide) and vortexed vigorously, followed by centrifugation at 10,000 g for 10 min at 24 °C. The fat pad was collected for squalene analysis[45].

2.2.6   Treatment in darkness

Six-week-old plants were transferred from the greenhouse to a growth chamber with 16 hour light /8 hour dark cycle, at 30 ◦C, 400 µmol·m^{-2}·s^{-1} light intensity and 40% humidity. After 3 weeks in chamber, FPS-SQS line and FPS-SQS-HPG line were grouped

as light or dark group respectively. For plants in the dark group, fully expanded lower leaves were covered in darkness while other leaves were exposed to light. For plants in light group, all leaves were exposed to light. Squalene content in tobacco leaf was measured with or without darkness treatment. Around 200 mg of leaf discs were collected in the same leaf on 0, 3, 6, 9, 12 and 15 days after covering leaves. The leaf discs were stored in -80 °C for further squalene analysis.

2.2.7 Photosynthesis and growth measurement

The measurement of CO_2 assimilation rate was conducted using a Licor Li-6400 Portable Photosynthesis System, following the manufacturer's instruction. The measurements were made from 2 hours after sunrise to 2 hours before sunset. Fully expanded leaves were measured under the steady condition of 400 μmol s^{-1} CO_2, 70% relative humidity, 2000 μmol m^{-2} s^{-1} light density, and 26 °C. Growth measurements were made 30 days after flowering. Five parameters were measured including plant height, leaf number, maximum leaf length, maximum leaf width and bottom stem diameter. Six plants for each group were weighted for their total biomass.

2.2.8 Protein modeling

Protein binding sites within oleosin were predict by PredictProtein. Comparative 3D structure modeling of oleosin and various derivatives were carried out with I-TASSER and RaptorX servers. Only the optimal truncated hydrophobic proteins are presented in Figure 2-2D.

2.2.9 GC-MS Analysis of fatty acid methyl ester

Total lipid of lipid droplets isolated from chloroplasts were analyzed in the form of fatty acid methyl esters according to previous study[60].

2.2.10 Stimulated Raman Scattering Microscopy

A high power Nd:YO4 oscillator (picoTRAIN, HighQ Laser, Austria) producing 7 ps pulse trains at 1064 nm (15W max) and 532 nm (9W max). 2W of the 1064 nm light was used as the Stokes beam. The 532 nm beam was directed to pump an optic parametric oscillator (Levante Emerald, APE GmbH, Germany) to produce 6 ps tunable wavelength pulse train as the pump beam and it is tuned to 813 nm for C-H resonance frequency at 2900 cm-1 for example. The Stokes beam is intensity-modulated by an acoustic optic modulator (3080-122, Crystal Technology) at 10 MHz with 80% modulation depth and then combined with pump beams with a long-pass beam combiner (1064dcrb, Chroma). The two beams were routed to a custom modified mirror-scanning microscope system (BX62WI/FV300, Olympus) attached with the same above Olympus inverted microscope as FLIM. Typical laser power at the sample plane was 80 mW for each beam and this allows for continuous imaging without causing any noticeable photodamage. The light transmitted through the sample was collected by a high numeric aperture condenser (1.45 NA O, Nikon) and filtered by an optical filter (CARS980/220, Chroma) to block the Stokes beam completely so that only amplitude modulation on the pump beams at 10 MHz due to the SRS process can be detected. Pump beam intensity was detected by a large area silicon PIN photodiode (FDS1010, Thorlabs) back-biased at 70V. A lock-in amplifier (SR844, Stanford Research Systems) with full scale sensitivity set at 100 µV was used to detect the intensity change in pump beam.

2.2.11 Shot-gun proteomics analysis

The intact chloroplasts for proteomics analysis were isolated as previously described[61]. The total plant protein and chloroplast protein was extracted, digested and

desalted as previously described[62]. Then the desalted peptides were fractionated with high

pH reversed-phase peptide fractionation kit (Thermo Fisher Scientific, San Jose, CA) and

then loaded to nLC-LTQ orbitrap mass spectrometer (Thermo Fisher Scientific, San Jose,

CA). DATA was analyzed as previously described[62].

2.3 Results and discussion

2.3.1   Derivation of equations for mathematic modeling

According to Fick's First Law and the Meyer-Overton rule, relation among Flux

density (J), diffusion coefficient (D), partition coefficient (K), diffusion distance (δ), and

squalene concentration in plastids (V_2) and in cytosol (V_1) are presented as:

$$J = \frac{DK}{\delta}(V_2 - V_1) \tag{1}$$

As mentioned in the main text, squalene total leaking rate is (b), so (b) can be

presented as:

$$b = \int_0^A JdA = \int_0^A \frac{DK}{\delta}(V_2 - V_1)dA \tag{2}$$

(A) is the total plastid membrane area. As samples are in the same growth

condition and the same growth stage, we regarded total plastid membrane (A) are the

same:

$$b = \frac{DKA}{\delta}(V_2 - V_1) \tag{3}$$

16

(a) is squalene synthesis rate. Based on the equation, squalene content = total synthesis - total leak. Assuming there is no enzyme in plastids to convert squalene. total squalene content (V) is:

$$\frac{dV}{dt} = a - b = a - \frac{DKA}{\delta}(V_2 - V_1) \tag{4}$$

Define constant (k) by:

$$k = \frac{DKA}{\delta} \tag{5}$$

As cytosol squalene epoxidase is every efficient, $V_1 << V_2$, $V_2 = V$. Equation (4) is simplified by:

$$\frac{dV}{dt} = a - kV \tag{6}$$

Finally, the equation can be transformed to:

$$V = \frac{a}{k} - (\frac{a}{k} - V_0)e^{-kt} \tag{7}$$

According to the derivation as shown above, we derived Equation (7) for squalene content (V) under dark treatment, where V_0 represents the squalene content at 0 day, (a) is a constant representing the squalene biosynthesis rate, and (k) is a constant relevant to transmembrane permeability:

17

$$V = \frac{a}{k} - (\frac{a}{k} - V_0)e^{-kt} \tag{7}$$

V thus responds to time (t) as shown in Equation (1), where a, k, and V_0 are all constants.

2.3.2   The diffusion and instability of squalene was demonstrated by modeling and treatment experiment in dark

The leakage of plastidic squalene was first predicted by computational modeling and further verified by treatment experiments in the dark. The FPS-SQS lines were engineered with overexpressed FPS and SQS enzymes targeted to the chloroplasts driven by constitutive promoters as previously described[53]. As shown in Figure 2-1A and B, the squalene content of leaves from young FPS-SQS lines were measured either under a 16 hour light/8 hour dark cycle (Figure 2-1A) or under completely dark conditions (Figure 2-1B). Under light conditions, the young leaves accumulated squalene to about 1 mg/g fresh weight. Under treatment in the dark, the squalene content significantly decreased for the first 3 days, assumingly due to the lack of production and leakage of squalene (Figure 2-1B). Both the increase and the reduction of squalene fit perfectly into a mathematical model according to the Fick's law of diffusion and the Meyer-Overton rule*13* (Equation 7). In plants, squalene is a sterol precursor and is convertd by downstream enzymes such as squalene epoxidase. Previous work also demonstrated that knocking-out downstream hopene cyclase in cyanobacteria could significantly increase squalene accumulation[36]. Considering that squalene epoxidase *in planta* is a highly efficient enzyme, all squalene transported to cytosol would be converted to downstream products for sterol synthesis[9].

18

The squalene accumulation thus results from the balanced impacts of squalene biosynthesis and leaking. Under darkness, the terpene biosynthesis should cease or be significantly slower[63] and the predominant impact for squalene content would be the leaking effects. In fact, the regression analysis showed that experimental data fits well into the equation 7 at an R^2 of 0.88 (Figure 2-1). The results indicated that the plastid synthetized squalene might 'leak' out of chloroplast by transporting itself through the membrane. In order to achieve a higher squalene production, we thus designed a synthetic droplet to prevent the leaking, store the squalene product, and co-compartmentalize biosynthesis and storage in chloroplast.
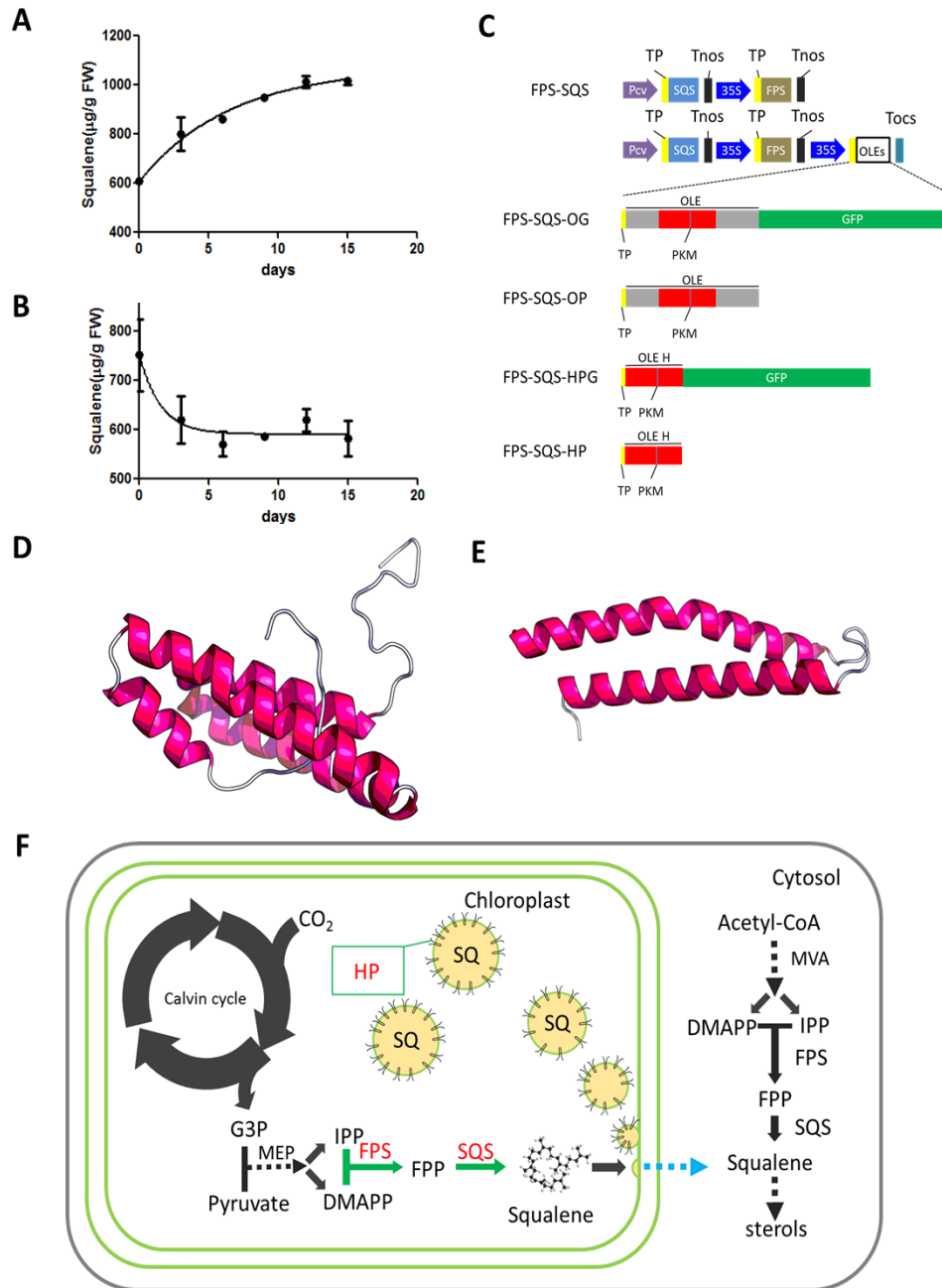
Figure 2-1. Treatment experiments in darkness and the design of synthetic droplet. (A) Squalene accumulation of FPS-SQS lines in leaves for 15 days in the condition of 16 h light/8 h dark with triplicates. (B) Squalene accumulation of FPS-SQS lines in leaves for 15 days of dark treatment with triplicates. (C) Protein design and constructs used to engineer co-compartmentation of squalene synthesis and storage. (D) Structure modeling for the full-length oleosin. (E) Protein structure modeling for the hydrophobic protein (HP) derived from oleosin. (F) The proposed mechanism of co-compartmentation of terpene storage and biosynthesis, where squalene biosynthesis and the synthetic droplets were both engineered in chloroplasts. MVA, mevalonate pathway. MEP, non-mevalonate pathway. SQ, squalene. LD, lipid droplets. OLE, oleosin. OLE H, oleosin hydrophobic domain. PKM, proline-knot motif. TP, chloroplast transit peptide. Pcv, cassava vein mosaic viral promoter. 35S, 35S Cauliflower mosaic viral promoter, Tnos, nos terminator. Tocs, ocs terminator. Error bars represent standard deviation. Reprinted with permission[46].

2.3.3 Biodesign to co-compartmentalize squalene synthesis and storage

Proper biodesign is the key to generate synthetic storage droplets at proper sub-cellular locations and to enable the co-compartmentation of synthesis and storage. Even though oleosin was previously engineered to enhance lipid accumulation, for the compartmentation of squalene biosynthesis and storage, phospholipid monolayer forming protein had to be targeted to chloroplast instead of ER. As shown in Figure 2-1C and 2-2, oleosin contains three domains, N-terminal amphiphilic domain, hydrophobic domain, and C-terminal amphiphilic domain[55]. According to the protein structure modeling, the hydrophobic domain containing a 'proline-knot motif' is believed to be required for stabilizing the phospholipid monolayer during lipid droplet formation, whereas N-terminal domain may contain the ER-targeting domain and lipase-binding domain (Figure 2-1D and Figure 2-2A)[64,65]. The principle of protein design requires the removal of ER-targeting domain and lipase-binding domain, whereas maintaining the 3 dimensional (3D) structure of the hydrophobic region to allow the interaction with the hydrophobic 'tail' of phospholipids to form synthetic droplets (Figure 2-2A). Protein modeling was carried out to evaluate the effects of different truncation sites and an optimal 3D structure was obtained based on the modeling (Figure 2-2B, C, and D). Based on the modeling and the cleavage site, we replaced the N-terminal domain with a transit peptide for chloroplast localization and removed the C-terminal domain to form a stabilized hydrophobic protein (HP) (Figure 2-1D, E and Figure 2-2). In an alternative design, the C-terminal domain was replaced with GFP for visualization and protein expression verification. The hydrophobic protein, without the original N- and C-terminals, would also prevent various modification and protein binding.

21

A

MADTARGTHHDIIGRDQYPMMGRDRDQYQMSGRGSDYSKSRQI
AKAATAVTAGGSLLVLSSLTLVGTVIALTVATPLLVIFSPILVPALITVALLI
TGFLSSGGFGIAAITVFSWIYKYATGEHPQGSDKLDSARMKLGSKAQ
DLKDRAQYYGQQHTGGEHDRDRTRGGQHTT

B

| Amphipathic | Hydrophobic | Amphipathic |

KAATAVTAGGSLLVLSSLTLVGTVIALTVATPLLVIFSPILVPALITVALLIT
GFLSSGGFGIAAITVFSWIY

C

D
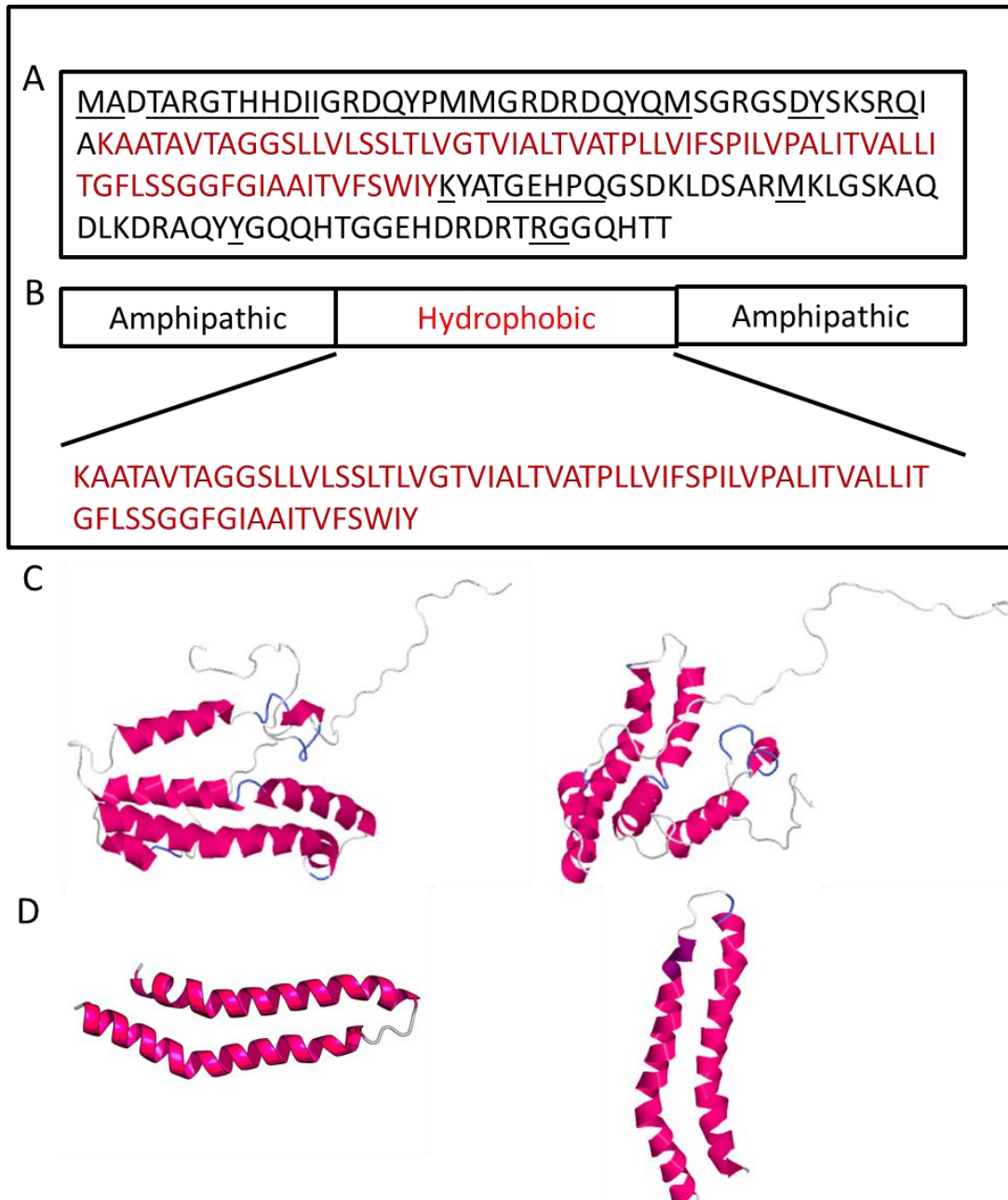
Figure 2-2. The detailed design of hydrophobic protein, where oleosin sequence and hydrophobic protein sequence were shown. (A) Amino acid sequence of full length oleosin where the hydrophobic domain of oleosin is marked in red. The underlined amino acids were predicted to be protein binding sites. (B) Simplified illustration of protein domains in oleosin. Oleosin is composed of two amphiphilic in the figure domains at N-terminus and C-terminus and one hydrophobic domain in the middle. (C) Comparative structure modeling of 3D structure of full length oleosin. (D) Optimal design of hydrophobic protein (HP) was derived from the hydrophobic domain of oleosin based on the 3D modeling of oleosin protein and further fusion with chloroplast signal peptides. Reprinted with permission[46].

We hypothesized that the designed hydrophobic protein would transit into chloroplasts and change the membrane topology to form a synthetic droplet for terpene storage, which could further synergize with chloroplast-targeted terpene biosynthesis for co-compartmentation and yield increase (Figure 2-1F).

In order to evaluate the effectiveness of the co-compartmentation and characterize the synthetic droplet formation in chloroplasts, several variants of designed proteins were included in the study: hydrophobic protein with chloroplast transit peptide, the hydrophobic protein fused to transit peptide at N terminal and GFP at the C terminal (HPG), the original oleosin protein (OP), and the oleosin protein fused with GFP (OG) (Figure 2-1C). These variants were transformed into plants with two strategies for transformation: the sequential transformation of the aforementioned designed proteins into an FPS-SQS plastid expression line, or the building of FPS-SQS-Synthetic Protein constructs to directly transform into tobacco for synthetic droplet formation and co-compartmentation of squalene synthesis and storage (Figure 2-1C, Table 2-1). The performance of transformants from both strategies was consistent and we hereby focused on analysis of the transgenic plants from the second strategy. The transgenic plants were first selected on antibiotics and verified for transgene incorporation with genomic DNA PCR. The positive T0 transformants were analyzed for squalene content (Table 2-1), where transformants from FPS-SQS-HP and FPS-SQS-HPG lines had the highest squalene yield. The copy number of transgenes from selected lines were verified by segregation experiments in T1 and T2 generations (Table 2-2). The single copy heterozygous lines from T1 were first used for imaging and squalene analysis (Figure 2-3 to 2-14).

23

Table 2-1. Screening of T0 transgenic tobacco lines for squalene content.

Constructs	# of Positive lines evaluated	Average (µg/g FW)	Max	Min
FPS-SQS	9	441.0	520.8	282.5
FPS-SQS-HP	18	785.3	1111.2	237.8
FPS-SQS-HPG	15	863.0	1116.0	232.8
FPS-SQS-OP	18	511.0	727.4	265.4
FPS-SQS-OG	12	377.3	818.4	204.2

Note: Positive tobacco transgenic lines generated from each construct were screened by antibiotics and genomic DNA PCR. All positive lines were further analyzed for squalene content at different developmental stages. The table presents the squalene content from lower leaves sampled approximately 80 days after plants were transferred to soil. The average, maximum, and minimum squalene content for all positive independent lines are presented. The detection limit of squalene is 50 µg/g FW. Reprinted with permission[46].

Table 2-2. G418 resistance test in T1 and T2 generation of representative transformant lines.

	T1			T2		
	Lines	G^R	G^S	Lines	G^R	G^S
FPS-SQS-HPG	1	207	76	1-x	21	14
	2	210	58	2-x	116	0
	3	126	35	3-x	33	6
	4	88	30	4-x	63	15
FPS-SQS-HP	1	60	21	1-x	41	14
	2	29	15	2-x	57	13
	3	46	15	3-x	42	12
	4	60	21	4-x	52	17
FPS-SQS	1	93	42	1-x	46	0
	2	97	38	2-x	87	0
	4	138	1	4-x	34	0
	5	53	46	5-x	42	12

Note: G^R and G^S represents the number of G418-resistant and G418-sensitive plants, respectively. Reprinted with permission[46].
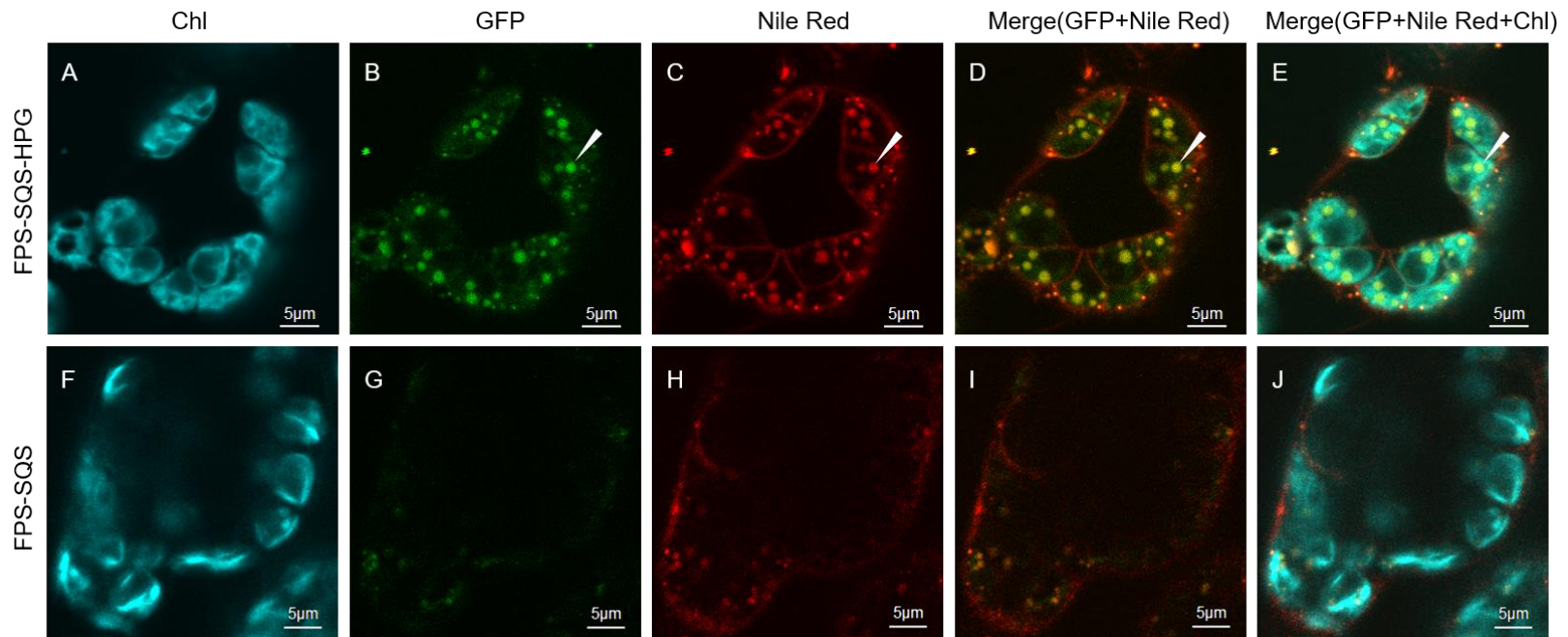
Figure 2-3. Subcellular localization of synthetic storage droplets in tobacco mesophyll cells of FPS-SQS-HPG line (A-E) and FPS-SQS line (F-H). (A, F) Chlorophyll auto fluorescence; (B, G) GFP fluorescence; (C, H) Nile red fluorescence; (D, I) overlap GFP signal with Nile red signal; (E, J) overlap chlorophyll auto-fluorescence signal, GFP signal and Nile red signal. White arrows show an example of synthetic droplets. Scale bar is 5 µm. Reprinted with permission[46].

2.3.4   Formation of lipid droplets in chloroplasts by expressing the hydrophobic protein

Confocal microscopy analysis confirmed the formation of synthetic storage droplets in chloroplasts. As shown in Figure 2-3A,B,C,D,E, the signals from GFP, Nile red staining, and chlorophyll auto-fluorescence were integrated for the FPS-SQS-HPG transformed lines, in which the GFP were fused to the C-terminus of hydrophobic protein. Nile red staining was carried out to visualize the accumulation of neutral lipid or hydrophobic hydrocarbons such as squalene[66]. The Nile red stained droplet structure could be clearly observed in the mesophyll cells in the FPS-SQS-HPG transformed plants, whilst no such structure was found in the FPS-SQS plants (Figure 2-3H). The GFP signal from FPS-SQS-HPG-transformed plants well overlapped with the Nile red fluorescence, indicating that the HPG protein promoted the formation of synthetic droplet containing squalene. In addition, the GFP and Nile red signals further overlapped with the auto-fluorescence signal from chlorophyll, indicating the synthetic storage droplet located within the chloroplast. Overall, the confocal microscopy analysis suggested that HPG and HP protein can be exploited to induce the formation of synthetic droplets in chloroplasts to store squalene.

The co-localization of FPS, SQS and HPG in chloroplasts was further verified by molecular methods. RT-PCR was carried out to verify the gene expression of FPS, SQS and HPG in different FPS-SQS-HPG lines (Figure 2-4). Shot-gun proteomics experiments were carried out to characterize the soluble protein in one of the transformants (FPS-SQS-HPG2), with a wild-type plant and a FPS-SQS line as control samples. As shown in Figure 2-5A, the shot-gun proteomics verified the expression of FPS and SQS in both the FPS-SQS line and the FPS-SQS-HPG line, whereas wild-type control does not have either

enzyme expressed. Considering that memberane-bound HPG protein is highly resistent to trypsin digestion for shot-gun proteomics, we used western blot to verify the expression of HPG protein. As shown in Figure 2-5B, HPG protein is co-expressed with FPS and SQS only in FPS-SQS-HPG lines, not in the FPS-SQS lines.
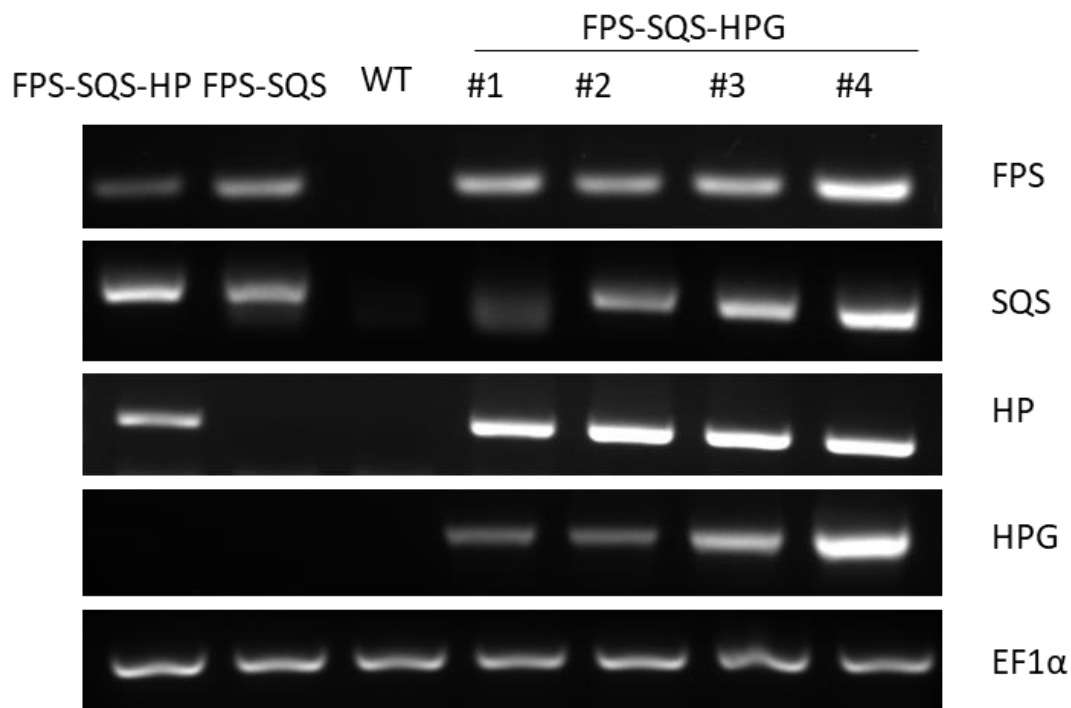
Figure 2-4. The expression of transgenes in T1 generation transgenic plants was measured by reverse-transcription PCR. EF1α was used as the internal control. From left to right is transgenic plants of FPS-SQS-HP, FPS-SQS, WT and four lines of FPS-SQS-HPG. Reprinted with permission[46].
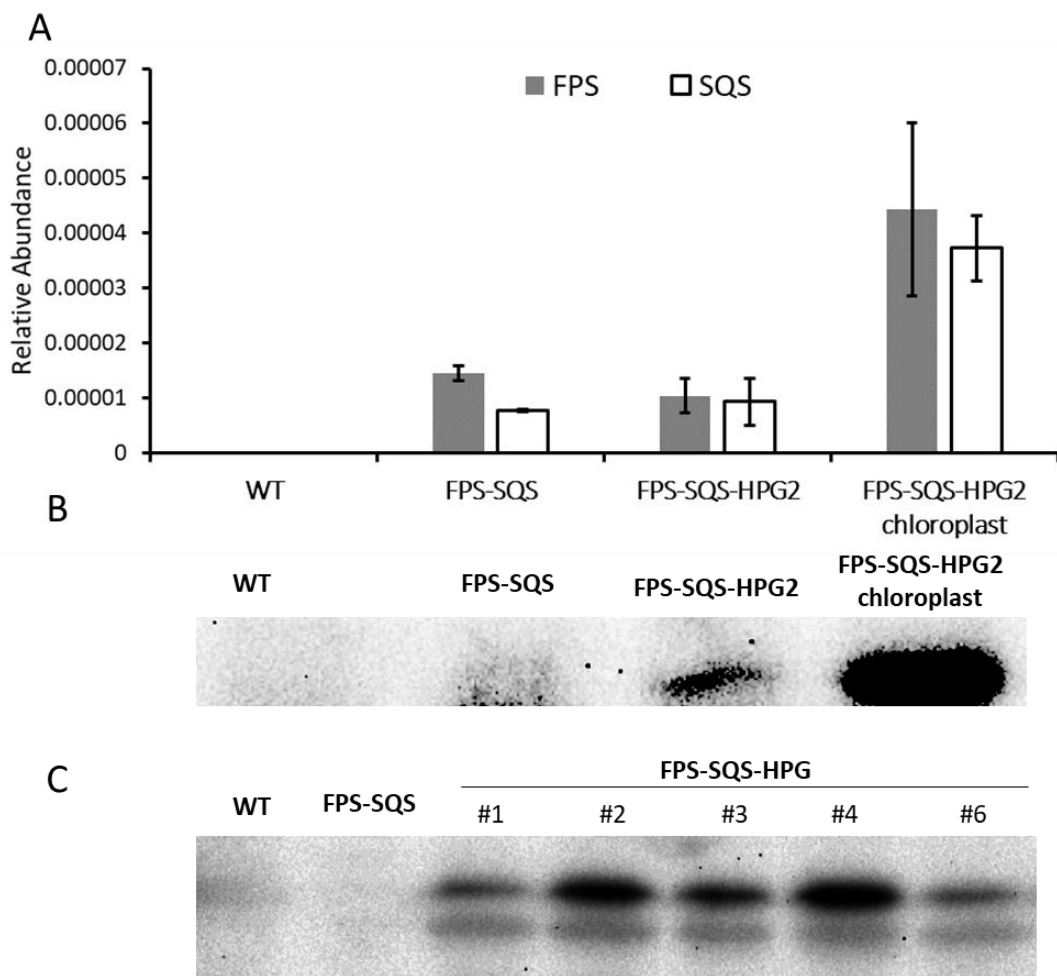
Figure 2-5. Protein expression and localization of FPS, SQS, and HPG in transgenic plants as measured by shot-gun proteomics analysis and Western blots. (A) Relative abundance of FPS and SQS in the soluble proteins from young leaves of WT, FPS-SQS, FPS-SQS-HPG2 and the chloroplast of FPS-SQS-HPG2. The proteins were digested by trypsin and analyzed by nLC-LTQ orbitrap mass spectrometer. The relative abundance was calculated based on spectrum counts and normalized based on the total signal. Error bars represents standard deviation. (B) Western blot of HPG protein. Equal amounts of protein from leaf extracts of WT, FPS-SQS, FPS-SQS-HPG2 and the chloroplast of FPS-SQS-HPG2 were subjected to SDS-PAGE, and the existence of HPG was probed with GFP-antibody. (C) Western blot of HPG protein for total protein extracted from leaves of WT, FPS-SQS and five lines FPS-SQS-HPG2 probed with GFP-antibody. Reprinted with permission[46].

29

The sub-cellular localization of FPS, SQS, and HPG proteins were further verified by proteomics and western blot analyses of the chloroplast protein for FPS-SQS-HPG2. As shown in Figure 2-5A and B, the chloroplast of FPS-SQS-HPG2 contained FPS, SQS, and HPG proteins, verifying the co-localization of these proteins for co-compartmentation of storage droplet and squalene biosynthesis. The expression of HPG protein in multiple transgenic lines was also verified (Figure 2-6C). Overall, the results from molecular analysis corroborated the confocal microscopy analysis to verify the co-compartmentation of squalene biosynthesis and storage.

2.3.5   The formation of synthetic droplets significantly enhanced squalene yield without compromising growth

The formation of the synthetic storage droplet in chloroplast significantly enhanced the squalene production, when the storage and biosynthesis of squalene were co-compartmentalized. (Figure 2-6A, Figure 2-7). As shown in Figure 2-5A, the squalene production in T1 plants of FPS-SQS-HPG and FPS-SQS-HP transformed tobacco was increased by about three folds as compared to that of FPS-SQS transformed tobacco. Paired-wise T tests revealed that the differences between FPS-SQS and FPS-SQS-HPG or FPS-SQS-HP were significant, with a P-value of less than 0.01. The highest squalene yield in HPG lines was about 2.6 mg/g FW (Fresh Weight) of tobacco leaves, representing the highest squalene yield reported. Although the FPS-SQS-OG-transformed lines also yielded higher squalene levels than that of the FPS-SQS-transformed lines (P=0.0286), FPS-SQS-HPG and FPS-SQS-HP lines both yielded squalene at about 1.5 folds higher than that of FPS-SQS-OG lines. The increased squalene yield in FPS-SQS-HPG-transformed lines was also verified in homozygous T2 lines (Figure 2-8). Overall, the results indicated that the

30

design of HP and HPG was effective in enabling the formation of synthetic storage droplet in chloroplasts. The co-compartmentation of biosynthesis and storage significantly enhanced the squalene production and achieved a significant increase of squalene yield.

Accumulating terpene at high levels often leads to growth defects[53]. To evaluate if the synthetic droplet strategy leads to similar growth defects, we compared the growth of FPS-SQS-OG, FPS-SQS-HP, FPS-SQS-HPG, and FPS-SQS-transformed lines. As shown in Figure 2-6B, no significant differences were found among these lines in all the growth parameters measured, including the height, leaf count, largest leaf length, largest leaf width, and stem diameter. In a similar way, total biomass and photosynthesis rate for the FPS-SQS-HPG transformed lines were similar to those of FPS-SQS-transformed lines (Figure 2-9). Even though squalene accumulation reached 2.6 mg/g FW in the FPS-SQS-HP and FPS-SQS-HPG-transformed lines, the substantially increased squalene accumulation did not result in a significant impact on plant growth. This is very different from the previous studies, showing that increased squalene production in trichomes led to dwarf plants and bleached leaves[53]. Overall, the results highlighted that synthetic storage droplet represented an effective strategy to enhance bioproduct accumulation without compromising plant growth.
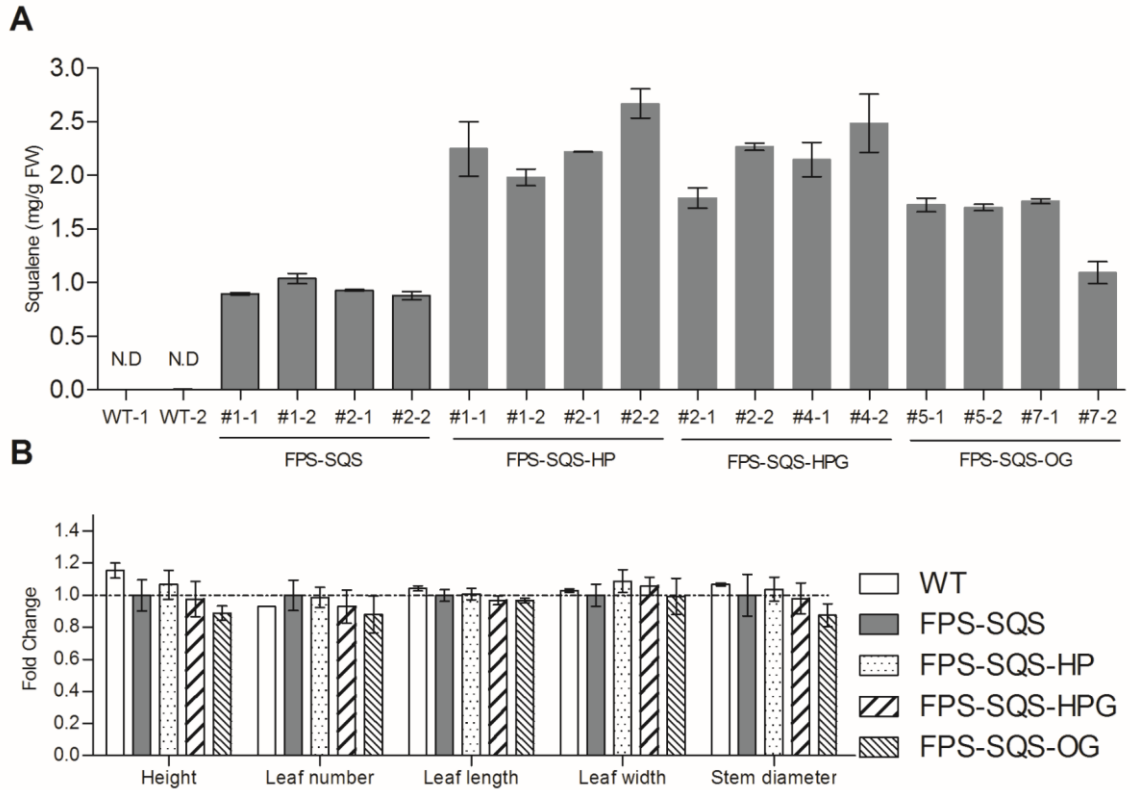
Figure 2-6. Squalene yield and plant growth parameters in transgenic lines with and without synthetic droplets. (A) Two independent transgenic lines for each construct and two individual plants for each line were analyzed for squalene content with 2 technical replicates. Lower leaves of 120 days plants were sampled for squalene analysis. (B) Plant growth parameters were measured including height, leaf count, largest leaf length, largest leaf width and stem diameter. The means and standard deviation of growth parameters fold change are compared to FPS-SQS line were presented. Plants were grown in 16 h light/8 h darkness cycle. Each squalene measurement had 2 replicates. Squalene detection limit was 50 μg/g FW. N.D not detectable. Error bar represents standard deviation. Reprinted with permission[46].
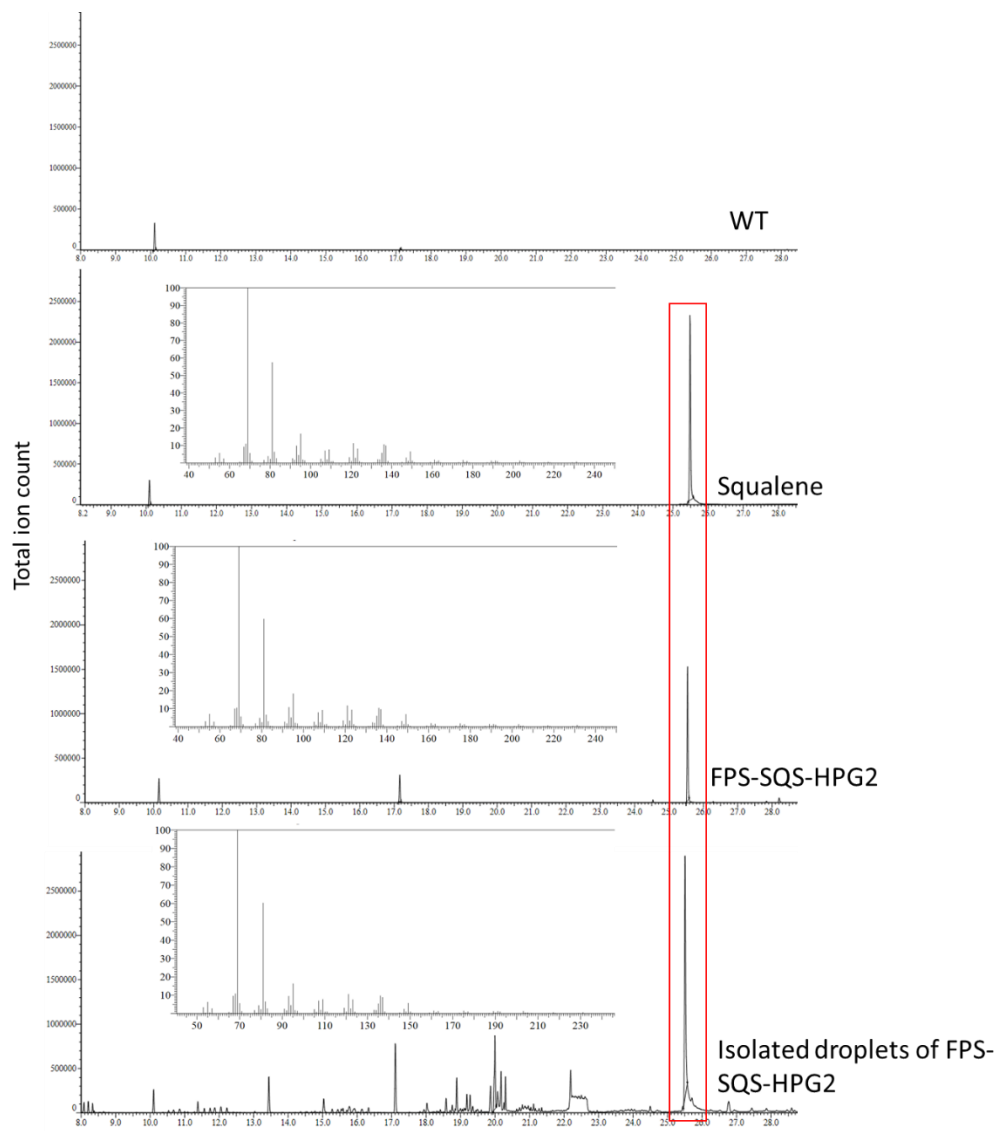
Figure 2-7. GC-MS comparison of hexane extracts of WT, squalene standard, FPS-SQS-HPG and the isolated droplets of FPS-SQS-HPG. The MS for 25.5 min peak is shown. (−)-α-Cedrene was used as internal standard. Reprinted with permission[46].
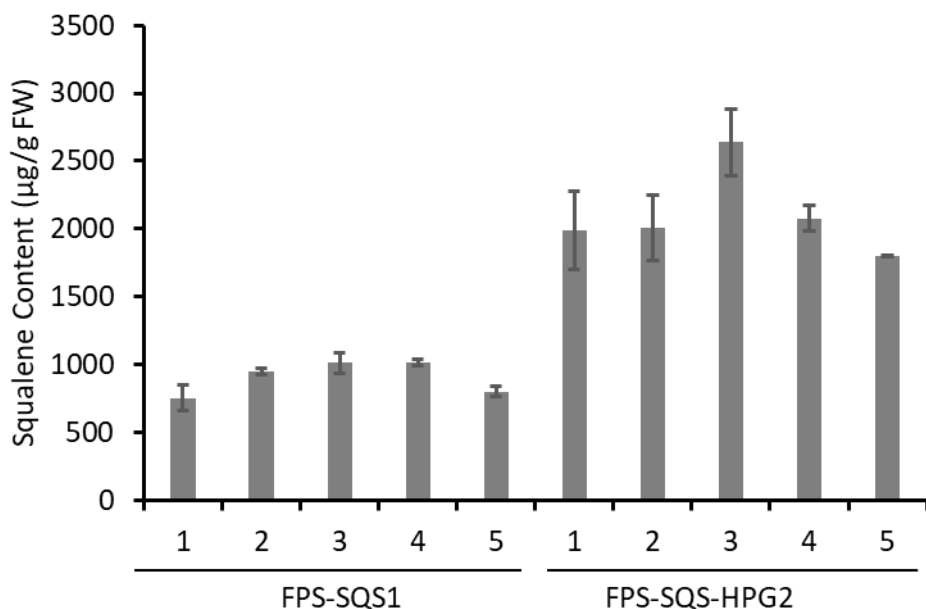
Figure 2-8. Squalene content of T2 homozygous lines. Squalene accumulation of mature T2 generation homozygous lines FPS-SQS-HPG2 and FPS-SQS1 were analyzed by GC-MS with 5 biological replicates and 2 technical duplicates for each line. Error bars represent standard deviation. Reprinted with permission[46].

2.3.6   The synthetic droplets contain squalene

Several strategies were used to further confirm that the synthetic storage droplet contains squalene. First, Raman microspectroscopy was used to analyze the subcellular chemical composition. As a label-free microscopic technique, the technology has been widely used because it offers high chemical specificity[67-70]. SRS (Stimulated Raman Scattering) in particular increased the signal to noise ratio and reduced the non-resonant background to allow fast image acquisition, which distinguished the fine molecular differences at ~300 nm resolution.
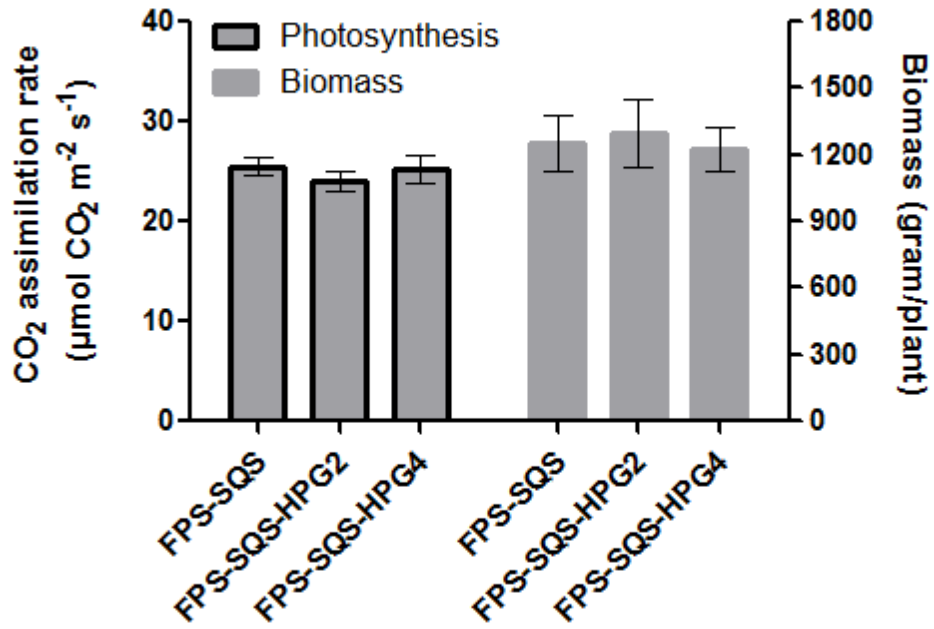
Figure 2-9. Plant above-ground biomass and photosynthesis. Plant above-ground biomass was measured before harvesting. The photosynthesis of transgenic plants (6 biological replicates) was measured after 2 months of growth, using a Li-Cor 6400 instrument. Error bars represent standard error. Reprinted with permission[46].
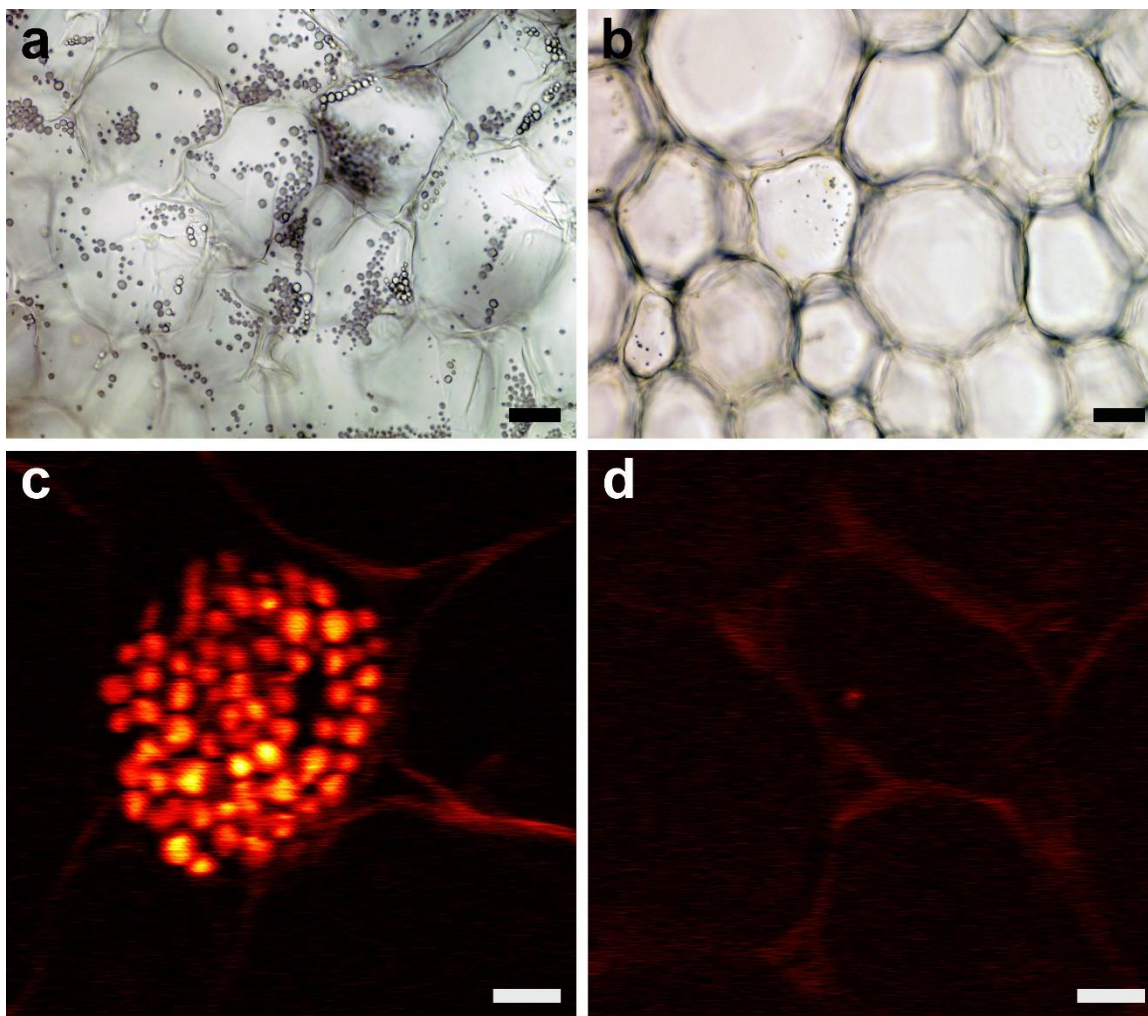
Figure 2-10. In vitro droplet composition analysis by Stimulated Raman Scattering microscopy. (a, b) Leaf vein section under light microscope for FPS-SQS-HPG and FPS-SQS lines respectively. (c, d) droplet imaging in leaf vein section for FPS-SQS-HPG and FPS-SQS lines respectively under Stimulated Raman Scattering microscopy at frequency 2900 cm-1. Scale bars are 10 µm. Reprinted with permission[46].

As shown in Figure 2-10A and B, the synthetic storage droplets could be clearly visualized in the bright field in the FPS-SQS-HPG transformed lines, whilst no significant droplet structure was found in the FPS-SQS lines. The SRS microscopy identified strong 2900 cm^{-1} spectrum in the synthetic droplet, corresponding to the C-H bond (Figure 2-10C), whilst no visible droplet or strong 2900cm^{-1} spectrum signals could be found in the FPS-SQS lines as control plants (Figure 2-10D). Replicate images of bright field and SRS microscopy are presented in Figure 2-11. The results indicate that the synthetic storage droplet contained high content of hydrocarbons with C-H bonds. Second, we isolated the droplets from chloroplasts of FPS-SQS-HPG-transformed lines and carried out squalene analysis. As shown in Figure 2-10A, the FPS-SQS-HPG lines contained a much higher concentration of droplets in chloroplasts. The further squalene analysis confirms that squalene levels are significantly higher in the isolated droplets from chloroplasts of FPS-SQS-HPG lines (Figure 2-12A). The retention time of GC (Gas Chromatograph) and fragmentation pattern of MS (mass spectrometry) both confirmed the compound identified in droplets as squalene (Figure 2-7). More importantly, we have analyzed the ratio of squalene vs. total lipid in synthetic droplets[60]. The significant increase of the ratio further proved the enrichment of squalene in the droplet. Overall, the results indicated that the synthetic storage droplets contained squalene and led to increased squalene storage.
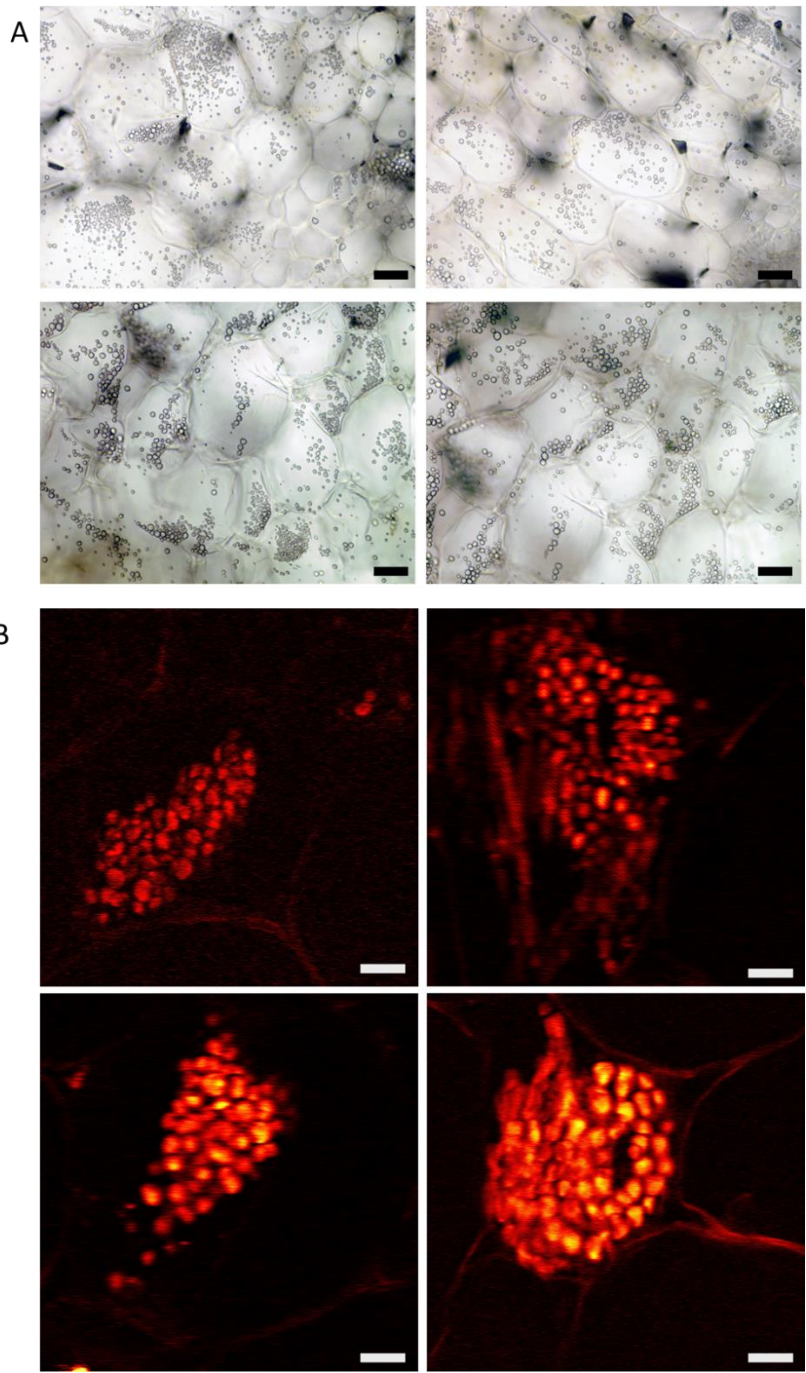
Figure 2-11. Multiple replicates of bright-field confocal microscopy and SRS Raman microscopy analyses of synthetic droplets in FPS-SQS-HPG2. Scale bars are 10 µm. Reprinted with permission[46].
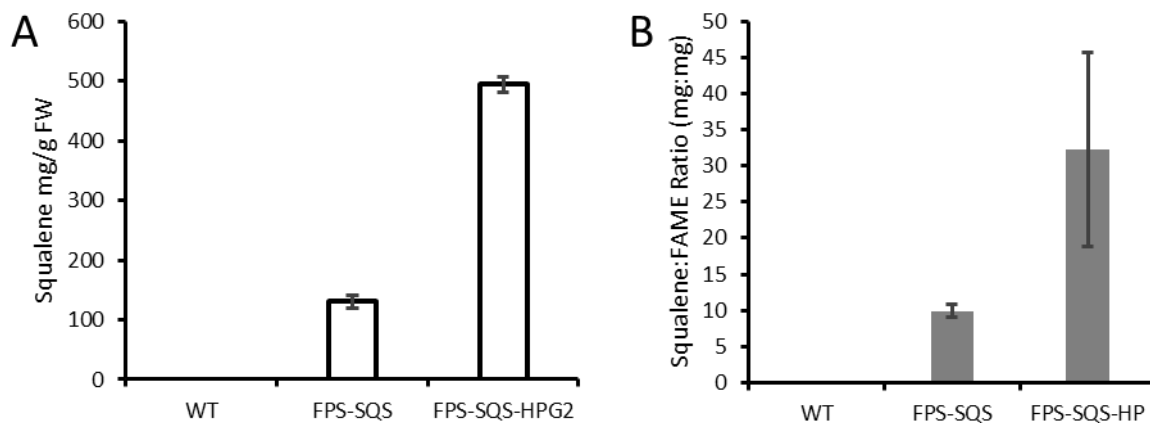
Figure 2-12. Squalene content and its relative abundance to total lipid content in isolated lipid droplets was measured. (A) Squalene content in synthetic droplet. Chloroplasts were first isolated, followed by lipid droplets isolation and squalene extraction and measurement (3 replicates). Fresh weight was counted as original leaf weight. Error bars represent one standard deviation. (B) The ratio of squalene vs. total lipid in synthetic droplet. The squalene and lipid were both extracted from the aforementioned droplets. Total lipid was measured after conversion to fatty acid methyl ester (FAME) using esterification by methanol. Reprinted with permission[46].
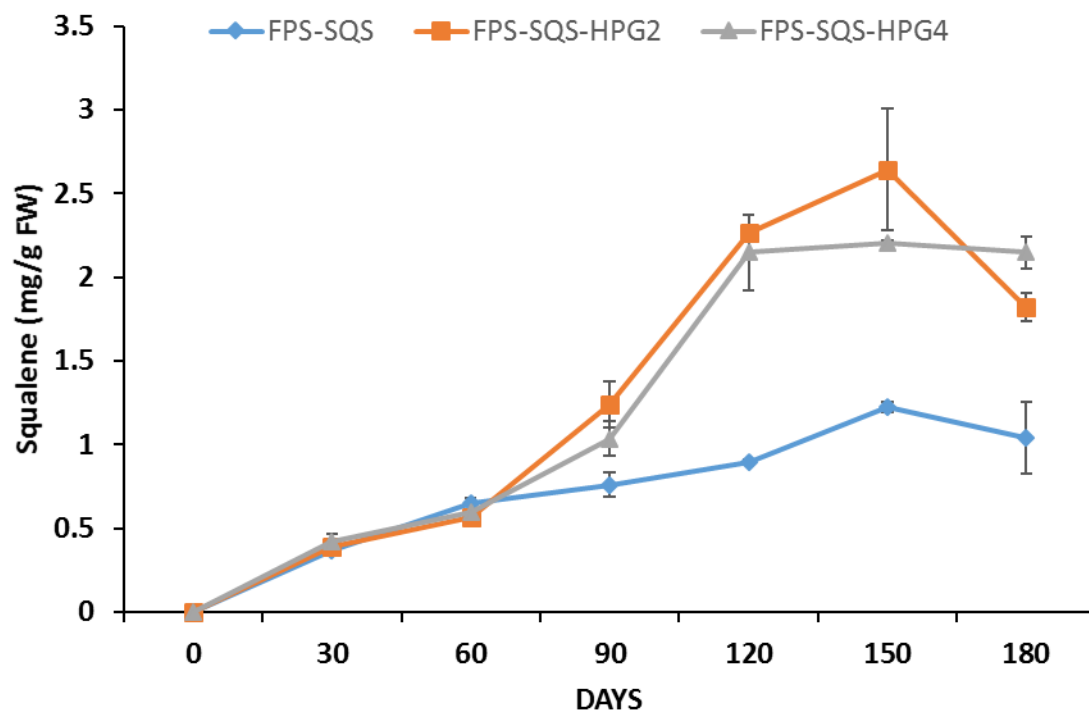
Figure 2-13. Change of squalene content throughout 180 days in four different T1 transgenic lines. Every 30 days, leaf samples were collected for squalene analysis. 0 day represents squalene content in 14-day germinated seedlings. Plant ages were counted from the day of moving from MS medium to soil. Transgenic plants were grown in standard greenhouse condition, where the temperature is controlled below 35°C, humidity below 60% and light for 14 h. N.D suggested not detectable. Detection limit was 50 µg/g FW. Error bars represent standard deviation. Reprinted with permission[46].

2.3.7 Synthetic droplets stabilized squalene in the plants

The mechanisms for enhanced squalene yield by the synthetic droplets were investigated with several approaches. Squalene accumulation throughout different developmental stages was first studied to evaluate the impact of synthetic storage droplets. During the vegetative growing stage, squalene accumulated to a similar level in FPS-SQS lines and FPS-SQS-HPG lines. However, when plants grew to the fully mature stage, the squalene levels in the FPS-SQS-HPG lines were significantly higher than those of FPS-SQS lines (Figure 2-13). The results suggested that the co-compartmentation of squalene biosynthesis with storage might have avoided the degradation of squalene at a later developmental stage, when squalene accumulated at a higher level. A higher squalene concentration in the chloroplasts might lead to more 'leakage' of squalene out of the chloroplasts, considering that the cross-membrane transport of the molecule is relevant to the concentration differences across the membrane. The synthetic droplet might have prevented the squalene from being transported out of the chloroplasts at higher concentration. The results were further validated by the treatment experiments in darkness.
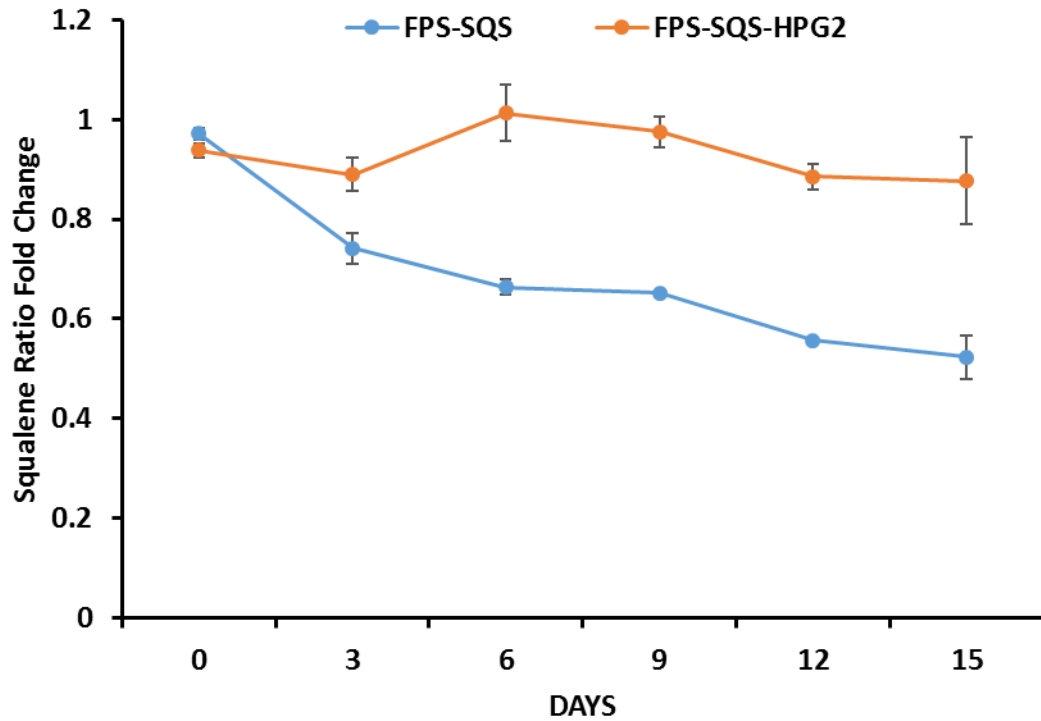
Figure 2-14. Squalene ratio under dark/light growth conditions. Squalene content was monitored in leaves of FPS-SQS line (control) and FPS-SQS-HPG in dark and light conditions with triplicates. The orange line represents the ratio of squalene content for FPS-SQS line under continuous darkness and regular light/darkness cycles. The blue line presents the ratio of squalene content for FPS-SQS-HPG2 line under darkness and regular light/darkness cycles. Error bars represent standard deviation. Reprinted with permission[46].

As shown in Figure 2-14, the squalene level in FPS-SQS lines dropped quickly after the dark treatment and gradually reached a stable level. In contrast, the squalene levels in the FPS-SQS-HPG lines did not show a significant difference between dark and light conditions. As mentioned earlier, squalene is a key intermediate for sterol biosynthesis, and thus can be modified by the endogenous squalene epoxidase in the cytosol[9]. In darkness, squalene biosynthesis either significantly slowed down, or stopped altogether. At the same time, the membrane's permeability allowed squalene to be transported outside of the chloroplast and convertd by downstream pathways in the cytosol, which led to the decrease of squalene level in FPS-SQS lines. The synthetic droplet effectively prevented squalene from 'leaking' out through the chloroplast membrane, and thus kept the squalene concentration unchanged in the FPS-SQS-HPG lines. The results suggested that the co-compartmentation of biosynthesis and storage might have improved storage of squalene compounds in the synthetic storage droplets, and thus prevented the conversion downstream. The results also verified our hypothesis that the leakage of squalene outside of chloroplasts and downstream conversion should be the main reason for decreased squalene accumulation in darkness treatment. The mechanisms of preventing downstream conversion also correlated with the recent findings that methylation of squalene compounds could increase the squalene accumulation significantly[71]. Overall, the synthetic storage droplet might have increased squalene levels by preventing the 'leakage' of squalene out of the chloroplasts and the downstream utilization.

2.4 Discussion

Overall, the study not only resulted in the highest reported yield of squalene in tobacco, but also provided a new solution to engineer bioproduct accumulation. Such a new approach involves the biodesign of a hydrophobic protein that enables creation of subcellular synthetic storage droplets, which could have broad applications. On one side, this design principle further allows co-compartmentation of biosynthesis and storage of various bioproducts at different subcellular locations including chloroplast, mitochondria, and such. The strategy could be applied to a wide range of bioproducts including alkenes, fatty acid esters, and terpenes. The strategy could also complement chloroplast-based protein production driven by oleosin-like peptides to co-manufacture high value compounds and proteins[72]. In particular, this study focused on squalene as a unique high-value bioproduct with broad market applications. The increased demand for squalene has been partially driven by its application as a vaccine adjuvant[73]. The co-compartmentation has enabled a record yield of squalene production over 2 mg/g FW, a level similar to that of lipid and starch accumulation in tobacco leaves[74, 75]. The yield highlighted the potential of the strategy to enable the bioproduct yield. Further pathway enhancement may lead to additional increase of squalene production to enable the sustainable production squalene as a vaccine adjuvant, cosmetic ingredient, nutraceuticals and other applications. The squalene concentration in cell biomass in this study remains among the highest in both microbial and plant study[76]. Considering the mature agriculture practices for tobacco, the tobacco-based biomanufacturing system thus has a significant commercial potential to both transform the declining tobacco industry and produce high value products like squalene. Moreover, the implementation of the synthetic droplet to compartmentalize the storage and

synthesis of various bioproducts at a designed subcellular location could enable broader applications of plant-based biomanufacturing.

Despite the progress, the mechanisms for the synthetic storage droplet to enhance squalene production needs to be further investigated. Squalene accumulation under different developmental stages and light/dark conditions indicated that the co-compartmentation might have prevented the 'leakage' of squalene into the cytosol for downstream modification. However, other mechanisms such as the removal of a concentration-dependent pathway feedback cannot be ruled out. Further study of intermediate accumulation and carbon flux might help to reveal the mechanisms more clearly. Nevertheless, the co-compartmentation of synthetic storage droplet with engineered metabolic pathways could be broadly applied to enhance the production of various target compounds in biomanufacturing.

3. INTER-COMPARTMENTAL 'PULL AND BLOCK' STRATEGY TO ACHIEVE HIGH SQUALENE YIELD

3.1 Introduction

Squalene is a triterpene produced by most organisms and is a precursor for the biosynthesis of steroids. Squalene has broad applications as vaccine adjuvants, skin moisturizers and emulsion enhancers due to its nontoxic nature, skin permeability, and high compatibility with human cells[17,77]. Squalene was first isolated and characterized from sharks (*Squaluss spp.*), and current production methods require sacrificing sharks for their livers[78], which imposes an unsustainable environmental tragedy to meet the increasing demand of squalene[79]. In order to address this challenge, extensive efforts have been taken to develop sustainable squalene production *in planta*, and to use squalene as a model compound for engineering terpene production. There are two biosynthesis pathways for terpene precursors in higher plants; a 2-C-methyl-d-erythritol 4-phosphate (MEP) pathway in chloroplast and a mevalonate (MVA) pathway in the cytosol. Both pathways produce the common precursors isopentenyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP) for the production of numerous terpenes. A heterologous overexpression strategy was developed, in which squalene synthase (SQS) and farnesyl pyrophosphate synthase (FPS) were compartmentalized into chloroplasts to convert DMAPP/IPP to squalene[53]. Although the compartmental strategy led to the increase of squalene production *in planta*, further engineering the high titer production of squalene is still hampered due to the downstream conversion[7,80]. In plants, squalene is oxidized to become 2,3-oxidosqualene by squalene epoxidase (SQE), which is a highly efficient enzyme localized in cytosol[81].

Recent advances in metabolic engineering and synthetic biology offered new opportunities for engineering squalene production by avoiding downstream conversion. Even though chloroplast compartmentation of squalene biosynthesis has led to a significant increase in squalene accumulation, various evidence suggested that squalene might leak out the plastid membrane to be convertd by SQE, which prevented further increase of squalene level. In order to address this issue, triterpene methyltransferases were introduced to chloroplasts, converting squalene to methylated squalene, which led to higher production of total triterpene[71]. In our previous study, synthetic storage droplets were engineered in chloroplasts together with the squalene biosynthesis pathway[46]. The strategy has been proven to be effective to increase squalene yield and stability. Both studies corroborated the fact that squalene can be transported out of chloroplasts and convertd by SQE. Despite the progress, the downstream degradation remained the major challenge for squalene production. The instability of squalene in cytosol has not been solved and still hampers squalene production. In order to address this barrier, we therefore explored the possibility of 'blocking' downstream conversion by down-regulating SQE activity.

Previously, several attempts have been made to inactivate SQEs. For example, *A. thaliana* mutants of *sqe1* or *sqe3* have trace levels of squalene accumulation[9,82]. Nevertheless, plants showed growth defects, where *sqe1* mutants were either lethal or drought sensitive. Moreover, the seeds of both mutants had embryo development defects, leading to seed abortion. The study in cyanobacteria also suggested that blocking downstream conversion could lead to increased terpene accumulation. A substantial gain of squalene yield was achieved by inactivation of squalene conversion enzymes *slr2089*, the hopene cyclase that converts squalene to hopene in *Synechococcus elongatus* PCC 7942.

47

The accumulation of squalene achieved 0.67 mg OD750^{-1} L^{-1} squalene [36], indicating that the downstream conversion is a key bottleneck for further increase of squalene yield in photosynthetic organisms. Furthermore, the silencing of SQE leads to the accumulation of squalene in *Chlamydomonas reinhardti*, but the overexpression of SQS in cytosol, together with SQE knockdown did not have a significant effect on squalene yield, as compared with SQE knockdown stains. The results highlighted the difficulty of further improvement of squalene yield simply by suppressing downstream conversion enzyme and overdressing squalene synthases in the same compartment[37]. In this study, we designed an inter-compartmental "pull and block" strategy, that squalene biosynthesis pathway was overexpressed in chloroplasts to "pull" squalene and squalene epoxidases in the cytosol was inactivated at transcriptional level to "block" squalene conversion, to enhance squalene production.

Overall, accumulation of squalene *in planta* is limited due to the downstream conversion. In this research, we identified the gene actively expressed in the leaf of *N. tabacum*, *SQE3*. *SQE3* was silenced by artificial microRNA to "block" downstream cytosolic squalene conversion. In addition, a squalene synthesis pathway was introduced into chloroplast to "pull" terpene common precursors IPP/DMAPP to squalene. The inter-compartmental "pull and block" strategy led to a significant increase of squalene level without compromising growth.

3.2 Methods and Materials

3.2.1   Plasmid Construction

The DNA fragment of the artificial microRNA 159a backbone was cloned from *Arabidopsis thaliana* genomic DNA by PCR using primers miR159a-F1 (5'-

48

ATCGAAGCTTACAGTTTGCTTATGTCAGATCC-3') and miR159a-R1 (5'-

CTGAGAATTCTAGAGCTCCCTTCAATCCAAAG-3'). The PCR products were

subsequently inserted into pUC19 and confirmed by DNA sequencing. The DNA fragment

of amiRNA-SQE was amplified from above plasmid using miR159-SQE-gib-F primer (5'-

ATCTTGATCTGACGATGGAAGGGTCGATGTACCTGGTCAAAACATGAGTTGAG

CAGGGTA-3') and miR159-SQE-gib-R primer (5'-

GAGAAGGTGAAAGAAGATGGGTCGATGTACCTGGTCAAAAGAAGAGTAAAAG

CCATTA-3'), containing the mature miR159-SQE reverse complementary sequence

(underlined). The DNA fragments of miR159-SQE, together with Cassava vein mosaic

virus promoter, and nos terminator were subsequently incorporated into the *pmeI* site of

pCAMBIA2300 and plasmid FPS-SQS, which was described in our previous research[46].

3.2.2   Plant Transformation and Growth Condition

The tobacco cultivar used in the research was *N. tabacum* Ti1068 which was

described previously [83]. Plant transformation was performed by a standard disc dissection

method[84]. Positive transformants were screened by 50 mg/L antibiotic G418 in the

shooting medium and 30 mg/L G418 in rooting medium. Plants were grown in a standard

greenhouse condition with temperature ranging from 24-35 degrees Celsius and 14 hrs

light/ 10 hrs dark with 50-80% humidity. T1 generation plants were screened by genomic

DNA PCR and RT-PCR.

3.2.3   Squalene Quantification

Squalene extraction was performed as described previously with minor

modifications[46]. Briefly, 0.5 g of tobacco leaf fresh samples or 0.1 g of lyophilized samples

were homogenized with 3 ml hexane containing 270 µg α-cedrene as internal standard for

2 hours. 1 ml of the extract was applied to a silica gel column (40-63 µm) and the eluent

was dried under nitrogen stream to 6 ml. Squalene concentration was analyzed on a

Shimadzu single quadrupole gas chromatograph-mass spectrometer (GCMS-QP2010 SE)

equipped with a Phenomenex Zebron ZB-35HT column (30 m ×0.25 mm, 0.25 µm film

thickness). GC-MS analysis was carried out under the following program: 30:1 split ratio,

injection temperature 220 ℃, initial oven temperature of 40 ℃ for 0.5 min, then raised to

140 ℃ at a rate of 30 ℃ min^{-1}, 150 ℃ at a rate of 5 ℃ min^{-1}, 280 ℃ at a rate of 25 ℃ min^{-1}, 300 ℃ at a rate of 5 ℃ min^{-1} and followed by 2 min hold at 300 ℃. Endogenous

squalene concentrations were quantified based on a standard curve established with known

squalene concentrations.

3.2.4   Terbinafine Treatment Analysis

Tobacco plants were treated with squalene epoxidase specific inhibitor terbinafine.

Two week old seedlings of wildtype and FPS-SQS lines were transferred to MS medium in

glass jars with gradient concentrations of terbinafine hydrochloride (Sigma) at 0, 0.5, 1,

and 3 µM, prepared from a 100 mM terbinafine stock solution in DMSO. DMSO

concentrations were the same in all media. The seedlings were grown at 22 ℃ with 14

hours light per day. Squalene content in leaves was quantified by GC-MS after 4 weeks of

growth.

3.2.5   Sterol Analysis

20 mg lyophilized leaf samples were extracted with 3 ml CHCl$_3$-MeOH (2:1),

containing 15 µg of 5α-cholestane as an internal standard (Sigma). Samples were sonicated

at 50 ℃ for 45 min. 1 ml of the extract was collected, dried in a rotary evaporator, and

applied to a silica gel column with hexane-EtOAc (2:1) and CHCl$_3$-MeOH (2:1). The

eluents were combined and dried using a rotor evaporator. The dried samples were

saponified with 1 ml of MEOH and 5% KOH at 80 ℃ for 1 h and subsequently hydrolyzed

with 4 N HCl in 1 ml MeOH. The samples were extracted with 1 ml hexane 3 times and

were dried under a nitrogen stream. The samples were dissolved in 60 µl of pyridine

followed by adding 40 µl of N-Methyl-N-(trimethylsilyl) trifluoroacetamide with 1%

trimethylchlorosilane, incubated at 50 ℃ for 1 h. Trimethylsilylated sterols were analyzed

by the GC-MS mentioned above with a GC-MS program, splitless mode, injection

temperature 220 ℃, initial oven temperature of 80 ℃ for 0.5 min, then raised to 220 ℃ at

a rate of 20 ℃ min-1, 320 ℃ at a rate of 5 ℃ min-1, and followed by 2 min hold at 320 ℃.

Endogenous sterol concentrations were determined by the peak areas for each sterol and

comparison to the internal standard.

3.2.6   Phylogenetic Analysis

NtSQEs sequences were identified through following databases: National Center

for Biotechnology Information and the SOL Genomics Network. The phylogenetic tree

was generated by Clustal Omega using amino acid sequence and was visualized by Jalview

3.3 Results

3.3.1   Chemical disruption suggested down regulation of SQE may lead to squalene

increase

Squalene conversion by squalene epoxidases was first demonstrated by

experimental treatment using an inhibitor. Plastidic squalene instability was predicted and

demonstrated in our previous research, which indicated that oily squalene is not stable in

chloroplasts and may diffuse out due to the squalene concentration difference between

chloroplast and cytosol and its permeability through membranes[46]. Thus, inactivation of

squalene downstream enzymes became critical for enhancing squalene accumulation

(Figure 3-1A). To investigate how inhibition of key enzymes of degradation pathways will

impact the squalene accumulation, tobacco seedlings were treated with terbinafine, an

allylamine that acts as a specific and non-competitive inhibitor of fungal and plant SQEs[85].

Wildtype plants and FPS-SQS transgenic plants, which overexpressed FPS and SQS in

chloroplast as previously decribed[46,53], were both transferred to MS medium with gradient

concentrations of terbinafine. As shown in Figure 3-1B, FPS-SQS plants, treated with

terbinafine at 0.5 µM and 1 µM, displayed higher squalene accumulation, compared with

those without inhibitor, while squalene is not detectable in wildtype plants with/without

terbinafine. The positive correlation of squalene contents and terbinafine concentrations in

FPS-SQS transgenic plants indicated the leaking effect of squalene. Moreover, the result

suggested that squalene epoxidases are key enzymes that prevent squalene accumulation at

high concentration. Thus, the key to increase squalene yield in tobacco is to inhibit the

enzyme activity using genetic tools. Notably, squalene level in FPS-SQS plants treated

with 3 µM terbinafine dropped significantly due to severe growth defects (data not shown).

This result indicated that complete inactivation of squalene epoxidases might lead to

growth deficiency, which is consistent with previous research results[9,82]. Overall,

enhancing squalene yield would require partially inactivation of squalene epoxidases.
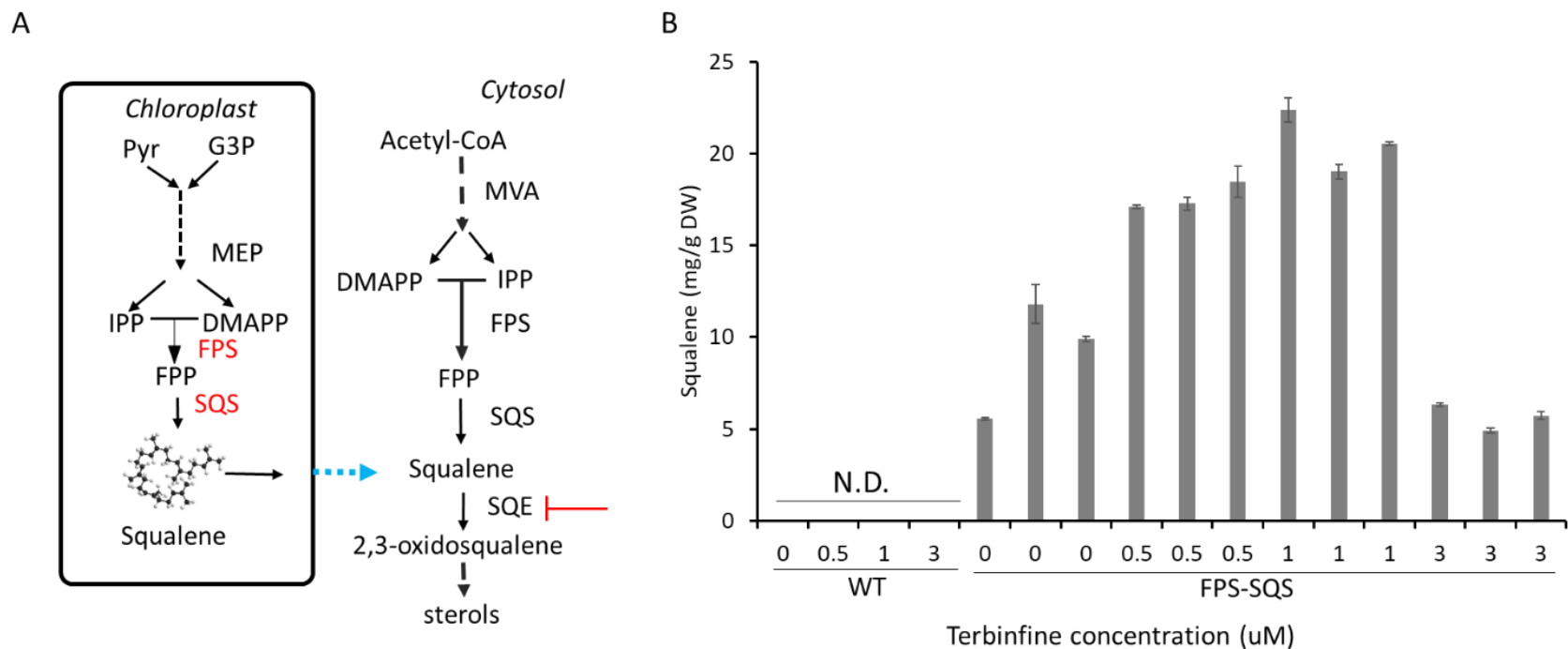
Figure 3-1. Scheme of the design for the squalene epoxidase inhibitor experiment. (A) The proposed mechanism of inter-compartmental 'pull and block' strategy, where squalene biosynthesis pathway was expressed in the chloroplast to 'pull' carbon into squalene, while cytosolic squalene epoxidase was inhibited to 'block' squalene conversion. (B) The squalene content was quantified in wildtype plants and FPS-SQS plants with the presence of gradient concentration of terbinafine. MVA, mevalonate pathway. MEP, non-mevalonate pathway. FPS, farnesyl pyrophosphate synthase. FPP, farnesyl pyrophosphate. SQS, squalene synthase. Pyr, pyruvate. G3P, Glyceraldehyde 3-phosphate. IPP, isopentenyl pyrophosphate. DMAPP, dimethylallyl pyrophosphate. DW, dry weight. WT, wildtype. Each squalene measurement had 3 technical replicates. Squalene detection limit was 500 µg/g DW. N.D. not detectable. Error bars represent standard error.

3.3.2 Identifying the target SQE gene to knock down

Bioinformatic approaches were used to identify candidate SQE genes from the tobacco genome for squalene conversion. Putative SQEs genes in the *N. tabacum* genome were predicted by BLAST using *A. thaliana* protein squalene epoxidases versus the *N. tabacum* TN90 Sierro 2014 database. These genes were designated as *NtSQEs*. There were total of 5 pairs of *NtSQE*s in the *N. tabacum* allotetraploid genome identified, named *NtSQE1, NtSQE2, NtSQE3 NtSQE4 and NtSQE5* (Figure 3-2A). NtSQEs peptide sequences were compared with squalene epoxidases from yeast, human, *A. thaliana* and corn. The resulting sequence alignment and phylogenetic analysis suggested that, similar to *A. thaliana*, NtSQEs were grouped into squalene epoxidases and squalene epoxidase-like proteins (Figure 3-2A). The squalene epoxidase-like proteins in *A. thaliana* were demonstrated to have no squalene epoxidase activity[9].

In order to target proper SQEs, transcriptional expression levels of different SQEs were quantified by reverse-transcriptional PCR using total RNA from the leaf samples of FPS-SQS plastic transformant lines. The result showed that SQE3 is the most actively expressing epoxidase in tobacco leaves (Figure 3-2B). Therefore, SQE3 was designated as the target for RNA silencing.
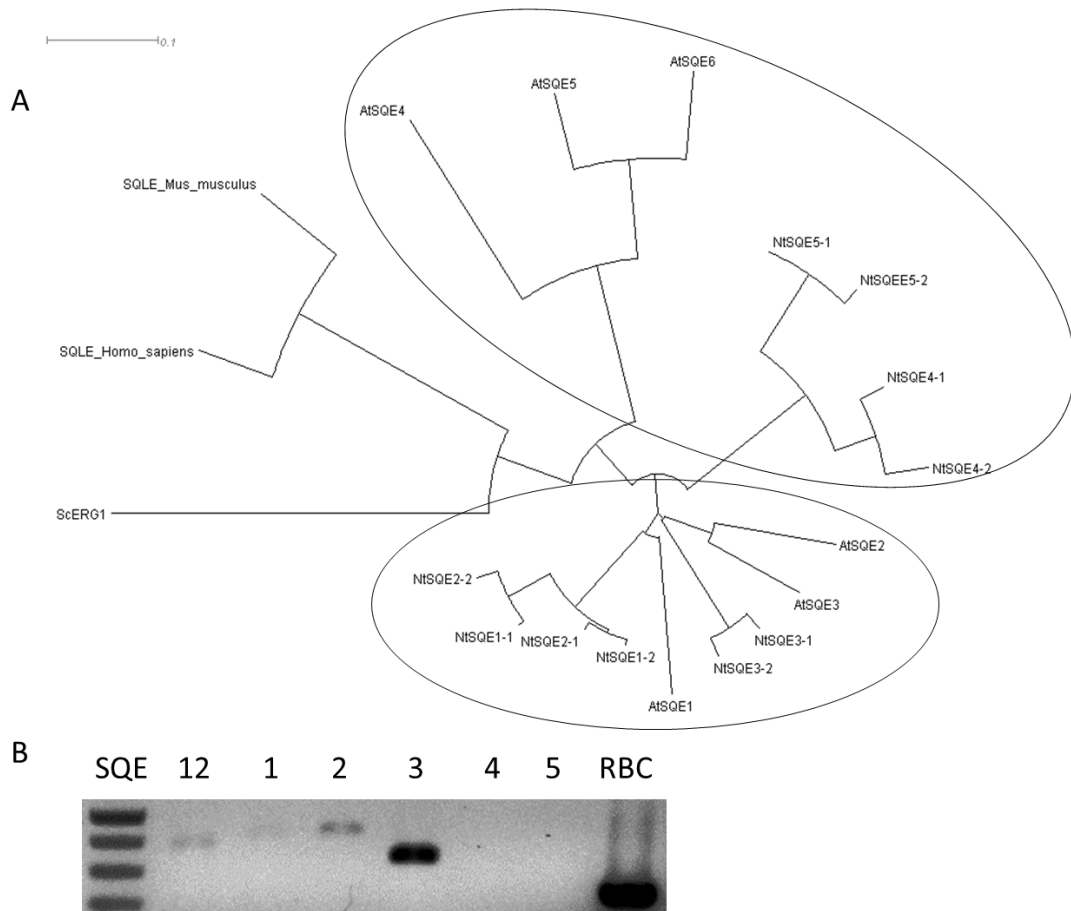
Figure 3-2. Multiple squalene epoxidases in the *N. tabacum* genome. (A) Phylogenetic tree of squalene epoxidases in Arabidopsis, yeast, animals and tobacco. *A. thaliana* (At), *N. tabacum* (Nt), *Homo sapiens* (Hs), *Mus musculus* (Mm), *Saccharomyces cerevisiae* (Sc). (B) Reverse-transcriptase PCR products of tobacco leaf cDNA from FPS-SQS plants.

3.3.3 Down regulation of SQE3 led to the squalene increase

In order to knockdown SQE3, we designed an artificial microRNA based on the gene sequence. *A. thaliana* artificial microRNA (amiRNA) 159a was used as a backbone containing 21 bps complementary with the SQE3 specific coding region (Figure 3-3). The amiRNA-SQE was transformed into plants in constructs FPS-SQS-amiRNA-SQE, amiRNA-SQE, and FPS-SQS, where FPS-SQS stands for the chloroplast-located over-expression of FPS and SQS and amiRNA-SQE is expected to knock down cytosolic expression of SQE3 (Figure 3-4A). To check effectiveness of RNA silencing, we examined the expression levels of SQE3 by semi-quantitative RT-PCR in T0 generation plants. The representative samples in Figure 3-4B showed that SQE3 mRNA was down-regulated in those lines that expressed amiRNA-SQE. In addition, the specificity of the amiRNA was investigated by RT-PCR on all the SQEs to see if the RNA expression levels of other SQE were influenced by amiRNA-SQE targeting to SQE3. Specific primers were designed based on the unique sequences of each SQEs. The RT-PCR results showed that SQE3 expression in FPS-SQS- amiRNA-SQE #4 plant was down-regulated, while other SQE1 and SQE2 expression did not change, compared to FPS-SQS plant (Figure 3-5).

```
                    cov    pid   1 [        .         .         .         .         :         .         .         .         .         .         1         .          . 120
 1 ScERG1      100.0% 100.0%   ------------------------------------------------------------------------------------------------------------------------
 2 HsSQLE       92.9%  27.9%   ------------------------------------------------MWTFLGIATFTYFYKKF-GDFITLA---------NREVLLCVLVFLSLGLVLSYRCRHRNGGLLGRQQSGSQFALFSD
 3 MmSQLE       92.9%  28.2%   ------------------------------------------------MWTFLGIATFTYFYKKC-GDV-TLA---------NKELLLCVLVFLSLGLVLSYRCRHRHGGLLGRHQSGAQFAAFSD
 4 AtSQE4       91.7%  21.9%   ---------------------------------------------------------MTYAWLWTLLAFVLTWMVFHLIKMKK--------------
 5 AtSQE5       91.9%  21.9%   ------------------------------------------------------MAFTNVCLWTLLAFMLTWTVFYVTNRGK--------------
 6 AtSQE6       92.1%  22.7%   ------------------------------------------------------MAFTHVCLWTLVAFVLTWTVFYLTNMKK--------------
 7 NtSQE5-1     91.1%  26.6%   -------------------------------------------------MELVILFIASLFGLLLLYNMIRRKIMEI--------------
 8 NtSQE5-2     53.4%  28.1%   ------------------------------------------------------------------------------------------
 9 NtSQE4-1     91.7%  27.3%   -------------------------------------------------MEMIIFLGSFIASLFGLLLLYNMIRRKIMEI--------------
10 NtSQE4-2     75.6%  24.9%   -------------------------------------------------MEMIIFLGSFIASLFGLLLLYNMIRRKIMEI--------------
11 AtSQE2       92.1%  26.0%   MKPFVIRNLPRFQSTLRSSL--LYTNHRPSSR--FSLSTRR----FTTGATYIRRWKATAAQTLKLSAVNSTVMMKPAKIALDQFIASLFTFLLLYILRRSSNKNK--------------
12 AtSQE3       91.3%  28.7%   -----------------------------------------------------MAPTIFVDHCILTTTFVASLFAFLLLYVLRRRSKTIH--------------
13 AtSQE1       91.7%  29.0%   -------------------------------------------ME--SQLWNWILPLLIS----SLLISFVAFYGFFVKPKRNG--------------
14 NtSQE3-1     91.9%  26.0%   MLLINIPNNPPFSLVFSKHNFRVSTNQTPPQLSQPSASSKKKKLALISFSKYLR---NTQLQFISTN--EKMVNFMMDKYIVPTFFVSLLGFLLLYILRP--RLRT--------------
15 NtSQE3-2     91.9%  26.3%   MLLINTPNNISFSLVFSKHKFPVSTNQTPPQLSQPSASSTKKKLALISFSKYLR---NTQLQFISTN--EKMVNFMMDKYIVPTFFVSLLGFLLLYILRP--RFRT--------------
16 NtSQE1-2     92.1%  29.5%   ------------------------------------------------MVDMGLSGSVLAVVFGFVAVLWFFI--QRNG--------------
17 NtSQE2-1     92.1%  29.3%   ------------------------------------------------MVDMGLSGSVLAVVFGFVAVLWFFI--QRNG--------------
18 NtSQE1-1     74.4%  29.1%   ------------------------------------------------MVDMGLSGSVLAVVFGFVAVLWFFI--QRNG--------------
19 NtSQE2-2     76.6%  28.5%   ------------------------------------------------MVDMGLSGSVLAVVFGFVAVLWFFI--QRNG--------------
   consensus/100%             ..............................................................................................................................
   consensus/90%              ..............................................................................................................................
   consensus/80%              ........................................................h...hhh..huhlh.h.hh....h.h..........
   consensus/70%              ........................................................hhh..shlhslhuhlhhhhhhhh..thth............
```

```
                    cov    pid  121         .         .         :         .         .         .         .         2         .         .         .          . 240
 1 ScERG1      100.0% 100.0%   --------------------------------------------MSA-VN-VAPELINADNTITYDAIVIGAGVIGPCVATGLARKGKKVLIVERDWAMPDRIVGELMQPGGVRALRSLGMIQSINNIE
 2 HsSQLE       92.9%  27.9%   ILSGLPFIGFFWAKSPPESENKEQLEARRRRKGTNISETSLIGTAACTSTSSQNDPEVIIVGAGVLGSALAAVLSRDGRKVTVIERDLKEPDRIVGEFLQPGGYHVLKDLGLGDTVEGLD
 3 MmSQLE       92.9%  28.2%   ILSALPLIGFFWAKSP-ESEKKEQLESKKCRKEIGLSETTLTGAATSVSTSFVTDPEVIIVGSGVLGSALAAVLSRDGRKVTVIERDLKEPDRIVGELQPGGYRVLQELGLGDTVEGLN
 4 AtSQE4       91.7%  21.9%   ---------------AAT--------------------G-DLEAEAEARRDGATDVIIVGAGVAGASLAYALAKDGRRVHVIERDLKEPQRFMGELMQAGGRFMLAQLGLEDCLEDID
 5 AtSQE5       91.9%  21.9%   ---------------KAT--------------------Q-LADAVVEEREDGATDVIIVGAGVGGSALAYALAKDGRRVHVIERDLREPERIMGEFMQPGGRLMLSKLGLEDCLEGID
 6 AtSQE6       92.1%  22.7%   ---------------KAT--------------------D-LADTVAEDQKDGAADVIIVGAGVGGSALAYALAKDGRRVHVIERDMREPERMMGEFMQPGGRLMLSKLGLQDCLEDID
 7 NtSQE5-1     91.1%  26.6%   ---------------RNI-----------------TS--NV--VVMKDHKSYGNTDVIIVGAGVAGSALACTLGKDGRRVHVIERDLTEPDRMVGELLQPGGYKLVELGLEDCLEDID
 8 NtSQE5-2     53.4%  28.1%   ----------------------------------------------------------------------------------------------------------------------
 9 NtSQE4-1     91.7%  27.3%   ---------------RNI-----------------AS--NV--VLIKDHKSYGKTDVIIVGAGVAGSALACTFGKDGRRVHVIERDLTEPHRMVGELLQPGGYKLVDLGLEDCLEDID
10 NtSQE4-2     75.6%  24.9%   ---------------RNI-----------------AS--NV--VLIKDHKSYGKTDVIIVGAGVAGSALACTFGKDGRRVHVIERDLTEPHRMVGELLQPGGYKLVDLGLEDCLEDID
11 AtSQE2       92.1%  26.0%   ---------------KNRGL--------------VVSQ--NDTVSKNLETEVDSGTDVIIVGAGVAGAALAHTLGKEGRRVHVIERDFSEQDRIVGELLQPGGYVLKLIELGLEDCVKKID
12 AtSQE3       91.3%  28.7%   ---------------G---S--------------VNVR--NG----TLTVKSGTDVDIIIVGAGVAGAALAHTLGKEGRRVHVIERDLTEPDRIVGELLQPGGYKLIELGLEDCVKDID
13 AtSQE1       91.7%  29.0%   ---------------LRHDRKTVS----------TVTS--D--VG-SVNITGDTVADVIVVGAGVAGSALAYTLGKDKRRVHVIERDLSEPDRIVGELLQPGGYKLLELGIEDCVEEID
14 NtSQE3-1     91.9%  26.0%   ---------------GPYYKKKDPKTSQKCDTQNVISS--NLTNGECKLEK-GSDADIIIVGAGVAGAALAHTLAKEGRRVLVIERDLTEPDRIVGELLQPGGYKLIELGLEDCVEDID
15 NtSQE3-2     91.9%  26.3%   ---------------PYYKKKDPKSIQKCETHNVISS--NLINGECKLEK-GSDADIIIVGAGVAGAALAHTLAKEGRKVLVIERDLTEPDRIVGELLQPGGYKLIELGLEDCVEDID
16 NtSQE1-2     92.1%  29.5%   ---------------GRKDKNNDPATT------STTTT--TVYDGECRSKDANDDADIIIVGAGVAGAALAHTLKEGRHVKVIERDLTEPDRIVGELLQPGGFLKLQELGLEDCVENID
17 NtSQE2-1     92.1%  29.3%   ---------------GRKDKNNDPATAT----TSTTTT--TVYDGECRSKDANDNADIIIVGAGVAGAALAHTLGKEGRRVKVIERDLTEPDRIVGELLQPGGFLKLQELGLEDCVENID
18 NtSQE1-1     74.4%  29.1%   ---------------GRKDKNNDPATT------STTTT--TVYDGECRSKDANDDADIIIVGAGVAGAALAHTLKEGRHVKVIERDLTEPDRIVGELLQPGGFLKLQELGLEDCVENID
19 NtSQE2-2     76.6%  28.5%   ---------------GRKDKNNDPATAT----TSTTTT--TVYDGECRSKDANDNADIIIVGAGVAGAALAHTLGKEGRRVKVIERDLTEPDRIVGELLQPGGFLKLQELGLEDCVENID
   consensus/100%             ..............................................................................................................................
   consensus/90%              ...................................................h...p......hIllGuGVhGsslAhshu+ct++VhllERDht..pRhhGEhhQsGGhhhL.pLGh.pslptlp
   consensus/80%              ..........................................t...s..hptp..sssDlIIVGAGVhGuALAhsLu+-GR+VhVIERDhpEPcRhhGEhhQPGGhhhL.cLGLtDslcsID
   consensus/70%              ..............h.h..........sp..s...s.hhppctssssDlIIVGAGVuGuALAtsLuK-GR+VpVIERDLpEP-RhVGELLQPGGaLhL.-LGLEDClEsID
```

Figure 3-3. Amino acid sequence alignment of squalene epoxidases from plants, yeast and animals. The *N. tabacum* SQE genes were aligned with other squalene epoxidases using Clustal omega. Identical sequences are marked as dark letters.

57

```
                  cov     pid 241      :        .        .        .       3        .        .        .        :        . 360
 1 ScERG1     100.0% 100.0% AYPVTGYTVFF-NG-EQVDIPYPYKADIPKVEKLKDLVKDGNDKVLEDSTIHIKDYEDDERERGVAFVHGRFLNNLRNITAQEPNVTRVQGNCIEILKDEKNEVVGAKVDID-GRGKVEF
 2 HsSQLE      92.9%  27.9% AQVVNGYMIHDQESKSEVQIPYPLSENNQ-------------------------------VQSGRAFHHGRFIMSLRKAAMAEPNAKFIEGVVLQLLEE-DDVVMGVQYKDKETGDIKEL
 3 MmSQLE      92.9%  28.2% AHHIHGYIVHDYESRSEVQIPYPLSETNQ-------------------------------VQSGIAFHHGRFIMSLRKAAMAEPNVKFIEGVVLQLLEE-DDAVIGVQYKDKETGDTKEL
 4 AtSQE4      91.7%  21.9% AQEAKSLAIYK-DG-KHATLPFPDD-KSF----------------------------------PHEPVGRLLRNGRLVQRLRQKAASLSNVQLEEGTVKSLIEE-EGVVKGVTYKNSAGEEI-TA
 5 AtSQE5      91.9%  21.9% AQKATGMTVYK-DG-KEAVASFPVDNNNF---------------------------------PFDPSARSFHNGRFVQRLRQKASSLPNVRLEEGTVKSLIEE-KGVIKGVTYKNSAGEET-TA
 6 AtSQE6      92.1%  22.7% AQKATGLAVYK-DG-KEADAPFPVDNNNF---------------------------------SYEPSARSFHNGRFVQQLRRKAFSLSNVRLEEGTVKSLLEE-KGVVKGVTYKNKEGEET-TA
 7 NtSQE5-1    91.1%  26.6% AQRVGGHVLYKKNG-EHIMLSYPLEK-FQ-----------------------------------AVDASGRCFHNGRFVQKMREKAATLPNVRLEKGTVTSLIEE-KGSVKGVNYKTKDGREL-TA
 8 NtSQE5-2    53.4%  28.1% ------------------------------------------------------------------------------------------------------------------------
 9 NtSQE4-1    91.7%  27.3% AQRVVGYTLYKKNG-EHIMLSYPLEK-FQ-----------------------------------AADMSGRCFHNGRFVQKMREKAATLPNVRLEQGTVTSLIEE-KGTVKGVNYKTKDGREL-SA
10 NtSQE4-2    75.6%  24.9% AQRVVGYTLYKKNG-EHIMLSYPLEK-FQ-----------------------------------AADMSGRCFHNGRFVQKMREKAATLPNVRLEQGTVTSLIEE-KGTVKGVNYKTKDGREL-SA
11 AtSQE2      92.1%  26.0% AQRVLGYVLFK-DG-KHTKLAYPLET-FD-----------------------------------S-DVAGRSFHNGRFVQRMREKALTLSNVRLEQGTVTSLIEE-HGTIKGVRYRTKEGNEF-RS
12 AtSQE3      91.3%  28.7% AQRVLGYALFK-DG-KHTKLSYPLDQ-FD-----------------------------------S-DVAGRSFHNGRFVQRMREKASLLPNVRMEQGTVTSLVEE-NGIIKGVQYRTKDGQEL-KS
13 AtSQE1      91.7%  29.0% AQRVVGYALFK-NG-KRIRLAYPLEK-FH-----------------------------------E-DVSGRSFHNGRFIQRMREKAASLPNVQLEQGTVLSLLEE-NGTIKGVRYKNKAGEEQ-TA
14 NtSQE3-1    91.9%  26.0% AQRVVGYALFK-DG-KSTNVSYPLEN-FH-----------------------------------S-DVAGRSFHNGRFIQKMREKAATFPNVRLEQGTVTSLIEE-NGSVKGIQYKTKAGQEL-KA
15 NtSQE3-2    91.9%  26.3% AQRVVGYALFK-DG-KSTNVSYPLEN-FH-----------------------------------S-DVAGRSFHNGRFIQKMREKAATFPNVRLEQGTVTSLIEE-NGSVKGIQYKTKAGQEL-KA
16 NtSQE1-2    92.1%  29.5% AQRVFGYALFK-DG-KSTRLSYPLEK-FH-----------------------------------A-DVSGRSFHNGRFIQRMREKAASLPNVKLEQGTVTSLLEE-NGTIRGVQYKNKSGEEL-KA
17 NtSQE2-1    92.1%  29.3% AQRVFGYALFK-DG-KSTRLSYPLEK-FH-----------------------------------A-EVSGRSFHNGRFIQRMREKAASLPNVKLEQGTVTSLLEE-NGTIRGVQYKNKSGEEL-KA
18 NtSQE1-1    74.4%  29.1% AQRVFGYALFK-DG-KSTRLSYPLEK-FH-----------------------------------A-DVSGRSFHNGRFIQRMREKAASLPNVKLEQGTVTSLLEE-NGTIRGVQYKNKSGEEL-KA
19 NtSQE2-2    76.6%  28.5% AQRVFGYALFK-DG-KSTRLSYPLEK-FH-----------------------------------A-EVSGRSFHNGRFIQRMREKAASLPNVKLEQGTVTSLLEE-NGTIRGVQYKNKSGEEL-KA
   consensus/100%                  ........................................................................................................
   consensus/90%   A..h.uhhla..pu.pph.hsaP.p...........................uh.hhpGRhl.phRphs...sNsph.pGsshpllc-.ps.lhGhphc.p.stt..p.
   consensus/80%   AQcshGhslac.sG.cph.lsaPlcp.........................c.uGRsFHpGRFlpphRptAhs.sNVph.pGsVhpLlEE.psslhGVpYKsctGp-h.ph
   consensus/70%   AQ+VhGYslaK.sG.cpsplsYPL-p..p..................s.-.uGRuFHNGRFlQ+hRcKAuoLPNV+LEpGTVpSLlEE.pGsl+GVpYKsKtGpEh.pu

                  cov     pid 361      .        .        .       4        .        .        .        :        . 480
 1 ScERG1     100.0% 100.0% KAHLTFICDGIFSRFRKELHPDHVPTVGSSFVGMSL--FNAKNPAPMHGHVILGSDHMPILVYQISPEETRILCAYNSPKVP----ADIKSWMIKDVQPFIPK--SLRPSFDEAVS-QG-
 2 HsSQLE      92.9%  27.9% HAPLTVVADGLFSKFRKSLVS-NKVSVSSHFVGFLM--KNAPQFKANHAELILA-NPSPVLIYQISSSETRVLVDIRGEMP-----RNLREYMVEKIYPQIPD--HLKEPFLEATD-NS-
 3 MmSQLE      92.9%  28.2% HAPLTVVADGLFSKFRKSLIS-SKVSVSSHFVGFLM--KDAPQFKPNFAELVLV-NPSPVLIYQISSSETRVLVDIRGELP-----RNLREYMAEQIYPQLPE--HLKESFLEASQ-NG-
 4 AtSQE4      91.7%  21.9% FAPLTVVCDGCYSNLRRSLVD-NTEEVLSYMVGYVT--KNSRLEDPHSLHIFS-KPLVCVIYQITSDEVRCVAEVPADSIPSISNGEMSTFLKKSMAPQIPETGNLREIFLKGIEEGLP
 5 AtSQE5      91.9%  21.9% LAPLTVVCDGCYSNLRRSLND-NNAEVLSYQVGFIS--KNCQLEEPEKLKLIMS-KPSFTMLYQISSTDVRCVFEVLPNNIPSISNGEMATFVKNTIAPQVPL--KLRKIFLKGIDEGE-
 6 AtSQE6      92.1%  22.7% LAPLTVVCDGCYSNLRRSLNDDNNAEIMSYIVGYIS--KNCRLEEPEKLHLILS-KPSFTMVYQISSTDVRCGFEVLPENFPSIANGEMSTFMKNTIVPQVPP--KLRKIFLKGIDEGA-
 7 NtSQE5-1    91.1%  26.6% YAPLTIVCDGCFSNLRRSLCN-PKMDIPSTYVGLIL--KDCQLPYANHGVLVMS-DPSPVTFYPISSTEIRCLVLIAGQKVPSIANGEMVNYLKNVVASQVPS--ELRDALIRAIGEEN-
 8 NtSQE5-2    53.4%  28.1% --------------------------MDIPSTYVGLIL--KDCQLPYANHGVLVMS-DPSPVTFYPISSTEIRCLVLIAGQKVPSIANGEMVNYLKNVVASQVPS--ELRDALIRAIGEEN-
 9 NtSQE4-1    91.7%  27.3% YAPLTIVCDGCFSNLRRSLCN-LKMDIPSTYVGLIL--KDCQLPYANHGVLVMS-DPSPVTFYPVSSTEVRCLVLIAGQNVPSIANGEMVNYLTNVVAPQVPS--ELRDALITAISEEN-
10 NtSQE4-2    75.6%  24.9% YAPLTIVCDGCFSNLRRSLCN-LKMDIPSTYVGLIL--KDCQLPYANHGVLVMS-DPSPVTFYPVSSTEVRCLVLIAGQNVPSIANGEMVNYLTNVVAPQVPS--ELRDALITAISEEN-
11 AtSQE2      92.1%  26.0% FAPLTIVCDGCFSNLRRSLCK-ENCELPFANHGVVLG-DPSPILMYPISSSEVRCLVDVPGQKLPPIANGEMAKYLKTRVAPQVPT--KVREAFITAVE-KG-
12 AtSQE3      91.3%  28.7% FAPLTIVCDGCFSNLRRSLCK-PKVEVPSNFVGLVL--ENCELPFPNHGVVLG-DPSPILFYPISSSEVRCLVDVPGSKLPSVASGEMAHHLKTMVAPQVPP--QIRDAFISAVE-KG-
13 AtSQE1      91.7%  29.0% FAALTIVCDGCFSNLRRSLCN-PQVEVPSCFVGLVL--ENCNLPYANHGVVLA-DPSPILMYPISSTEVRCLVDVPGQKVPSIANGEMKNYLKTVVAPQMPH--EVYDSFIAAVD-KG-
14 NtSQE3-1    91.9%  26.0% HAPLTVVCDGCFSNLRRSLCN-PKVDIPSCFVGLVLELENNQLPYPNHGVILA-DPSPILFYPISSTEIRCLVDVPGQKLPSLANGDMANYLKTMVAPQVPP--ELHDAFITAID-KG-
15 NtSQE3-2    91.9%  26.3% HAPLTVVCDGCFSNLRRSLCN-PKVDIPSCFVGLVLELENDQLPYPNHGVILA-DPSPILFYPISSTEIRCLVDVPGQKLPSLANGDMANYLKTMVAPQVPP--ELHDAFITAID-KG-
16 NtSQE1-2    92.1%  29.5% YAPLTIVCDGCFSNLRRTLCD-PKVEVPSCFVGLVL--ENCQLPHANHGVILA-DPSPILFYPISSTEVRCLVDVPGQKVPSISNGEMAKYLKSVVAPQVPP--EIKDAFIAAID-KG-
17 NtSQE2-1    92.1%  29.3% YAPLTIVCDGCFSNLRRTLCD-PKVEVPSCFVGLVL--ENCQLPHANHGVILA-DPSPILFYPISSTEVRCLVDVPGQKVPSISNGEMGKYLKSIVAPQVPP--EIKDAFIAAID-KG-
18 NtSQE1-1    74.4%  29.1% YAPLTIVCDGCFSNLRRTLCD-PKVEVPSCFVGLVL--ENCQLPHANHGVILA-DPSPILFYPISSTEVRCLVDVPGQKVPSISNGEMAKYLKSVVAPQVPP--EIKDAFIAAID-KG-
19 NtSQE2-2    76.6%  28.5% YAPLTIVCDGCFSNLRRTLCD-PKVEVPSCFVGLVL--ENCQLPHANHGVILA-DPSPILFYPISSTEVRCLVDVPGQKVPSISNGEMGKYLKSIVAPQVPP--EIKDAFIAAID-KG-
   consensus/100%                  .........................pl.S..VGh.....ss....s..h.llhs.p...hhhY.losp-hRhhh.h.s.........tph.pahhp.h.s.hP...plh..h.tu.t.t..
   consensus/90%   hA.LThlsDGhaSphR+pL......pl.S.hVGhl...css....spph.llhu.pP..hhhY.lSSs-hRslh.l.sp.......tph.pahhp.lhsQlP...pl+c.hltuht.tt.
   consensus/80%   hAPLTlVCDGhaSphR+oLss..ph-lsSpaVGhlh..csspL.hsp+uhllhu.cPSPlhhY.ISSoElRCLs.l.upplPslusG-Mhpahpphls PQlP...cl+-shltAlp.ps.
   consensus/70%   aAPLTlVCDGCFSNLRRoLss.sps-lsSsaVGhlL..cNspLPhsNHGcllhu.DPSPlhhY.ISSoElRCLV-lsGpplPSluNGEMspYLKshVAPQVPs..cl+-uFlpAls.ps.
```
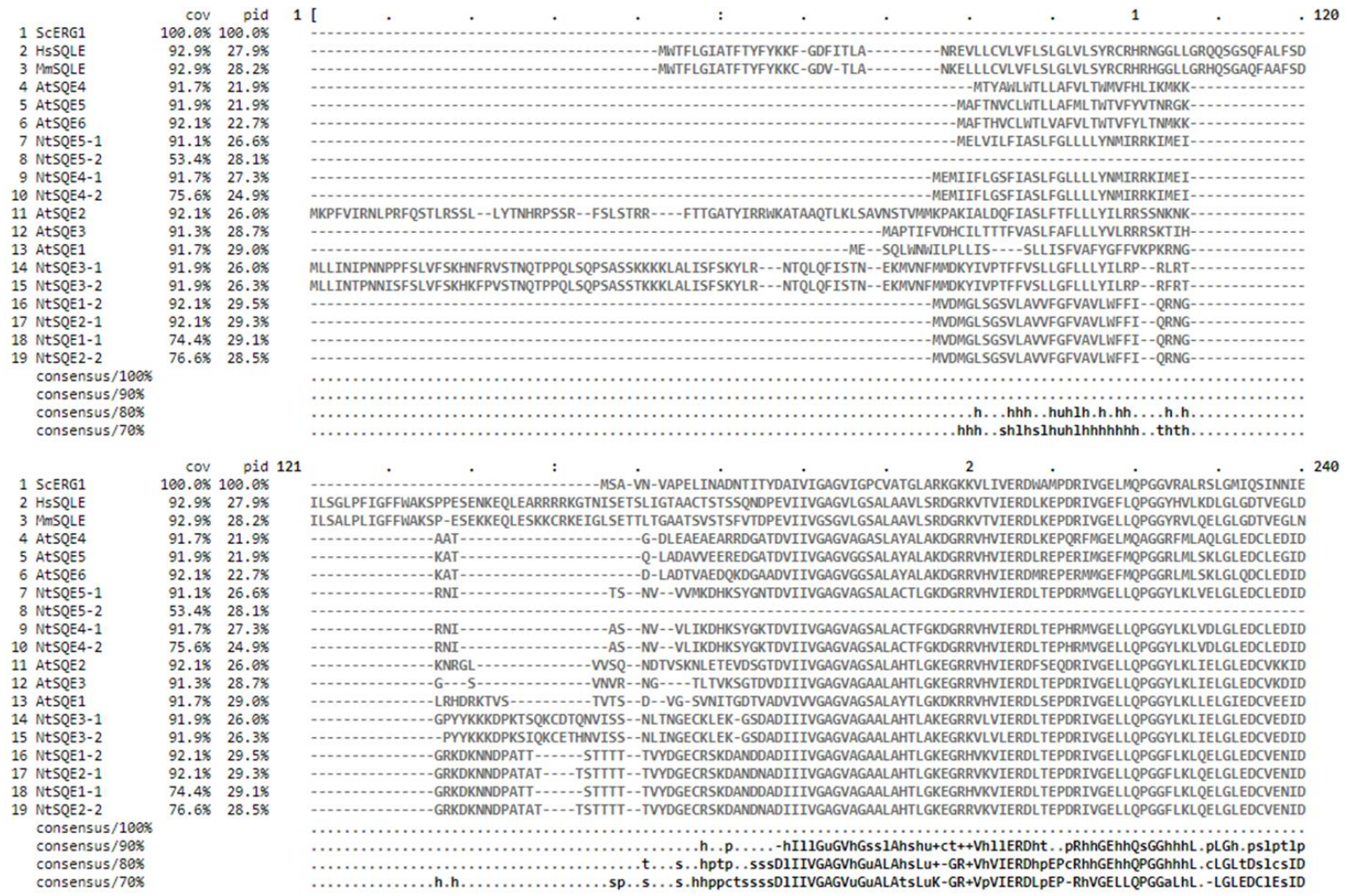
Figure 3-3. Continued

58

```
                 cov     pid 481        .        5        .        .        .        :        .        .        .        .     6 600
 1 ScERG1    100.0% 100.0%   KFRAMPNSYLPARQNDVTGMCVIGDALNMRHPLTGGGMTVGLHDVVLLIKKIGDL-DFSDREKVLDELLDYHFERKSYDSV-INVLSVALYSLFAAD-SDNLKALQKGCFKYFQRGGDCV
 2 HsSQLE     92.9%  27.9%   HLRSMPASFLPPSSVKKRGVLLLGDAYNMRHPLTGGGMTVAFKDIKLWRKLLKGIPDLYDDAAIFEAKKSFYWARKTSHSFVVNILAQALYELFSAT-DDSLHQLRKACFLYFKLGGECV
 3 MmSQLE     92.9%  28.2%   RLRTMPASFLPPSSVNKRGVLILGDAYNLRHPLTGGGMTVALKDIKLWRQLLKDIPDLYDDAAIFQAKKSFFWSRKRTHSFVVNVLAQALYELFSAT-DDSLHQLRKACFLYFKLGGECV
 4 AtSQE4     91.7%  21.9%   EIKSTATKSMSSRLCDKRGVIVLGDAFNMRHPIIASGMMVALSDICILRNLLKPLPNLSNTKKVSDLVKSFYIIRKPMSAT-VNTLASIFSQVLVATTDEAREGMRQGCFNYLARGDFKT
 5 AtSQE5     91.9%  21.9%   HIKAMPTKKMTATLSEKKGVILLGDAFNMRHPAIASGMMVLLSDILILRRLLQPLSNLGNAQKISQVIKSFYDIRKPMSAT-VNTLGNAFSQVLVASTDEAKEAMRQGCYDYLSSGGFRT
 6 AtSQE6     92.1%  22.7%   HIKVVPAKRMTSTLSKKKGVIVLGDAFNMRHPVVASGMMVLLSDILILRRLLQPLSNLGDANKVSEVINSFYDIRKPMSAT-VNTLGNAFSQVLIGSTDEAKEAMRQGVYDYLCSGGFRT
 7 NtSQE5-1   91.1%  26.6%   NIRRMQFRSMAANTCTIPGSLLIGDAFNMRHPLSAGGMTVALSDIVVLRHLLKPLPNFTDANAV---LQRFYTLRKPVAST-INILAGVTYQVFSPSLDPSRKEIREAFYGYLRLGGIFT
 8 NtSQE5-2   53.4%  28.1%   NIRRMQFRSMAANTCTIPGSLLIGDAFNMRHPLSAGGMTVALSDIVVLRHLLKPLPNFTDANAV---LQRFYTLRKPVAST-INILAGVTYQVFSPSLDPSRKEIREAFYGYLRLGGIFT
 9 NtSQE4-1   91.7%  27.3%   NIRTMQFRSMPATCSTIPGSLLIGDAFNMRHPLSAGGMTVALSDIVVLRHLLKPLTDFNDANAVSKYLQRFYTLRKPLAST-INILAGVTYQVFSPSLDPSRKEIREVFYGYLRVGGIFT
10 NtSQE4-2   75.6%  24.9%   NIRTMQFRSMPATCSTIPGSLLIGDAFNMRHPLSAGGMTVALSDIVVLRHLLKPLTDFNDANAVSKYLQRFYTLRKYILLV-AYC-----FHFLLPREF---CALPRCFM----------
11 AtSQE2     92.1%  26.0%   NIRTMPNRSMPADPIPTPGALLLGDAFNMRHPLTGGGMTVALADIVVLRDLLRPIRNLNDKEALSKYIESFYTLRKPVAST-INTLADALYKVFLASSDEARTEMREACFDYLSLGGVFS
12 AtSQE3     91.3%  28.7%   NIRTMPNRSMPADPIHTPGALLLGDAFNMRHPLTGGGMTVALSDIILNPLVDLTNKESLSKYIESFYTLRKPVAST-INTLAGALYKVFLASPDDARSEMRRACFDYLSLGGVCS
13 AtSQE1     91.7%  29.0%   NIKSMPNRSMPASPYPTPGALLMGDAFNMRHPLTGGGMTVALADIVVLRNLLRPLRDLSDGASLCKYLESFYTLRKPVAAT-INTLANALYQVFCSSENEARNEMREACFDYLGLGGMCT
14 NtSQE3-1   91.9%  26.0%   HIRTMPNRSMPAAPYPTPGALLLGDSFNMRHPLTGGGMTVALSDIAVLRNLLKPLNDLNDADELCKYLESFYTLRKPVAST-INTLAGALYKVFCASPDQARKEMREACFDYLGLGGTCS
15 NtSQE3-2   91.9%  26.3%   HIRTMPNRSMPAAPYPTPGALLLGDSFNMRHPLTGGGMTVALSDIAVLRNLLKPLDNLNDADELCKYLESFYTLRKPVAST-INTLAGALYKVFCASPDQARKEMREACFDYLSLGGTCS
16 NtSQE1-2   92.1%  29.5%   NIRTMPNRSMPAAPHPTPGALLMGDAFNMRHPLTGGGMTVALSDIVVLRDLLKPLRDLNDAPTLCRYLESFYTLRKPVAST-INTLAGALYKVFCASPDQARKEMRDACFDYLSLGGVFS
17 NtSQE2-1   92.1%  29.3%   NIRTMPNRSMPAAPHPTPGALLMGDAFNMRHPLTGGGMTVALSDIVVLRDLLKPLRDLNDAPTLCRYLESFYTLRKPVAST-INTLAGALYKVFCASPDQARKEMRDACFDYLSLGGVFS
18 NtSQE1-1   74.4%  29.1%   NIRTMPNRSMPAAPHPTPGALLMGDAFNMRHPLTGGGMTVALSDIVVLRDLLKPLRDLNDAPTLCRYLESFYTLRKVNIILNVSSHKQKIWSWYLLST----------------------
19 NtSQE2-2   76.6%  28.5%   NIRTMPNRSMPAAPHPTPGALLMGDAFNMRHPLTGGGMTVALSDIVVLRDLLKPLRDLNDAPTLCRYLESFYTLRKVKIILNLSSHKQKIWSRYLFSTLRIPSEVRALC-----------
   consensus/100%             ph+.h..p.hsst...h.G.h1hGDuhNhRHPh.uuGMhVhhtDlh1hhphltsl.sh.st.tl...h.paa..RK....h.h........phh..p..............
   consensus/90%              pl+th..p.hsst...h.Gsl1hGDuaNMRHPl.uuGMhVhLtDIh1hRpLLps1.sh.st.tl...hppFa..RK....h.ls.ht.hh.phh..s.....pth.thh..........
   consensus/80%              pI+sM..+pMsus...h.Gsl1hGDAFNMRHPLouGGMTVALuDIh1LRpLLpPL.shsDs.tl.phlppFYhlRKshtuh.1NhLuthhaplh.so.s.shpthRpsha.Yht.Gs.hs
   consensus/70%              pIRoMPs+SMPAsshspsGsLL1GDAFNMRHPLouGGMTVALSDIllLRcLL+PL.sLsDsstlschlpSFYslRKshsus.lNsLusshaplahsS.DpuhpthRcusatYhphGG.hs
```

```
                 cov     pid 601        .        .        .        .        .        :        .     ] 697
 1 ScERG1    100.0% 100.0%   NKPVEFLSGVLPKPLQLTRVFFAVAFYTIYLNMEERGFLGLPMALLEGIMILITAIRVFTPFLFGELIG----------------------------
 2 HsSQLE     92.9%  27.9%   AGPVGLLSVLSPNPLVLIGHFFAVAIYAVYFCFKSEPWITKPRALLSSGAVLYKACSVIFPLIYSEMKYMVH--------------------------
 3 MmSQLE     92.9%  28.2%   TGPVGLLSILSPHPLVLIRHFFSVAIYATYFCFKSEPWATKPRALFSSGAVLYKACSILFPLIYSEMKYLVH--------------------------
 4 AtSQE4     91.7%  21.9%   RGLMTILGGMNPHPLTLVLHLVAITLTSMGHLLSPF---PSPRRFWHSLRILAWALQMLGAHLVDEGFKEMLIPTNAAAYRRNYIATTTV--------
 5 AtSQE5     91.9%  21.9%   SGMMALLGGMNPRPISLIYHLCAITLSSIGHLLSPF---PSPLRIWHSLRLFGLAMKMLVPHLKAEGVSQMLFPVNAAAYSKSYMAATAL--------
 6 AtSQE6     92.1%  22.7%   SGMMALLGGMNPRPLSLVYHLCAITLSSIGQLLSPF---PSPLRIWHSLKLFGLAMKMLVPNLKAEGVSQMLFPANAAAYHKSYMAATTL--------
 7 NtSQE5-1   91.1%  26.6%   NGAIGLLAGLNPTPLGFAFQFIAVAVYTFGRLLLPF---PTPNAILRGARLLYVAIDMIFPIIKAEGIRQMFFPASLPAYYSAPALAHKSDQ------
 8 NtSQE5-2   53.4%  28.1%   NGAIGLLAGLNPTPLGFAFQFIAVAVYTFGRLLLPF---PTPNAILRGARLLYVAIDMIFPIIKAEGIRQMFFPASLPAYYSAPALAHKSDQ------
 9 NtSQE4-1   91.7%  27.3%   NGAIGLLAGLNPRPLSLPFHFIAVAVYTFGRLLLPF---PTPKGILRAATLLYVACGMIFPIIKAEGIKHMFFPAIFPA-------------------
10 NtSQE4-2   75.6%  24.9%   ------------------LHVV---------------------------------------------------------------------------
11 AtSQE2     92.1%  26.0%   SGPVALLSGLNPRPLSLVHFFAVAIYAVCRLMLPF---PSIESFWLGARIISSASSIIFPIIKAEGVRQMFFPRTIPAIYRAPP-------------
12 AtSQE3     91.3%  28.7%   SGPVALLSGLNPRPMSLVLHFFAVAIFGVGRLLVPL---PSVKRLWLGARLISSASGIIFPIIKAEGVRQMFFPRTIPAIYRAPPTPSSSSPQ----
13 AtSQE1     91.7%  29.0%   SGPVSLLSGLNPRPLTLVCHFFAVAVYGVIRLLIPF---PSPKRIWLGAKLISGASGIIFPIIKAEGVRQMFFPATVPAYYYKAPTVGETKCS----
14 NtSQE3-1   91.9%  26.0%   TGPVALLSGLNPSPLSLVLHFFAVAIYGVGRLLVPF---PSPKRLWIGARLISAASSIIFPIIKAEGVRQMFFPTTIPAYHRAPPVNKGSDQIKQNK
15 NtSQE3-2   91.9%  26.3%   TGPVALLSGLNPSPLSLVLHFFAVAIYGVGRLLVPF---PSPKRLWIGARLISAASGIIFPIIKAEGVRQMFFPTTIPAYHRAPPVNKGSD------
16 NtSQE1-2   92.1%  29.5%   TGPVSLLSGLNPRPLSLVCHFFAVAIYGVGRLLLPF---PSPKRMWIGARLISGASGIIFPIIKAEGVRQMFFPATVPAYYRSPPEVKL--------
17 NtSQE2-1   92.1%  29.3%   TGPVSLLSGLNPRPLSLVCHFFAVAIYGVGRLLLPF---PSPKRMWIGARLISGASGIIFPIIKAEGVRQMFFPATVPAYYRSPPEVKL--------
18 NtSQE1-1   74.4%  29.1%   ------------------------------------------------------------------------------------------------
19 NtSQE2-2   76.6%  28.5%   ------------------------------------------------------------------------------------------------
   consensus/100%             ................................................................................................
   consensus/90%              ................................................................................................
   consensus/80%              tt.hthLuhh.PpPh.h.hphhulsh.shh..h......s...thh.uhhlh..A.thhhs.lhsEhh............................
   consensus/70%              sG.huLLuGhsPpPLsLhhHhhAlslhshhhhh.sh...spP.thh.uhtll.hA.thlhPhlhuEhht.hh.........................
```
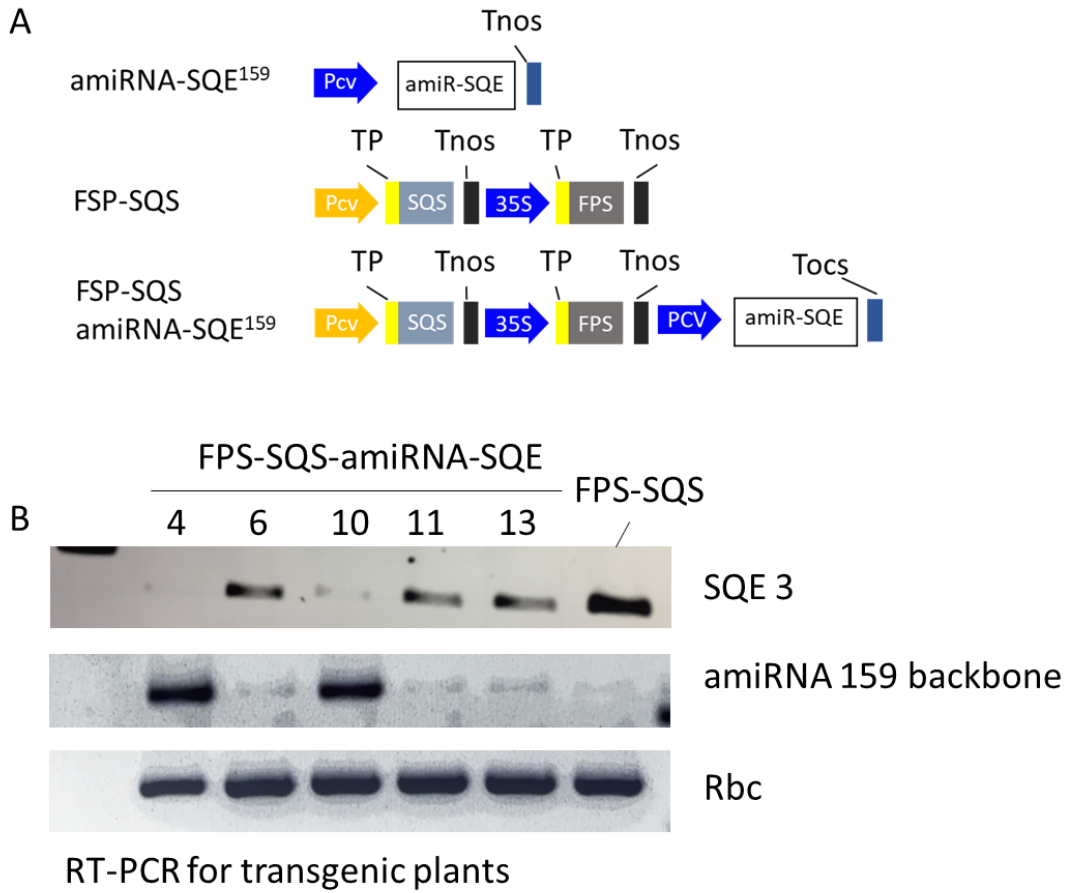
Figure 3-3. Continued

Figure 3-4. Plasmid designs and reverse-transcriptase PCR of transgenic plants. (A) plasmid designs and constructs used to silence squalene epoxidases in tobacco. (B) RT-PCR for SQE3 consensus sequence and specific sequences for each squalene epoxidase in tobacco from cDNA of leaf. TP, chloroplast transit peptide. Pcv, Cassava vein mosaic viral promoter. 35S, 35S Cauliflower mosaic viral promoter. Tnos, nos terminator.
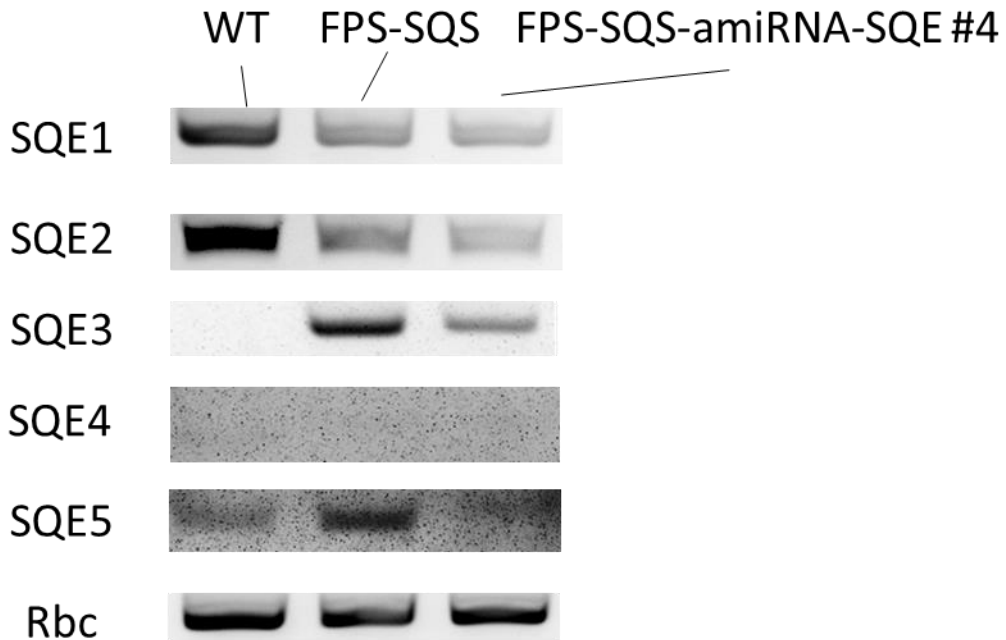
Figure 3-5. RNA silencing specificity analysis. RT-PCR for SQE1, SQE2, SQE3, SQE4, and SQE5 using cDNA of leaf samples as templates. Rbc, Rubisco small subunit as internal control.

The T0 generation transgenic plants were analyzed for squalene content, where squalene was not detectable in wildtype and amiRNA-SQE plants. The FPS-SQS-amiRNA-SQE plants showed higher yield of squalene compared with FPS-SQS lines (Figure 3-6). Seeds from most of the T0 lines could geminate except #12 and #15. The transgene copy number from positive transformants were verified by segregation of T1 generation plants in selection medium.
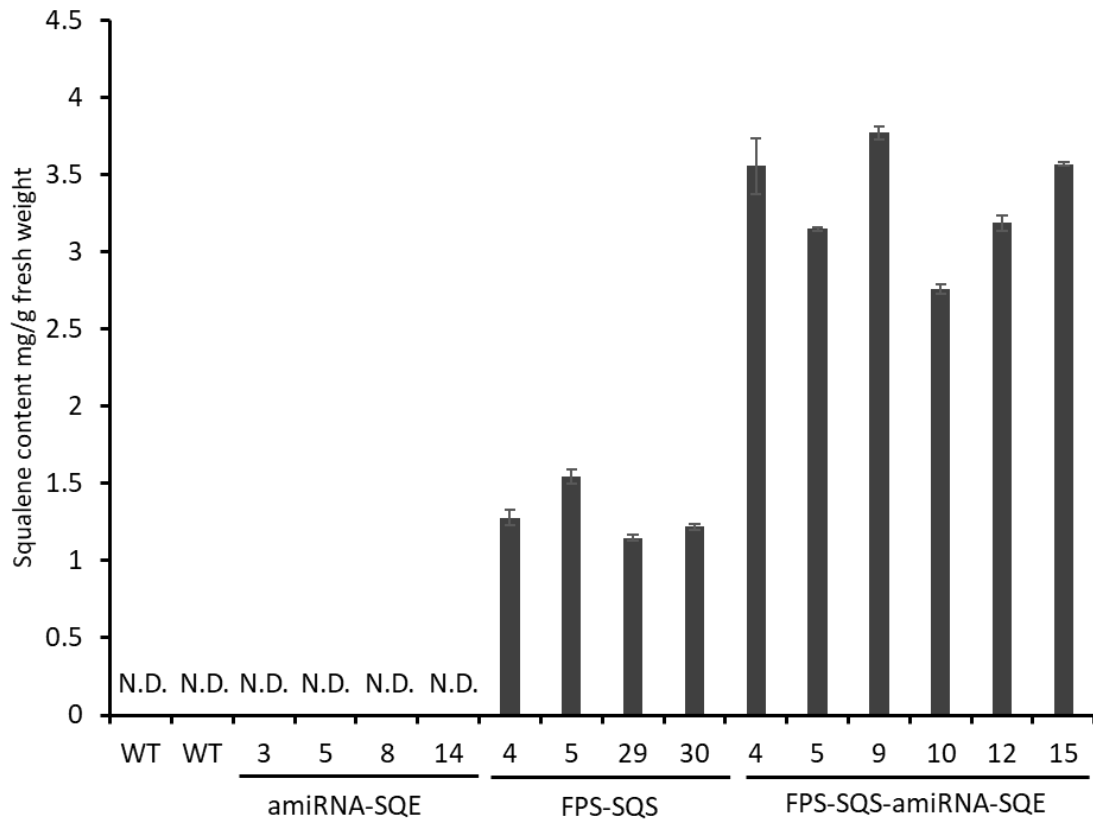
Figure 3-6. Squalene yield in T0 generation plants. Squalene contents were measured in lower leaves of 80 days old plant. Squalene content in each plant was measured by 3 individual replicates. Squalene detection limit was 50 μg/g fresh weight. N.D. not detectable. Error bars represent standard error.

The inactivation of squalene epoxidases significantly increased squalene level, when SQE3 was down-regulated. Squalene accumulation in single transgene copy T1 generation FPS-SQS-amiRNA-SQE3 showed 4 folds increase, compared to FPS-SQS plants (Figure 3-7). The highest one reached 3.98 mg/g fresh weight. The result indicated that SQEs activity was effectively down-regulated and the decrease of SQE3 expression significantly enhanced squalene production in plants. Moreover, the increase of squalene yield in squalene epoxidase silenced lines also proved squalene diffuses out of chloroplasts. In a word, the inter-compartmental "pull and block" is an effective strategy to enhance terpene yield.

3.3.4   Down-regulation of SQE3 leads to minor changes in sterol levels

In many studies, inactivation of sterol biosynthesis, including squalene epoxidases led to sterol content changes and growth defects[9,81,82]. To investigate whether sterol content or growth were influenced by the silencing of SQE3, the sterol content of wildtype, FPS-SQS, and multiple FPS-SQS-amiRNA-SQE plants were measured. As shown in Figure 3-8, sterol contents in FPS-SQS plants increased significantly, compared with wildtype plants, while sterols in FPS-SQS-amiRNA-SQE plants were similar sterol level in wildtype plants. The results suggest that enzyme activity of squalene epoxidases was partially inactivated.

We measured the total above-soiled biomass in young tobacco plants (Figure 3-9). The results showed the biomass in the early stage of FPS-SQS-amiRNA-SQE #10 plants had no significant difference and #4 plants had slight difference (P=0.055) by pair-wise t-test, compared with FPS-SQS plants. In a word, the sterol biosynthesis in SQE silenced plants was effectively downregulated without significant impact on the plant growth.
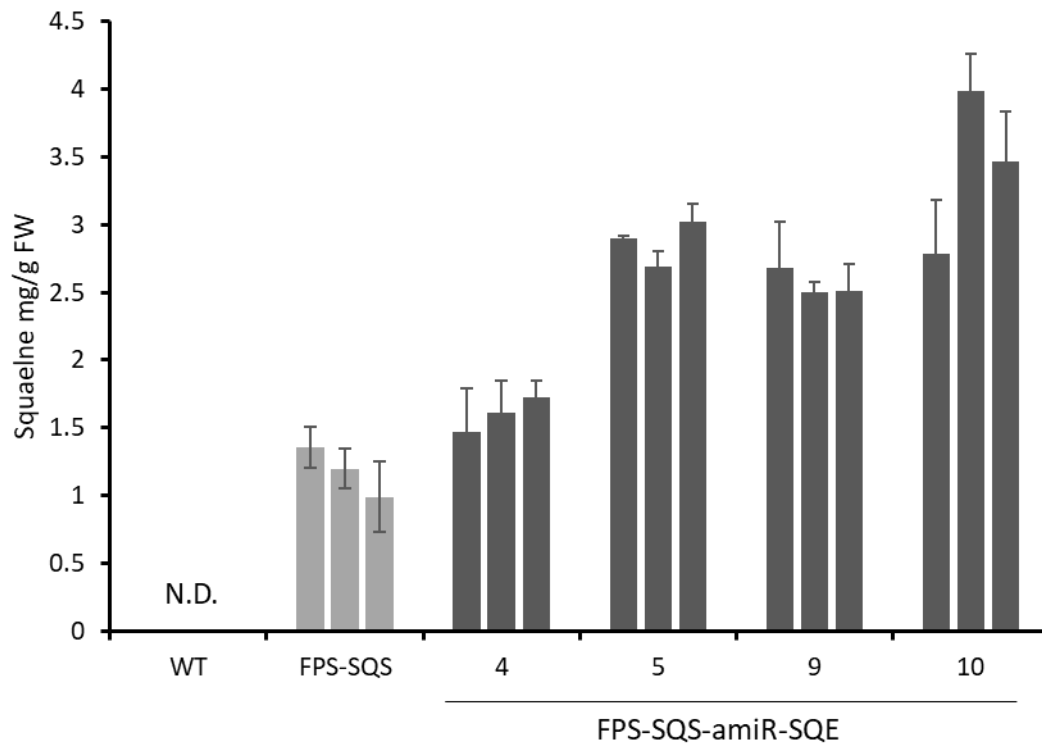
Figure 3-7. Squalene yield in T1 generation plants. Each transgenic event has 3 individual plants and each plant was analyzed for squalene content for 2 technical replicates. The figure represents squalene content in lower leaves of 90 days old plants. N.D. not detectable. Error bars represent standard error.
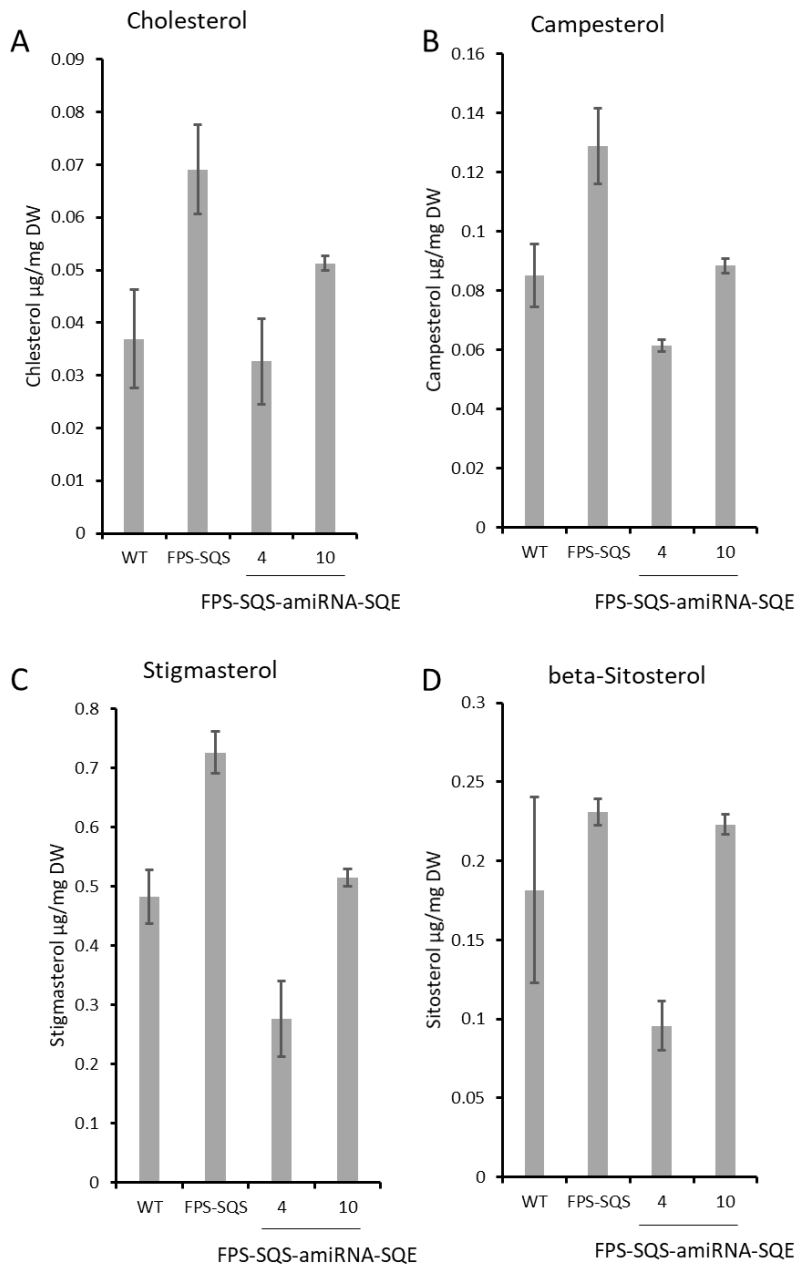
Figure 3-8. Sterol concentrations in T1 generation plants. Endogenous concentrations of cholesterol, campesterol, stigmasterol and beta-sitosterol were measured in tobacco leaf samples in 30 days old plants. Each transgenic event has 3 individual plants
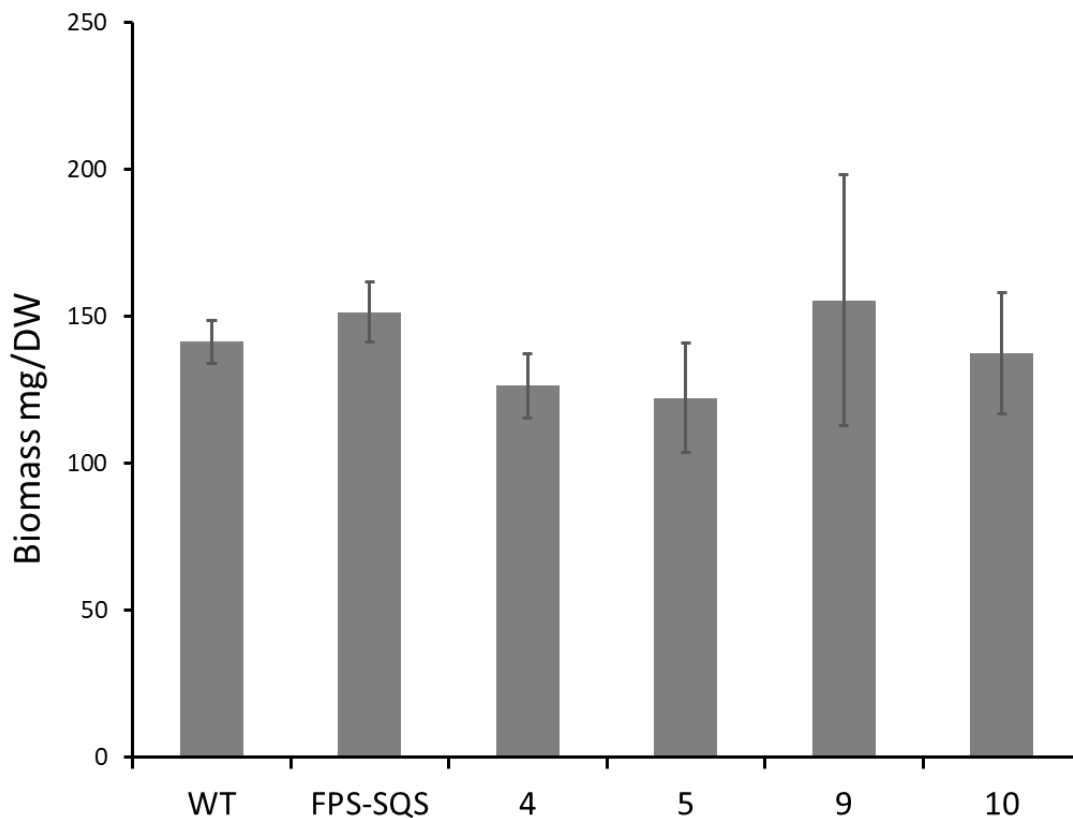
Figure 3-9. Total leaf biomass measurement. All leaves were collected from 30 days old T1 generation plants and were lyophilized for 48 hours in pre-weighed tubes. Total leaf biomass was calculated as the weight of lyophilized samples. Error bars represent standard error.

3.4 Discussion

The study presented a new inter-compartmental 'pull and block' strategy to increase squalene yield. The results not only reported the highest squalene yield in tobacco, but also highlighted the importance of preventing downstream conversion for natural products when engineering the bioproduct accumulation. The study verified the hypothesis that squalene epoxidases are key players for downstream conversion, even though squalene biosynthesis engineering is compartmentalized in chloroplasts. The results are also highly

66

consistent with our previous findings that squalene is membrane permeable and can 'leak' out of chloroplast[46]. Considering membrane permeability, in this engineering strategy it is crucial to 'block' the downstream conversion in the cytosol, even though the biosynthesis of squalene occurs mostly in chloroplasts. The inter-compartmental 'pull and block' strategy thus led to 3.98 mg/g fresh weight of squalene yield by inactivation of SQEs in cytosol and over-expressing FPS and SQS in chloroplasts.

The strategy well corroborated with the other study to prevent the downstream conversion of squalene. As aforementioned, methylation of squalene to prevent the downstream conversion, led to significant increase of methylated squalene. The study highlighted the importance of prevention of downstream conversion yet led to a product that is similar to but not exactly the same as squalene. Nevertheless, this study presented a strategy for plants to to produce the exact same squalene product as sharks and humans produce, which has a direct market value for various applications in cosmetics and pharmaceutical industries. In addition, we developed an alternative strategy to engineer a synthetic storage organelle within chloroplast to prevent the 'leaking' of squalene. The result led to efficient accumulation of squalene, yet the yield is still lower than those in this study. Overall, the inter-compartmental 'pull and block' strategy represents the most effective ones for blocking the downstream and enhancing squalene yield *in planta*.

Another important feature for the inter-compartmental 'pull and block' strategy is that the strategy did not significantly compromise plant growth. Over-accumulation of squalene in the chloroplast leads to the increase of sterols (Figure 3-8), which again proves the 'leaking' of squalene out of chloroplast membrane and the conversion of squalene by SQE into sterols. The selection of leaf-expressing SQE3, and the down-regulation of its

expression led to a slight decrease of sterol levels in line 10, but strong decrease in line 4. In most of high squalene plants (lines FPS-SQS-amiRNA-SQE 10), the squalene level remained similar with wildtype. The strategy thus represents an effective approach to increase squalene level without compromising plant growth. The high squalene productivity without compromising the squalene yield represented a unique set of engineered plants with potential for field implementation and production.

The inter-compartmental 'pull and block' strategy also represented a completely different approach for improving squalene levels. Such approach can be readily integrated with other strategies to further improve squalene level *in planta*. For example, the strategy can be integrated with synthetic organelles to further create down-stream sink and block conversion[46]. Such integration may lead to synergistic increase of squalene level. Moreover, the inactivation of downstream conversions could be integrated with other strategies to engineer higher yield of terpenes upstream to create a 'push, pull, and block' strategy. For instance, the inactivation of downstream degradation could be integrated with carbon concentrating mechanisms, higher photosynthesis rate, carbon redistribution strategies, and other ways to increase carbon fixation and squalene yield. Together, these strategies could increase squalene yield by increasing carbon flux toward terpene, enhancing the storage of end products, and reducing the loss rate[32,47,86]. Overall, the success of the 'pull and block' strategy offered a sustainable feedstock for high value squalene to meet the increasing demands in pharmaceutical, nutraceutical and cosmetic industries. The implementation of these technologies in tobacco also enabled the transition of tobacco industry into biomanufacturing platforms.

4. CONCLUSION AND PERSPECTIVE

The conversion of bioproducts within production systems limits the yield of metabolic engineering bioproducts. High yields of bioproducts can only be achieved with the fulfillment of three important conditions: an efficient biosynthesis, a sufficient sink capacity and a minimized conversion. In this study, two strategies were developed to enlarge the sink capacity and inhibit the conversion of terpenes to enhance terpene yield, with squalene as a model compound, in *N. tabacum*. Diffusion and instability of squalene with the presence of squalene biosynthesis pathway in chloroplast was demonstrated by the mathematic modeling and the dark treatment experiment. (1) In order to prevent the squalene leaking out of chloroplast to be oxidized by cytosol squalene epoxidases, a co-compartmentation strategy was established, in which synthetic lipid droplets were designed and implemented in chloroplast to enlarge the sink capacity of squalene, leading to significant increase of squalene accumulation. (2) An inter-compartmental "pull and block" strategy was developed. Squalene conversion enzymes were identified and silenced to inhibit the downstream conversion of squalene. Simultaneously, squalene biosynthesis pathway was overexpressed in chloroplasts to synthesize squalene.

Overall, there are several critical challenges to be addressed to increase bioproduct yield to a commercially viable level. First, biosynthesis pathways for unconventional bioproducts, such as terpenes, are highly regulated and subject to feedback inhibition, downstream conversion. Though many efforts have been made to investigate the regulation mechanism in terpene biosynthesis pathways, the regulation is still not very clear. Second, carbon partition toward terpene biosynthesis is limited. More efforts need to be made to

investigate the limitation of carbon partitioning from the angle of metabolic network. In a word, it is common for many researches that a ceiling emerges when titer of products research certain levels. The two strategies in the research can be integrated with each other and be integrated with other approaches of enhancing biosynthesis, to further enhance the production of terpenes. The research in this dissertation provide a new perspective to break through the ceiling in metabolic engineering.

REFERENCES

1.   Dixon, R. A. Natural products and plant disease resistance. *Nature* **411**, 843-847 (2001).

2.   Klayman, D. L. Qinghaosu (artemisinin): an antimalarial drug from China. *Science* **228**, 1049-1055 (1985).

3.   Ajikumar, P. K. *et al.* Isoprenoid pathway optimization for Taxol precursor overproduction in Escherichia coli. *Science* **330**, 70-74 (2010).

4.   O'Connor, S. E. & Maresh, J. J. Chemistry and biology of monoterpene indole alkaloid biosynthesis. *Natural product reports* **23**, 532-547 (2006).

5.   Bar-Even, A. & Tawfik, D. S. Engineering specialized metabolic pathways—is there a room for enzyme improvements?. *Current opinion in biotechnology* **24**, 310-319 (2013).

6.   Heinig, U., Gutensohn, M., Dudareva, N. & Aharoni, A. The challenges of cellular compartmentalization in plant metabolic engineering. *Current opinion in biotechnology* **24**, 239-246 (2013).

7.   Wang, X., Ort, D. R. & Yuan, J. S. Photosynthetic terpene hydrocarbon production for fuels and chemicals. *Plant biotechnology journal* **13**, 137-146 (2015).

8.   Gimpel, J. A., Specht, E. A., Georgianna, D. R. & Mayfield, S. P. Advances in microalgae engineering and synthetic biology applications for biofuel production. *Current opinion in chemical biology* **17**, 489-495 (2013).

9. Rasbery, J.M., Shan, H., LeClair, R.J., Norman, M., Matsuda, S.P. & Bartel, B. Arabidopsis thaliana squalene epoxidase 1 is essential for root and seed development. *Journal of Biological Chemistry* **282**, 17002-17013 (2007).

10. Banerjee, A. & Sharkey, T. Methylerythritol 4-phosphate (MEP) pathway metabolic regulation. *Natural product reports* **31**, 1043-1055 (2014).

11. Xiao, Y. *et al.* Retrograde signaling by the plastidial metabolite MEcPP regulates expression of nuclear stress-response genes. *Cell* **149**, 1525-1535 (2012).

12. Melis, A. Terpene Hydrocarbons Production in Cyanobacteria. *Cyanobacteria*, 187-198 (2017).

13. Lindberg, P., Park, S. & Melis, A. Engineering a platform for photosynthetic isoprene production in cyanobacteria, using Synechocystis as the model organism. *Metabolic engineering* **12**, 70-79 (2010).

14. Evans, J. R. Improving photosynthesis. *Plant physiology* **162**, 1780-1793 (2013).

15. Melis, A. Photosynthesis-to-fuels: from sunlight to hydrogen, isoprene, and botryococcene production. *Energy & Environmental Science* **5**, 5531-5539 (2012).

16. Sikkema, J., de Bont, J. A. & Poolman, B. Mechanisms of membrane toxicity of hydrocarbons. *Microbiological reviews* **59**, 201-222 (1995).

17. Reddy, L. H. & Couvreur, P. Squalene: A natural triterpene for use in disease management and therapy. *Advanced Drug Delivery Reviews* **61**, 1412-1426 (2009).

18. Köksal, M., Jin, Y., Coates, R. M., Croteau, R. & Christianson, D. W. Taxadiene synthase structure and evolution of modular architecture in terpene biosynthesis. *Nature* **469**, 116-120 (2011).

19. White, N. J. Qinghaosu (artemisinin): the price of success. *Science* **320**, 330-334 (2008).

20. Chan, H.-T. *et al.* Cold chain and virus free chloroplast-made booster vaccine to confer immunity against different polio virus serotypes. *Plant biotechnology journal*, **14**, 2190-2200 (2016).

21. Yuan, J. S. *et al.* Molecular and genomic basis of volatile-mediated indirect defense against insects in rice. *The Plant Journal* **55**, 491-503 (2008).

22. Gershenzon, J. & Dudareva, N. The function of terpene natural products in the natural world. *Nature Chemical Biology* **3**, 408-414 (2007).

23. Seemann, M., Tse Sum Bui, B., Wolff, M., Miginiac-Maslow, M. & Rohmer, M. Isoprenoid biosynthesis in plant chloroplasts via the MEP pathway: Direct thylakoid/ferredoxin‐dependent photoreduction of GcpE/IspG. *FEBS letters* **580**, 1547-1552 (2006).

24. Vranová, E., Coman, D. & Gruissem, W. Network analysis of the MVA and MEP pathways for isoprenoid synthesis. *Annual review of plant biology* **64**, 665-700 (2013).

25. Ghirardo, A. *et al.* Metabolic flux analysis of plastidic isoprenoid biosynthesis in poplar leaves emitting and nonemitting isoprene. *Plant Physiology* **165**, 37-51 (2014).

26. Wright, L. P. *et al.* Deoxyxylulose 5-phosphate synthase controls flux through the methylerythritol 4-phosphate pathway in Arabidopsis. *Plant Physiology* **165**, 1488-1504 (2014).

27.   Estévez, J. M., Cantero, A., Reindl, A., Reichler, S. & León, P. 1-Deoxy-D-xylulose-5-phosphate synthase, a limiting enzyme for plastidic isoprenoid biosynthesis in plants. *Journal of Biological Chemistry* **276**, 22901-22909 (2001).

28.   Banerjee, A. *et al.* Feedback inhibition of deoxy-D-xylulose-5-phosphate synthase regulates the methylerythritol 4-phosphate pathway. *Journal of Biological Chemistry* **288**, 16926-16936 (2013).

29.   Gao, X. *et al.* Engineering the methylerythritol phosphate pathway in cyanobacteria for photosynthetic isoprene production from CO 2. *Energy & Environmental Science* **9**, 1400-1411 (2016).

30.   Wang, X. *et al.* Enhanced limonene production in cyanobacteria reveals photosynthesis limitations. *Proceedings of the National Academy of Sciences* **113**, 14225-14230 (2016).

31.   Martin, V. J., Pitera, D. J., Withers, S. T., Newman, J. D. & Keasling, J. D. Engineering a mevalonate pathway in Escherichia coli for production of terpenoids. *Nature biotechnology* **21**, 796-802 (2003).

32.   Bentley, F. K., Zurbriggen, A. & Melis, A. Heterologous expression of the mevalonic acid pathway in cyanobacteria enhances endogenous carbon partitioning to isoprene. *Molecular plant* **7**, 71-86 (2014).

33.   Kumar, S. *et al.* Remodeling the isoprenoid pathway in tobacco by expressing the cytoplasmic mevalonate pathway in chloroplasts. *Metabolic engineering* **14**, 19-28 (2012).

34.   Dudareva, N. & Pichersky, E. Metabolic engineering of plant volatiles. *Current opinion in biotechnology* **19**, 181-189 (2008).

35.   Mahmoud, S. S., Williams, M. & Croteau, R. Cosuppression of limonene-3-
       hydroxylase in peppermint promotes accumulation of limonene in the essential oil.
       Phytochemistry **65**, 547-554 (2004).

36.   Englund, E. *et al.* Production of squalene in Synechocystis sp. PCC 6803. *PLoS
       One* **9**, e90270 (2014).

37.   Kajikawa, M. *et al.* Accumulation of squalene in a microalga Chlamydomonas
       reinhardtii by genetic modification of squalene synthase and squalene epoxidase
       genes. *PloS One* **10**, e0120446 (2015).

38.   Posé, D. *et al.* Identification of the Arabidopsis dry2/sqe1‑5 mutant reveals a
       central role for sterols in drought tolerance and regulation of reactive oxygen
       species. *The Plant Journal* **59**, 63-76 (2009).

39.   Li, J. *et al.* Gene replacements and insertions in rice by intron targeting using
       CRISPR–Cas9. *Nature plants* **2**, 16139 (2016).

40.   Jiang, W. *et al.* Demonstration of CRISPR/Cas9/sgRNA-mediated targeted gene
       modification in Arabidopsis, tobacco, sorghum and rice. *Nucleic acids research* **41**,
       e188-e188 (2013).

41.   O'Connor, S. E. Engineering of secondary metabolism. *Annual review of genetics*
       49, 71-94 (2015).

42.   Wu, S. *et al.* Redirection of cytosolic or plastidic isoprenoid precursors elevates
       terpene production in plants. *Nature biotechnology* **24**, 1441-1447 (2006).

43.   Malhotra, K. *et al.* Compartmentalized metabolic engineering for artemisinin
       biosynthesis and effective malaria treatment by oral delivery of plant cells.
       Molecular plant **9**, 1464-1477 (2016).

44. Pasoreck, E. K. *et al.* Terpene metabolic engineering via nuclear or chloroplast genomes profoundly and globally impacts off‑target pathways through metabolite signalling. *Plant biotechnology journal* **14**, 1862-1875 (2016).

45. Winichayakul, S. *et al.* In Vivo Packaging of Triacylglycerols Enhances Arabidopsis Leaf Biomass and Energy Density. *Plant Physiology* **162**, 626-639 (2013).

46. Zhao, C. *et al.* Co-Compartmentation of Terpene Biosynthesis and Storage via Synthetic Droplet. *ACS synthetic biology* **7**, 774-781 (2018).

47. Keasling, J. D. Manufacturing molecules through metabolic engineering. *Science* **330**, 1355-1358 (2010).

48. Banerjee, A. & Sharkey, T. D. Methylerythritol 4-phosphate (MEP) pathway metabolic regulation. *Natural Product Reports* **31**, 1043-1055 (2014).

49. Brennan, T. C., Turner, C. D., Krömer, J. O. & Nielsen, L. K. Alleviating monoterpene toxicity using a two‑phase extractive fermentation for the bioproduction of jet fuel mixtures in Saccharomyces cerevisiae. *Biotechnology and bioengineering* **109**, 2513-2522 (2012).

50. Aziz, N., Paiva, N. L., May, G. D. & Dixon, R. A. Transcriptome analysis of alfalfa glandular trichomes. *Planta* **221**, 28-38 (2005).

51. Besser, K. *et al.* Divergent Regulation of Terpenoid Metabolism in the Trichomes of Wild and Cultivated Tomato Species. *Plant Physiology* **149**, 499-514 (2009).

52. Zulak, K. G. & Bohlmann, J. Terpenoid biosynthesis and specialized vascular cells of conifer defense. *Journal of Integrative Plant Biology* **52**, 86-97 (2010).

53.   Wu, S. *et al.* Engineering triterpene metabolism in tobacco. *Planta* **236**, 867-877 (2012).

54.   Hansch, C. & Dunn, W. J. Linear relationships between lipophilic character and biological activity of drugs. *Journal of Pharmaceutical Sciences* **61**, 1-19 (1972).

55.   Abell, B. M., Hahn, M., Holbrook, L. A. & Moloney, M. M. Membrane topology and sequence requirements for oil body targeting of oleosin. *The Plant Journal* **37**, 461-470 (2004).

56.   Horn, P. J. *et al.* Identification of a new class of lipid droplet-associated proteins in plants. *Plant Physiology***162**, 1926-1936 (2013).

57.   Walther, T. C. & Farese, R. V., Jr. Lipid droplets and cellular lipid metabolism. *Annual review of biochemistry* **81**, 687-714 (2012).

58.   Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature methods* **6**, 343-345 (2009).

59.   Lung, S.-C., Smith, M. D. & Chuong, S. D. Isolation of Chloroplasts from Plant Protoplasts. *Cold Spring Harbor Protocols* **2015**, pdb. prot074559 (2015).

60.   Xie, S., Sun, S., Dai, S. Y. & Yuan, J. S. Efficient coagulation of microalgae in cultures with filamentous fungi. *Algal Research* **2**, 28-33 (2013).

61.   van Wijk, K. J., Peltier, J.-B. & Giacomelli, L. Isolation of chloroplast proteins from Arabidopsis thaliana for proteome analysis. *Plant Proteomics: Methods and Protocols*, 43-48 (2007).

62.   Zhang, Y. *et al.* Application of an improved proteomics method for abundant protein cleanup: molecular and genomic mechanisms study in plant defense. *Molecular & Cellular Proteomics* **12**, 3431-3442 (2013).

63.   Yuan, S. *Investigating the Molecular Basis of Volatilemediated Plant Indirect Defense against Herbivorous Insects Using Functional and Comparative Genomics* PhD thesis, University of Tennessee, (2007).

64.   Abell, B. M. *et al.* Role of the proline knot motif in oleosin endoplasmic reticulum topology and oil body targeting. *The Plant Cell* **9**, 1481-1493 (1997).

65.   Abell, B. M., High, S. & Moloney, M. M. Membrane protein topology of oleosin is constrained by its long hydrophobic domain. *Journal of Biological Chemistry* **277**, 8602-8610 (2002).

66.   Weiss, T. L. *et al.* Colony organization in the green alga Botryococcus braunii (Race B) is specified by a complex extracellular matrix. *Eukaryotic Cell* **11**, 1424-1440 (2012).

67.   Agarwal, U. P. Raman imaging to investigate ultrastructure and composition of plant cell walls: distribution of lignin and cellulose in black spruce wood (Picea mariana). *Planta* **224**, 1141-1153 (2006).

68.   Gierlinger, N. & Schwanninger, M. Chemical imaging of poplar wood cell walls by confocal Raman microscopy. *Plant Physiology* **140**,   1246-1254 (2006).

69.   Schmidt, M. *et al.* Label-free in situ imaging of lignification in the cell wall of low lignin transgenic Populus trichocarpa. *Planta* **230**, 589-597 (2009).

70.   Sun, L., Simmons, B. A. & Singh, S. Understanding tissue specific compositions of bioenergy feedstocks through hyperspectral Raman imaging. *Biotechnology and bioengineering* **108**, 286-295 (2010).

71. Jiang, Z., Kempinski, C., Bush, C. J., Nybo, S. E. & Chappell, J. Engineering triterpene and methylated triterpene production in plants provides biochemical and physiological insights into terpene metabolism. *Plant Physiology*, 01548 (2015).

72. Vidi, P.-A., Kessler, F. & Bréhélin, C. Plastoglobules: a new address for targeting recombinant proteins in the chloroplast. *BMC biotechnology* **7**, 4-15 (2007).

73. O'Hagan, D. T., Rappuoli, R., De Gregorio, E., Tsai, T. & Del Giudice, G. MF59 adjuvant: the best insurance against influenza strain diversity. *Expert review of vaccines* **10**, 447-462 (2011).

74. Andrianov, V. *et al.* 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 Biotechnology Journal* **8**, 277-287 (2010).

75. Matheson, N. & Wheatley, J. Starch changes in developing and senescing tobacco leaves. *Australian Journal of Biological Sciences* **15**, 445-458 (1962).

76. Xu, W., Ma, X. & Wang, Y. Production of squalene by microbes: an update. *World Journal of Microbiology and Biotechnology* **32**, 195-202 (2016).

77. Fox, C. B. Squalene emulsions for parenteral vaccine and drug delivery. *Molecules* **14**, 3286-3312 (2009).

78. Kim, S.-K. & Karadeniz, F. Biological importance and applications of squalene and squalane. *Adv. Food Nutr. Res* **65**, 223-233 (2012).

79. Maisashvili, A., Bryant, H. L. & Richardson, J. W. Economic feasibility of tobacco leaves for biofuel production and high value squalene. *International Food and Agribusiness Management Review* **19**, 145-162 (2016).

80. Dahl, R. H. *et al.* Engineering dynamic pathway regulation using stress-response promoters. *Nature biotechnology* **31**, 1039-1046 (2013).

81. Wentzinger, L. F., Bach, T. J. & Hartmann, M.-A. Inhibition of squalene synthase and squalene epoxidase in tobacco cells triggers an up-regulation of 3-hydroxy-3-methylglutaryl coenzyme A reductase. *Plant Physiology* **130**, 334-346 (2002).

82. Laranjeira, S. *et al.* Arabidopsis Squalene Epoxidase 3 (SQE3) complements sqe1 and is important for embryo development and bulk squalene epoxidase activity. *Molecular plant* **8**, 1090-1102 (2015).

83. Keene, C. K. & Wagner, G. J. Direct demonstration of duvatrienediol biosynthesis in glandular heads of tobacco trichomes. *Plant physiology* **79**, 1026-1032 (1985).

84. McCormick, S. *et al.* Leaf disc transformation of cultivated tomato (L. esculentum) using Agrobacterium tumefaciens. *Plant Cell Reports* **5**, 81-84 (1986).

85. Ryder, N. Terbinafine: mode of action and properties of the squalene epoxidase inhibition. *British Journal of Dermatology* **126**, 2-7 (1992).

86. Melis, A. Carbon partitioning in photosynthesis. *Current opinion in chemical biology* **17**, 453-456 (2013).